소셜미디어 기반 의사결정 지원을 위한 이벤트 템플릿 추출

허정[O], 류법모, 최윤재, 김현기
지식마이닝연구팀, 한국전자통신연구원
jeonghur, pmryu, mp2893, hkk@etri.re.kr

Event Template Extraction for the Decision Support based on Social Media

Jeong Heo[O], Pum-Mo Ryu, Yoon-Jae Choi, Hyun-Ki Kim
Knowledge Mining Research Team
Electronics and Telecommunications Research Institute

요 약

본 논문은 소셜 미디어 기반 의사결정 지원 시스템인 '소셜위즈덤'에 포함된 이벤트 템플릿 추출에 대해서 소개한다. 의사결정 지원 시스템은 경제적, 사회적 중요사항을 결정할 수 있도록 관련 정보와 인사이트(Insight)를 제공하는 정보시스템을 이른다. 기존 시스템은 단지 특정 키워드 빈도나 공기하는 키워드들의 관계만을 제공하였다. 그러나, 소셜위즈덤은 이벤트로 정의되는 주체(Subject), 이벤트 속성(Event-Property), 객체(Object)의 트리플(Triple) 집합인 템플릿을 추출하여 이를 기반으로 이벤트 정보를 함께 제공한다. 템플릿 추출은 고정밀 언어분석의 관계추출 기술과 온톨로지에 기반한 템플릿 제약 및 필터링 규칙을 이용하였다. 수작업으로 구축한 평가데이터로 평가한 결과, 템플릿 추출 성능(F-Score)은 뉴스 0.544, 블로그 0.3386, 트위터 0.3251이고 전체 통합 성능은 0.4648이었다. 필터링 성능(Accuracy)은 뉴스 0.7257, 블로그 0.6122, 트위터 0.6207이고 전체 통합 성능은 0.722이었다.

주제어: 이벤트, 템플릿, 의사결정지원, 소셜 미디어

1. 서론

소셜 미디어와 모바일 기기의 활성화로 다양한 형태의 정보가 폭발적으로 축적되고 있다. 로그, 텍스트 콘텐츠, 멀티미디어 콘텐츠 등 다양한 형태의 대용량 데이터를 처리하고 관리하는 것의 중요성이 부각되면서 '빅데이터(Big Data)' 분석에 대한 요구가 급증하고 있으며 많은 연구가 진행되고 있다. 빅데이터 분석은 도구, 플랫폼, 분석방법 등 포괄적인 기술적 개념으로 사용되고 있다[1].

빅데이터 분석의 대표적인 사례는 IBM에서 개발한 DeepQA인, Watson이 있다[2]. Watson은 2011년 2월 TV 퀴즈 프로그램인, 'Jeopardy! 퀴즈쇼'에서 인간 챔피언들과의 경쟁에서 이기면서 크게 주목을 받았다. Watson은 고성능의 하드웨어(90대의 IBM Power 750서버)를 기반으로 대용량의 콘텐츠(약 200억 페이지 분량)를 대상으로 고정밀 자연어 분석을 통해 구조화된 지식을 구축하여, 질문에 대한 정답을 제시하는 질의응답 시스템(Question Answering System)이다.

다른 사례로는 Recorded Future가 있다[3]. Recorded Future는 웹 데이터를 기반으로 웹 인텔리전스 및 예측 분석을 전문으로 하는 회사이다. 비구조화 텍스트 콘텐츠를 대상으로 기 정의한 이벤트에 대해서 정보를 추출, 분석하고, 시각화하여 제시하고 있다. 이벤트별 긍/부정 정보와 사용자의 관심도, 흥미도, 중요도의 관점에서 모멘텀(Momentum) 정보를 시간 축을 기준으로 제시하고 있다.

2. 의사결정 지원 시스템

의사결정 지원 시스템은 경제적, 사회적 중요사항을 결정할 수 있도록 관련 정보와 인사이트(Insight)를 제공하는 정보시스템을 이른다.

빅데이터 분석의 중요한 소스인 소셜 미디어에는 정치, 경제, 사회문화적인 이슈에 대한 다양한 의견 및 사회적 행동 패턴이 잠재되어 있다. 이와 같이 소셜 미디어에 잠재되어 있는 중요 정보를 파악하여 분석할 수 있다면, 기업 및 공공단체에서 선행적인 의사결정을 통해 많은 기회를 창출할 수 있다. 이와 같은 특성으로 인해 많은 의사결정 지원 시스템은 소셜 미디어를 핵심 분석 대상으로 인식하고 있다. 소셜 미디어 상의 여론 추이 및 특정 브랜드, 인물, 정책 등의 호불호를 파악하거나, 리스크를 감지하기 위해서 다양한 유형의 소셜 미디어 분석 기술이 연구되고 있다.

[4]는 지진과 태풍 같은 이벤트가 트위터 상에서 시공간적으로 전파되는 양상을 파악하는 확률 모델에 대해서 연구하였다. [5]는 선거철에 소셜 미디어 상의 유권자 의견을 분석하여 선거의 결과를 예측하는 기술에 대해서 소개하고 있다.

응용사례로는 다음소프트의 소셜 매트릭스를 비롯한 다양한 시스템들[1]이 있다. 소셜 매트릭스는 자연어처리 기술과 텍스트마이닝 기술을 바탕으로 블로그, 트위터 문서를 분석하여 사용자의 요구에 따라 해당 키워드의

1) 트루스토리(솔트룩스),펄스K(코난테크놀로지),씨날(그루터)

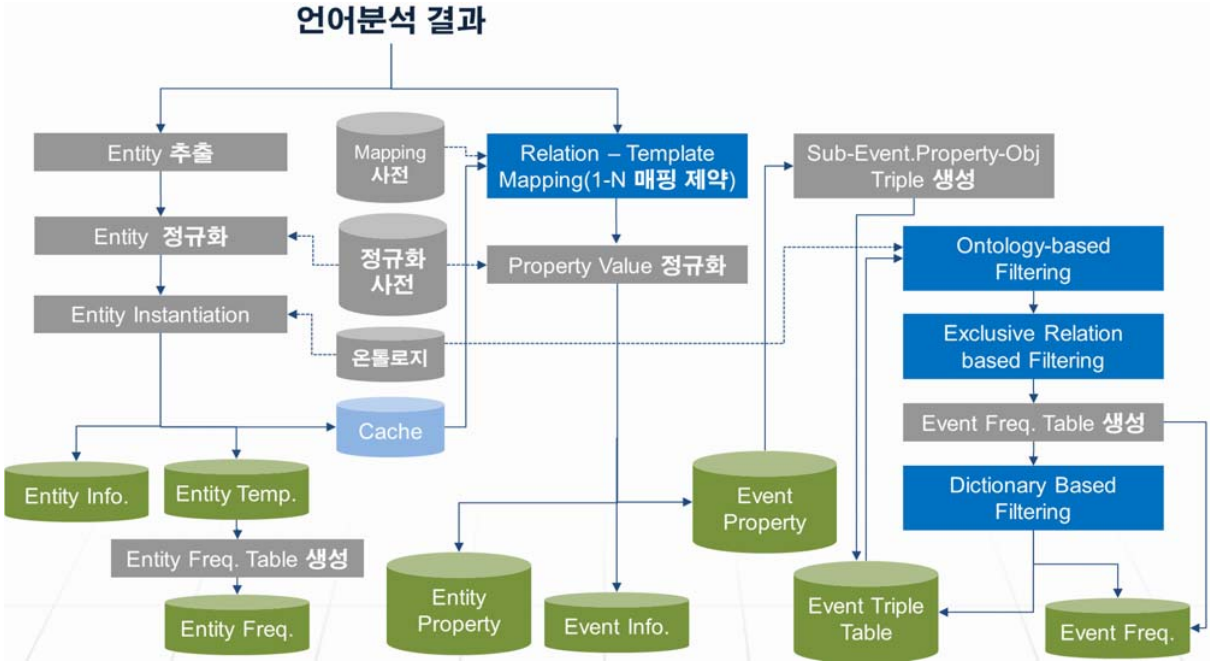

그림 1. 이벤트 템플릿 추출 및 필터링 구조도

빈도 추이 및 감성정보들을 모니터링할 수 있다. 소셜 매트릭스는 단지 개체 단위의 정보만을 제공하고 있으며, 개체들 간의 연관성에 기반한 이벤트 단위의 정보 분석은 지원하지 않는다. 이로 인해 개체들 간의 의미적 연관성 파악이 어려워 의사결정 지원에 한계가 있다. 이런 단점을 보완하기 위해서 본 논문에서는 개체들 간의 의미적 관계를 정의한 이벤트들을 템플릿 구조로 정의하고 소셜 미디어에서 이벤트 템플릿을 추출하는 기술에 대해서 소개한다.

3. 이벤트 템플릿

본 논문에서는 이벤트를 '주체(Subject)', '이벤트 유형(Event-Property)', '객체(Object)'로 구성된 트리플(Triple)의 집합인 템플릿(Template)으로 정의한다. 이벤트 유형은 이벤트 이름과 속성 이름으로 구성된다.

이벤트 템플릿 선정은 기업과 공공분야를 대상으로 의미적 가치를 고려하여 선정하였으며, 31개의 이벤트 템플릿으로 구성된다. <표 1>은 이벤트 템플릿의 목록으로써, 속성에서 **볼드체**로 표시된 속성명은 주체에 해당되는 속성이다.

4. 템플릿 기반 이벤트 추출

템플릿 기반 이벤트 추출은 고정밀 언어분석을 기반으로 수행된다. 언어분석은 형태소분석, 개체명인식, 관계추출이 수행된다. 개체명인식은 약 180개의 다양한 개체명을 분류하고 태깅하는 ETRI 개체명 인식기를 이용하였다[6]. 관계추출은 인식된 개체명들 간의 관계를 분석하는 기술로써, N-ary Relation 결과를 제공한다. 관계 추출된 개체명들은 온톨로지(Ontology)에 기반한 개체연결(Entity Linking or Entity Instantiation)을 수행하여 개체의 클래스(Class)와 인스턴스(Instance) 정보를 파악한다. 또한, 관계 분석된 개체명들 간의 관계는 이벤트 속성으로 매핑이 수행된다. 매핑관계는 1:N으로 이벤트 매핑 모호성(Ambiguity)이 발생한다. 모호성 해소는 개체들의 온톨로지 연결정보에 기반한 규칙으로 처리한다. 모호성 해소와 함께 이벤트 템플릿 필터링을 수행하여 오류에 해당하는 관계를 제거한다. 추출된 이벤트의 속성값은 의미적으로 동일한 객체(Object)로 인식될 수 있으나, 표현형태가 다른 많은 이형태로 구성된다. 예를 들면, 'iphone 4S'와 '아이폰 4s'는 동일한 객체이지만 그 표현 형태는 다르다. 이를 처리하기 위해서 정규화를 수행한다. 정규화 모듈은 개체명 정규화, 날짜 정규화, 지역명 정규화, 가격 정규화로 구성된다. 개체명 정규화는 위키피디아 Redirection정보를 기반으로 (반)자동으로 구축한 사전을 이용한다. 날짜와 가격 정규화는 정규표현식(Regular Expression)을 기반한 규칙을 이용하고, 지역명 정규화는 주소사전(Gazetteer) 기반의 규칙을 활용한다.

개체명 분할 문제는 속성값의 모호성 해소와 밀접하게 관련된다. 예를 들면, "빙그레가 러시아 스낵시장에 본격 진출한다."의 문장에서 '빙그레(Subject) - CompanyExpansion.expansiontype - 스낵시장(Object)'의 트리플을 추출하였을 때, 객체(Object)의 값이 '러시아 스낵시장'이 되는 것이 보다 명확하다. 이를 위해서, 분할되어 개체명 태깅이 된 두 개체들을 결합할 수 있도록 개체명 태그열 규칙을 이용하여 개체명을 결합한다.

5. 이벤트 제약 및 필터링

이벤트 속성 매핑이 완료되면, 템플릿은 주체, 이벤트 속성, 객체의 트리플 집합으로 구성된다. 구성된 이벤트 템플릿은 일관성(Consistency) 유지를 위하여 제약규칙(Constraint Rule)에 따른 필터링(Filtering)을 수행한다. 제약규칙은 크게 네 가지 유형으로 나뉜다.

가) 온톨로지에 기반한 제약
나) 배타적 관계(Exclusive Relation)에 기반한 제약
다) 필수 속성(Required Property)에 기반한 제약
라) 사전(Dictionary)에 기반한 제약

온톨로지에 기반한 제약은 개체와 온톨로지 연결정보에 따른 제약으로써, 개체 카테고리(Entity Category[2)])와 개체 유형(Entity Type[3)]) 에 따른 제약으로 구분된다. 템플릿의 속성별로 유효한 개체 카테고리가 있다. 예를 들면, ProductRecall.company의 속성값으로 COMPANY 카테고리만이 유효하다. COMPANY가 아닌 다른 카테고리의 경우 필터링 된다. 또한, 템플릿의 속성별로 유효한 개체 유형이 있다. 예를 들면, ProductRelease.product의 속성값은 인스턴스 유형만이 유효하다. 즉, 개체 유형이 클래스인 경우는 필터링 된다.

배타적 관계에 해당하는 이벤트 템플릿은 다음과 같다.

- zStockRise.company - zStockDecline.company
- zPriceRise.product - zPriceDecline.product
- zPriceRise.fee - zPriceDecline.fee
- zPolicySupport.organization- zPolicyOpposition.organization
- zPolicySupport.person zPolicyOpposition.person

이벤트 템플릿은 하루 단위로 통합되어 빈도정보를 저장한다. 따라서 배타적 관계에 해당하는 템플릿은 빈도 정보에 의해서 하나의 이벤트 템플릿을 선택하고, 선택되지 않은 템플릿은 필터링 된다.

이벤트 템플릿 별로 주체를 제외한 객체 속성들 중, 반드시 값이 있어야 하는 속성을 정의하고 있다. 예를 들면, EmploymentChange.positiontitle은 반드시 값이 채워져야 한다. 그러나 해당 속성값이 채워지지 않았을 경우, 해당 템플릿은 필터링 된다.

사전(Dictionary)에 기반한 제약은 템플릿 추출에서 고빈도로 발생하는 오류를 사전에 규칙으로 기술하고, 이를 기반으로 템플릿을 제약한다.

2) Entity Category : Entity가 연결(Linking)된 온톨로지 클래스(Class) 이름

3) Entity Type : Entity가 온톨로지의 Class인지, Instance인지 여부

표 1. 이벤트 템플릿 목록

이벤트 명	개요	속성 목록
Acquisition	인수합병	**acquirer**, beingacquired, date, price
CompanyCompetitor	경쟁기업	**company1**, company2
CompanyEarning Announcement	기업실적	**company**, date, earning
CompanyExpansion	기업확장	**company**, date, expansiontype
CompanyInvestment	기업투자	**company**, date, target, price
CompanyLegalIssue	법적이슈 (기업)	**company_plaintff**, **person_plaintiff**, date, company_sued
CompanyMeeting	사업미팅	**company**, date, meetingtype
CompanyTechnology	보유기술	**company**, technology
CreditRating	신용평가	**rated_org**, date, rating_org
EmploymentChange	직책변동	**person**, date, positiontitle, organization
ProductRecall	제품리콜	**product**, date, company
ProductRelease	제품출시	**product**, company, date, price
zAdvertisingStart	광고	**product**, date, company
zAnnounce	발표	**organization**, **person**, date, location
zAward	수상	**person**, **organization**, prize, date, product
zEventOpen	행사	**event**, date, organization, person, location
zInvestigate	조사	**person**, **event**, date, organization
zLegalAct	법적규제	**organization**, date, legalaction
zMarketShare	시장 점유율	**company**, date, ratequantity, product
zPersonDie	부고	**person**, date, reason
zPersonTravel	여행출장	**Person**, date, destination
zPolicyEnforce	정책시행	**policy**, date, organization
zPolicyOpposition	정책반대	**policy**, date, organization, person
zPolicySupport	정책지지	
zPriceDecline	가격하락	**product**, **fee**, date, organization, pricechange
zPriceRise	가격상승	
zRecommend	추천	**product**, person, organization
zStockDecline	주가하락	**company**,date, stockrisecontent
zStockRise	주가상승	
zStockList	주식상장	**company**, date
zVote	투표/선거	**vote**, date

6. 평가

평가를 위해서, 뉴스, 블로그, 트위터로 구성된 평가데이터를 구축하였다. <표 2>는 평가데이터를 구성하는 소스별 분포와 성능에 대해서 기술하고 있다. 평가데이터는 문서 생성 시간(Document Created Time)이 없는 관계로 이벤트 제약 규칙 중, 나)는 적용하지 않고 평가하였다. (그림 2)는 평가방법으로써, 필터링 전/후의 정밀도(Precision)과 재현율(Recall) 계산방법, 필터링의 정확도(Accuracy) 계산방법을 벤다이어그램으로 설명하고 있다.

표 2. 소스 별 평가데이터 분포와 성능

소스	문장 수	템플릿 수	문장별 템플릿 수	Filtering 전(Micro-Average)			Filtering 후(Micro-Average)			Filtering Accuracy
				Precision	Recall	F-Score	Precision	Recall	F-Score	
뉴스	1,118	2,060	1.84	0.7234	0.4519	0.5563	0.7429	0.4291	0.5440	0.7257
블로그	806	1,058	1.31	0.5986	0.2495	0.3522	0.6160	0.2335	0.3386	0.6122
트위터	806	757	0.94	0.6276	0.2404	0.3477	0.6395	0.2180	0.3251	0.6207
전체	2,730	3,875	1.42	0.6824	0.3584	0.4700	0.7073	0.3461	0.4648	0.7220

평가결과, 제약에 따른 필터링은 정밀도 향상에는 도움이 되나, 재현율의 성능을 저하시켰다. F-Score도 재현율의 성능저하로 인해 필터링 후, 성능이 하락하는 경향을 보였다. 그러나 재현율보다는 정밀도의 성능이 중요한 질의응답이나 정보추출 분야에서는 제약규칙에 따른 필터링이 중요하다.

블로그와 트위터보다 뉴스가 문장당 템플릿 수가 많으며 성능도 우수하였다. 뉴스는 문법적으로 올바른 문장으로 구성되기 때문에 비문이 많은 블로그와 트위터에 비해서 좋은 성능을 보였다. 또한, 필터링의 성능도 우수하였다.

오류가 사용자에게 제시되지 않아야 하는 질의응답과 정보추출 분야에서는 빅데이터에 기반한 정밀도 중심의 접근법이 효율적일 수 있다. 이를 위해서는 재현율의 성능 저하는 있지만, 정밀도의 성능 개선이 있는 필터링이 중요하며, 비교적 비문이 적은 뉴스나 백과사전(위키피디아) 등을 활용하는 것이 바람직할 것이다.

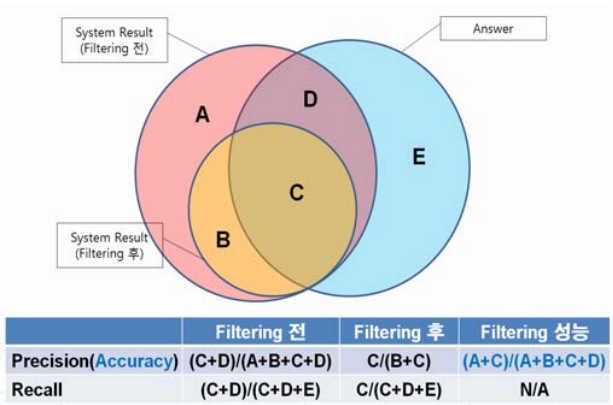

그림 2. 평가 방법

7. 오류 분석

이벤트 템플릿 추출 오류를 분석한 결과, 크게 4종류의 오류로 분류할 수 있다.

가) 온톨로지에 기반한 개체 연결 문제
나) 개체명 범위(Entity Boundary) 인식 문제
다) 개체명 정규화 문제
라) 날짜 정규화 문제

온톨로지에 기반한 개체연결 문제는 관계로 추출된 개체가 온톨로지의 클래스로 연결시 발생하는 모호성 오류이다. 예를 들면, “강릉시가 녹색생활 실천 문화 확산을 위해 환경부에서 도입하고 있는 그린카드제를 시행해 주목된다.”라는 문장에서 ‘강릉시’가 ORGANIZATION 클래스로 연결되어야 하지만, LOCATION으로 연결되는 경우이다.

개체명 범위 인식 문제는 개체명의 범위를 어디까지 볼 것인가 하는 문제이다. 예를 들면, “LG도 무안경 3D 기술을 보유하고 있지만, 시야각이 좁고 대형화하기 어려운 한계가 있다.”라는 문장에서 TECHNOLOGY 개체명의 범위로 ‘무안경 3D 기술’까지 인식되어야 하지만, ‘3D 기술’로 인식된 경우이다.

개체명 정규화 문제는 개체명의 의미적 모호성으로 발생하는 문제이다. 예를 들면, “MS가 최근 개발 과정에서 결함이 발견되어 SBS 제품 업데이트를 리콜한다고 지난 금요일 밝혔다.”에서 ‘SBS’는 ‘Small Business Serve’의 약어로써, PRODUCT에 해당된다. 그러나, 본 시스템에서는 ‘SBS’를 ‘서울방송’으로 정규화하였다. 이 오류는 ‘SBS’에 대한 의미정보인 개체명 태그나 온토로지에 기반한 개체연결 정보로 의미적인 모호성을 해소한다면 해결할 수 있다. 그러나, 정규화 사전에 관련 의미정보를 부착하는 것은 시간적으로 많은 노력이 요구된다.

날짜 정규화 문제는 다양한 형태의 날짜 표현을 ‘YYYYMMDD(날짜)’과 ‘YYYYMMDD~YYYYMMDD(범위)’형태로 정규화하면서 발생하는 문제이다. 날짜 표현은 절대적 표현(Absolute Expression)과 상대적 표현(Relative Expression)으로 구분될 수 있다. 절대적 표현은 ‘YYYY년 MM월 DD일’과 같이 명확한 시간적 지점을 표현한 형태이고, 상대적 표현은 기준 시간에 대한 상대적 시간을 표현한 것으로써, ‘어제’, ‘작년’등과 같은 표현이다. 상대적 표현은 일반적으로 문서 생성 시간(Document Created Time)을 기준시간으로 한 표현과 문장 내에서 기술된 절대시간을 기준으로 한 표현으로 구분된다. 그러나 본 시스템에서는 문서가 수집된 시간(Crawling Time)을 기준시간으로 하여 정규화 하였다. 이로 인해 문장 내에 표현된 절대적 시간 표현을 기준으로 상대적 시간이 표현된 경우, 날짜 정규화 오류가 발생하였다. 예를 들면, “농림수산식품부는 10일 농수산물의 원산지표시에 관한 법률 시행령이 법제처 심의를 거쳐 10월 10일 공포됨에 따라 6개월 후부터 본격 시행될 예정이라고 밝혔다.”(해당 문장이 포함된 문서의 수집 날짜는 2011년 10월 10일)에서 ‘농수산물의

원산지 표시에 관한 법률'의 시행일은 '10월 10일'에서 6개월 후의 날짜로 정규화 되어야 한다. '10월 10일'의 정규화된 날짜는 문서 수집 날짜를 기준으로 '2011년 10월 10일'이 되고, 이를 기준시간으로 하여 '6개월 후'인 '2012년 04월 11일'로 정규화 되어야 한다. 시간표현(Temporal Expression)에 대한 연구는 TAC(Text Analysis Conference) - KBP 2011(Knowledge Base Population)에서 TSF(Temporal Slot Filling)의 핵심 기술로 활발히 연구되고 있다[7].

8. 응용시스템(소셜위즈덤)

이벤트 템플릿 추출 결과는 소셜위즈덤 서비스에서 시간별 이벤트 추이 변화를 제공한다. 특정 개체(키워드)와 시간범위가 질의로 입력되었을 경우, 개체와 관련된 주요한 빈도/중요도 정보, 긍/부정 정보와 이벤트 정보를 제공한다. 해당 정보는 소스별 차이를 파악할 수 있도록 소스별로 구분하여 제공하고 있다.

(그림 3)은 소셜위즈덤의 화면을 갈무리한 것이다. '삼성전자' 키워드에 대한 한 달간4)의 정보를 제공하고 있다. 화면의 좌측 상단에 소스를 선택할 수 있고, 선택된 소스에 대한 이벤트 순위화 목록이 좌측 하단에 표시된다. 상위 5개의 이벤트는 우측 하단 시간별 주요 이벤트 창에 표현된다. 원의 사이즈는 이벤트 빈도, 속성값의 빈도/중요도 등에 따라 변경된다. 표현된 원의 중심을 클릭할 경우, 해당 이벤트들이 표현된 문장들이 하단에 제공된다.

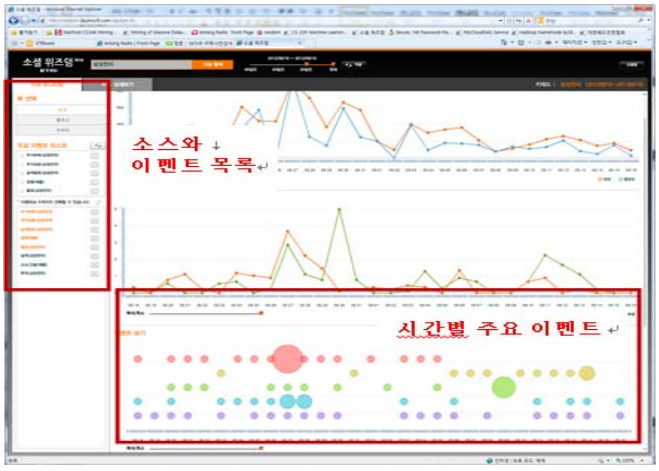

그림 3. 소셜위즈덤 화면

9. 결론 및 향후 연구

본 논문에서는 의사결정 지원 시스템을 위한 이벤트 추출 기술에 대해서 소개하였다. 기업과 공공분야에 필요한 31개의 이벤트 템플릿을 수작업으로 선정하여 정의하였고 템플릿 추출을 위해 고정밀 언어분석 기술과 온톨로지에 기반한 이벤트 제약 및 필터링 규칙을 적용하였다. 소스 별로 구분된 평가데이터를 대상으로 평가한 결과, 성능(F-Score)이 뉴스 0.544, 블로그 0.3386, 트위터 0.3251이고 전체 통합 성능은 0.4648이었다. 비문이 적은 뉴스의 성능이 우수하였다. 필터링의 성능(Accuracy)은 뉴스 0.7257, 블로그 0.6122, 트위터 0.6207이고 전체 통합 성능은 0.722이다. 필터링의 성능도 비문이 적은 뉴스에서 우수하였다. 재현율보다 정밀도가 중요한 질의응답이나 정보추출에서는 비교적 비문이 적은 뉴스나 백과사전 등을 대상으로 하는 것이 효율적이며, 다양한 제약을 통한 필터링이 효과적임을 알 수 있었다.

본 연구에서 구현된 이벤트 템플릿 추출 결과는 소셜미디어 상의 주요한 이슈, 이벤트, 감성정보 등을 제공하는 의사결정 지원 시스템인 소셜위즈덤 서비스에 포함되어 있다. 기존 시스템들의 경우, 주요 개체에 대한 빈도 및 감성 정보만을 제공하여 개체들 간의 의미적 연관성에 기반한 정보를 파악하기 쉽지 않은 단점이 있었다. 그러나 본 논문에서는 이벤트 템플릿에 기반하여 키워드와 연관된 이벤트를 순위화하여 제시함으로써, 시간대별 개체들 간의 연관성을 파악할 수 있어서 의사결정에 많은 도움을 제공할 것으로 판단된다.

향후 연구계획은 이벤트의 추가, 확장이 용이할 수 있도록, 이벤트 추출 기술에 대한 적응성(Adaptation) 향상을 위한 오픈 정보 추출(Open IE) 기술에 대한 연구를 진행할 예정이다. 또한, 정보추출에서 가장 중요한 두 가지 정보인 시간정보(Temporal Information)와 지역(Spatial Information)정보를 이용한 정보 시각화 기술에 대해서 연구할 예정이다. 특히, 시간정보를 기반으로 주요 개체에 대한 이벤트들을 타임라인 상에 연대기로 요약하여 제공하는 기술은 빅데이터에 기반한 정보추출 및 요약에서 중요한 기술이 될 것이다.

4) 2012.08.18. ~ 2012.09.16.

참고문헌

[1] 채승병, "정보호수 속에서 금맥 찾기: 빅 데이터(Big Data) 분석과 활용", SERI 경영노트 제 91호, 2011년 2월
[2] IBM Watson, http://www-03.ibm.com/innovation/us/waston/
[3] Recorded Future, http://www.recordedfuture.com
[4] Takeshi Sakaki, "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors", WWW 2010, 2010.
[5] Panagioris T. Metaxas, "How (Not) to Predict Elections", 2011 ieee third international conference on social computing, 2011
[6] C. Lee, "Named Entity Recognition with Structural SVMs and Pegasos algorithm". Korean Journal of Cognitive Science, Vol.21, No.4, 2010
[7] Heng Ji, Ralph Grishman and Hoa Trang Dang, "An Overview of the TAC2011 Knowledge Base Population Track", In Proceedings to the Third Text Analytics Conference(TAC2011), 2011