

# 구조화된 저널리즘 서비스를 위한 과학 칼럼 정보 지식화 프레임워크 설계\*

## Design of a Knowledge Framework for Structured Journalism Service based on Scientific Column Database

최 성 필 (Sung-Pil Choi)\*\*\*  
김 혜 선 (Hye-Sun Kim)\*\*\*\*  
김 지 영 (Ji-Young Kim)\*\*\*\*\*

### 목 차

- |                       |                     |
|-----------------------|---------------------|
| 1. 서 론                | 5. 과학향기 지식베이스 구축 방안 |
| 2. 연구의 목적 및 필요성       | 6. 차세대 과학향기 서비스 모델  |
| 3. 관련 연구 동향           | 7. 결론 및 향후 연구 방향    |
| 4. KISTI의 과학향기 콘텐트 분석 |                     |

### 초 록

본 논문에서는 과학 저널리즘의 대표적인 성공 사례로 평가되고 있는 KISTI의 과학향기 칼럼에 대한 반자동 지식화 방안과 이를 기반으로 과학기술 인포그래픽 기반의 서비스 체제를 제안한다. 전문가나 특정 분야 과학자를 대상으로 하는 전문 정보와는 달리 과학향기 콘텐트는 일반 대중을 대상으로 알기 쉬운 과학 상식을 제공하고 있다. 이러한 특성을 바탕으로 과학향기 데이터베이스를 지식화하기 위한 방법론 즉, 지식 유형, 지식 추출 방법 및 절차 등을 세부적으로 살펴본다. 또한 과학향기 지식베이스를 기반으로 과학기술 인포그래픽 서비스 체제를 새롭게 정의하고 이에 대한 세부적인 구성도, 방법론 및 특징 등을 기술한다. 이를 통해서 미래의 과학 저널리즘 서비스가 나아가야 할 발전적 방향을 제안한다.

### ABSTRACT

This paper proposes a noble service architecture based on scientific infographic as well as semi-automatic knowledge process for 'KISTI's Scent of Science' database, which has been highly credited as a science journalism service in Korea. Unlike other specialized scientific databases for domain experts and scientists, the database aims at providing comprehensible and intuitive information about various important scientific concepts which may seem not to be easily understandable to general public. In order to construct a knowledge-base from the database, we deeply analyze the traits of the database and then establish a semi-automatic approach to identify and extract various scientific intelligence from its contents. Furthermore, this paper defines a scientific infographic service platform based on the knowledge-base by offering its detailed structure, methods and characteristics, which shows a progressive future direction for science journalism service.

키워드: 과학저널리즘, 지식화, 인포그래픽, 구조화된 저널리즘, 지식서비스

Science Journalism, Knowledge Processing, Infographic, Structured Journalism, Knowledge Service

\* 본 논문은 경기대학교 교내과제(2014-005)로 수행된 “기술 용어의 서술적 표현 다양성 자동 인식 연구” 과제에서 일부 지원을 받아 수행된 연구의 성과물임.

\*\* 경기대학교 인문대학 문헌정보학과 조교수(spchoi@kgu.ac.kr) (제1저자)

\*\*\* 한국과학기술정보연구원 정보서비스실 실장(hskim@kisti.re.kr) (공동저자)

\*\*\*\* 한국과학기술정보연구원 정보서비스실 선임연구원(yes@kisti.re.kr) (교신저자)

논문접수일자: 2015년 1월 26일 최초심사일자: 2015년 2월 9일 게재확정일자: 2015년 2월 17일  
한국문현정보학회지, 49(1): 341-360, 2015. [<http://dx.doi.org/10.4275/KSLIS.2015.49.1.341>]

## 1. 서 론

### 1.1 정보 유통 체제의 혁신과 구조적 저널리즘

정보 기술의 급진적인 변혁은 고수준의 미디어 콘텐트(뉴스기사, 분석정보, 칼럼 정보 등)를 지속적으로 생산하여 공급해 왔던 국내외 언론기관 및 미디어 전문기관들에게 존립의 위기감과 동시에 새로운 도약의 기회를 시사하고 있다. 기존의 오프라인 유통체제와 종이기반 정보를 고수해 오던 많은 저널리즘 관련 기관 및 업체들은 이제 온라인 미디어 체제로의 혁신이라는 무거운 난제를 해결해야만 생존할 수 있다. 지난 5월 유출된 미국 최대 일간지 뉴욕타임즈(the New York Times)의 ‘혁신 보고서(Innovation)’는 디지털 시대에서의 전통적인 저널리즘 정보 유통 체계고수에 대한 심각한 자기반성과 아울러 단순 정보 생산자의 역할에서 벗어나 온라인 시대에서의 성공적인 혁신을 위한 네 가지 요구사항을 <표 1>과 같이 제시하고 있다(The New York Times 2014; 강정수 2014).

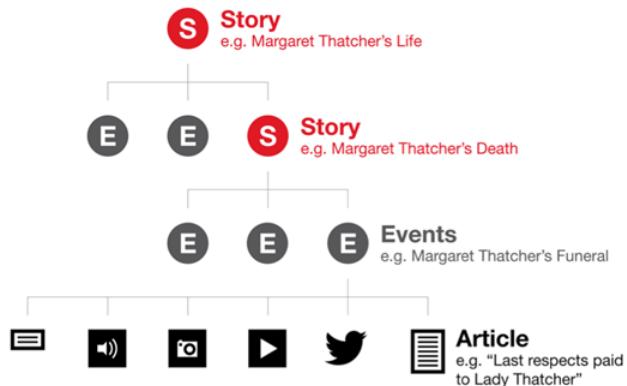
<표 1>에서 나타난 바와 같이 기존의 고루한

유통 체제를 고수하고 있는 언론 기관들이 Huffington Post(Huffington 2015), BuzzFeed(Peretti 2015), VoxMedia(Bankoff 2015) 등과 같은 신생 인터넷 저널리즘 유통 기업들의 도전에 대응하기 위해서는 (1) 새로운 기술의 신속한 적용을 통한 이용자의 편의성 증대 및 흥미 유도, (2) 능동적이고 적극적인 뉴스룸(News Room) 체제 구축을 위한 협력 및 융합, (3) 페이지 저널리즘 우선주의에서의 신속한 탈피를 통한 디지털 저널리즘으로의 체제 전환 등을 꼽고 있다.

본 논문은 <표 1>의 요구 사항 중에서 뉴스 도달거리 확장 방안 마련을 위한 구조화된 저널리즘(Structured Journalism) 체제 구축에 주목한다. 강정수(2014)에 따르면 구조화된 저널리즘이란 저널리즘 데이터 구축 과정에서 발생하는 정보 내부의 핵심적인 단위 요소를 분리하여 독립적인 객체로 인식한다. 결론적으로 단일 뉴스를 복수의 의미적 객체로 해체하는 과정을 통해서 해체된 객체들의 재조합을 통한 새로운 스토리라인(Storyline)을 구성할 수 있게끔 해준다. Tarling(2013)은 이러한 구조적 저널리즘을 쉽게 설명하기 위해서 <그림 1>과 같은 예를 제시한다.

<표 1> NYT 혁신보고서의 4가지 혁신 요구사항

혁신요구사항	내용
뉴스 도달거리 확장 방안 마련	- 온라인 뉴스기사 서비스에 적합한 구조화된 저널리즘 (Structured Journalism) 체제 구축 - 소셜 미디어 기반 서비스 체제 강화 - 뉴스 소비의 개인화 지원
편집국과 비즈니스 팀 사이의 협력	- 독자 및 청중 분석 데이터, 뉴스 유통 관련 통계 정보 등의 상호 공유
편집국에 전략팀 배치	- 내/외부 정보 수집, 기술 혁신 흐름 분석 등을 통한 새로운 프로젝트 제안 및 실행이 가능한 전략팀을 편집국 내에 배치하여 상시 운영
디지털 우선 전략 (Digital First) 수립	- 종이신문의 제약에서 벗어나 디지털 뉴스 생산에 가능한 모든 역량 투여 - 최종적으로 디지털 뉴스 중에서 가장 홀륭한 뉴스를 선별해 다음 날 (아침) 종이신문에 배치하는 체제 구축



〈그림 1〉 구조화된 저널리즘의 예(Tarling 2013)

하나의 스토리로서의 대처 수상의 인생(Margaret Thatcher's Life)은 많은 하위 스토리를 가지게 되며 그 중 그녀의 죽음(Margaret Thatcher's Death)에 대한 스토리는 다시 복수의 이벤트들로 구성될 수 있다. 이러한 이벤트들은 단일 뉴스기사 전체 혹은 일부로서 표현되며 구조적 저널리즘을 구현하기 위해서는 기사가 생성될 시점에서 표시(annotation)되어야 한다. 이런 형태로 뉴스를 구조화한 결과물을 Tarling (2013)은 객체 지향 뉴스(Object Oriented News)라고 정의하였으며 현재 영국의 세계적인 언론 기관인 BBC는 이러한 개념을 기반으로 “A Model for Story-telling in News”라는 연구 과제를 수행 중에 있다고 밝혔다.

기존의 나열식 뉴스 기사 서비스에 비해, 위에서 제시한 구조적 저널리즘 기반 서비스의 가장 큰 장점은 단일 기사 아이템 단위의 분절된 뉴스 전달이 아닌 맥락 중심(context centric)에서의 통합된 스토리 제공을 통해 이용자가 보다 쉽고 재미있게 온라인 뉴스 서비스를 즐길 수 있다는 점이다. 이용자는 현재 읽고 있는 뉴스기사를 이해하기 위해서 그 배경이나

관련지식을 추가적으로 검색하거나 파악할 필요가 없다. 특정 사건에 대한 연관 기사 혹은 칼럼들을 직관적이고 즉각적으로 확보할 수 있음으로 해서 마치 해당 사건을 잘 알고 있는 전문가 혹은 담당 기자가 이용자에게 직접 이야기를 해 주듯이 온라인 뉴스 서비스를 활용할 수 있다.

이러한 구조적 저널리즘을 현실화하기 위해서는 기반 체제나 기술들이 뒷받침 되어야 한다. 특히 기존 뉴스 기사 컨텐트 관리 시스템의 대대적인 개편이 필수적이다. 새롭게 작성되어 등록되는 뉴스 기사에 다양한 형태의 부가 정보 즉, 분류 정보, 장르 정보, 연관 기사 정보 등이 효율적으로 부착되어야 하며, 뉴스 기사의 내용에 존재하는 주요 용어, 문장, 단락 등에 추가적인 구조 정보 등이 추가될 수 있도록 시스템을 개편해야 한다. 특히 단일 뉴스 기사가 단절되고 독립된 하나의 아이템이 아니라 전체 뉴스 기사 데이터베이스 내에서 다른 기사들과 서로 긴밀하게 연동된 형태로 구성될 수 있도록 컨텐트 연계 체제 지원이 가능해야 한다.

### 1.2 사이언스 저널리즘(Science Journalism)

박종인(2012)은 사이언스 저널리즘을 자연 과학과 의학, 공학 기술 영역에서 발생하는 새로운 사실이나 새롭게 부각되는 사실 중 사회와 사회 구성원에게 중요하고 필요한 정보를 지속적으로 제공해 주는 것으로 정의하고 있으며 사이언스 저널리즘의 중요한 역할 중의 하나가 사회 및 사회 구성원을 대상으로 한 과학적 계몽 활동이라고 주장하고 있다. 이러한 노력의 일환으로 각 신문사들은 과학 기술 분야 셋션을 독립적으로 운영하고 있으며, 예전부터 “과학동아”라는 전문 잡지를 발간해 오던 동아일보의 경우에는 동아 사이언스(김두희 2015)라는 전문 과학기술 분야특화 정보서비스를 제공하고 있다. 특히 한국과학기술정보연구원(KISTI)에서 운영하고 있는 “KISTI의 과학향기”(KISTI 2015) 서비스는 생활 속 과학이야기를 모토로 내세우며 어렵고 복잡한 과학적 개념을 쉽고 재미있게 풀어 쓴 에세이 형식의 칼럼 기사, 과학 상식 등을 지속적으로 구축하여 제공하고 있다.

그럼에도 불구하고, 과학 영역은 일반 대중이 이해하기가 쉽지 않은 분야이며 이러한 난점을 극복하기 위해서는 다양한 형태의 정보 제공 방법론이 대두되어야 한다. 하지만 현재로선 이러한 정보 제공 방법을 과학기술 분야 뉴스를 작성하는 기자나 칼럼니스트에 온전히 의존하고 있는 상황이다. 따라서 정확하고 새로운 과학적 지식을 쉽게 전달하고 이러한 지식을 체계화함으로써 앞 절에서 언급한 구조적 저널리즘 기반의 정보 서비스 구축을 위한 방안이 절실히 요구되고 있다(박종인 2012).

본 논문의 구성은 다음과 같다. 우선 2장에서는 사이언스 저널리즘의 지식화 연구에 대한 필요성과 더불어 전체적인 연구 목적을 기술한다. 더불어 3장에서는 저널리즘 정보 지식화 및 차세대 서비스 모델 분야에 대한 사례 연구 결과를 보여주고, 4장에서는 현재까지 구축되어 서비스되고 있는 KISTI의 과학향기 콘텐트를 심층적으로 분석한 결과를 설명한다. 이어 5장에서는 본 논문에서 제안하는 지식베이스 기반 차세대 사이언스 저널리즘 서비스 모델을 도출한다. 추가적으로 6장은 5장에서 제안된 차세대 서비스 모델에 필요한 과학향기 지식베이스 구축 방안을 집중적으로 살펴보고 7장에서 결론 및 향후 연구 방향을 모색해 본다.

## 2. 연구의 목적 및 필요성

앞 절에서 지적하였듯이 사이언스 저널리즘은 과학의 대중화 및 사회 구성원들에 대한 과학적 계몽이라는 차원에서 그 중요성이 강조되어 왔으나, 일반 대중에게는 생소하고 어려운 과학 정보를 어떻게 하면 효과적으로 전달할 수 있을까에 대한 정보 서비스적인 측면에서의 연구는 거의 시도되지 않았다. 특히 텍스트로 표현된 뉴스 기사나 칼럼 등에 대한 체계적 분석을 통한 구조적 저널리즘 기반의 서비스 모델에 대한 연구는 거의 이루어지지 않았다.

이러한 상황에서, 본 논문은 과학기술분야 전문가 칼럼, 신문 기사 등을 바탕으로 이를 지식화 및 구조화함으로써 새로운 형태의 구조화된 저널리즘 기반 서비스 구현을 위한 통합 아키텍처를 제안한다. 이를 위해서 동아 사이언

스와 더불어 국내에서 많은 이용자를 확보한<sup>1)</sup> 사이언스 저널리즘 서비스인 KISTI의 과학향기 데이터베이스를 표본 선택하여 분석하였으며, 과학향기 칼럼 텍스트에 내재된 다양한 지식 요소들을 발굴하고 이를 활용하는 통합 모델을 제안한다. 이와 관련하여, 텍스트에 내재된 지식을 자동으로 추출하기 위해 현재까지 수행된 대표적인 연구 분야는 정보 추출(Information Extraction) 분야로, 특정 문장에 표현된 두 개체명(named entity)을 인식하고 이들 간의 관계를 추정함으로써, 텍스트 내에 존재하는 관계형 지식을 자동으로 추출하는 방법론을 주로 다룬다 (Choi et al. 2014). 본 논문에서는 이러한 관계형 지식뿐만 아니라 절차적 지식(Song et al. 2011), 정의형 지식(Kobyliński and Przepiórkowski 2008) 등을 과학향기 칼럼에서 추출할 수 있는 방안을 살펴본다. 추출된 다양한 유형의 지식들을 통합하고 구조화함으로써 위에서 언급한 구조적 저널리즘을 가장 효과적으로 실현할 수 있는 인포그래픽 기반 차세대 서비스 모델을 제안한다.

### 3. 관련 연구 동향

#### 3.1 저널리즘 정보 지식화 요소 기술

비정형 텍스트를 정형 지식으로 변환하는 기술 중에 대표적인 기술이 정보 추출(Information Extraction)이다. 정보 추출이란 텍스트의 내

용 전개에 핵심적인 역할을 수행하는 다양한 개체명(named-entity)들을 식별하여 추출하고 이들 간의 의미적 연관 관계를 식별하는 기술이다(Soderland 1999). 효과적인 정보 추출을 위해서는 용어 인식, 개체명 인식, 관계 추출, 문서 분류, 문서 요약 등의 세부적인 기술들이 통합적으로 활용되어야 한다. 현재 위의 세부 기술들은 많은 연구가 이루어지고 있으며 일부는 실용적인 수준에 도달한 상황이다. 하지만 위 기술들을 그대로 저널리즘 정보 지식화에 적용하기에는 아직 한계가 있으며, 체계적으로 구성된 워크플로우를 바탕으로 반자동 지식화 체계 구축(Semi-Automatic Knowledge Construction System)이 필요하다(Choi et al. 2012a).

#### 3.2 저널리즘 정보 서비스 기술

현재 뉴스 기사를 활용한 새로운 정보 서비스가 지속적으로 등장하고 있으며, 이들 각각은 나름대로의 독특한 특장점을 보유하고 있다.

우선 NewsMental<sup>2)</sup>은 신문 기사에서 핵심 어를 추출하고 자동 요약 서비스와 더불어 특정 기사에 대한 감성 및 영향력 분석을 커뮤니티 오피니언 제공 기법을 이용하여 제공하고 있다. FAST<sup>3)</sup>는 신문 기사에 대한 영향력 분석을 통해서 특정 기사에 대한 향후 페이지 뷰 횟수 예측 서비스를 제공한다. NHN은 유사 신문 기사를 통합해서 보여주는 뉴스 클러스터링 베타 서비스<sup>4)</sup>를 진행 중이고 이와 함께 뉴스

1) <<http://scent.ndsl.kr/sctApplicationSite.do>>

2) News Mental, <<http://www.newsmental.com/>>

3) FAST, <<http://fast.qcri.org/>>

라이브러리<sup>5)</sup> 서비스를 통해서 동아일보, 경향신문, 매일경제, 한겨례신문 등의 언론사가 보유한 창간 시점부터 현재까지의 모든 신문기사를 디지털화 하고 이에 대한 통합 검색 서비스를 제공하고 있다. (주)다음소프트는 소셜 매트릭스라는 서비스를 통해 기사 검색에 활용되는 특정 탐색어에 대한 소셜 미디어에서의 연관관계 분석, 탐색어에 대한 통계 추이 및 긍정/부정 분석 그리고 탐색어 여론 분석 서비스 등을 제공하고 있다.<sup>6)</sup>

### 3.3 연구 동향 시사점

대부분의 저널리즘 정보 서비스들은 지식화

요소 기술들의 불완전성으로 인해 검색, 군집화, 요약, 분류 등과 같은 비교적 제한된 형태의 분석 서비스를 제공하고 있다. 이러한 시점에서 최근 들어 등장한 인포그래픽(Infographic) 기반의 저널리즘 서비스를 예의 주시할 필요가 있다. 일반적으로 저널리즘을 위한 인포그래픽의 장점으로 흥미유발, 정보습득시간 경감, 기억 지속시간 연장 및 빠른 확산을 들고 있다.<sup>7)</sup> <그림 2>에서 보는 바와 같이, 이러한 인포그래픽의 종류에는 통계 기반, 타임라인 기반, 프로세스 기반, 위치 및 지리 기반, 구조 및 원리 설명 기반 인포그래픽 등이 있다.

현재 대부분의 언론사 및 뉴스미디어 관련 서비스 업체들은 텍스트로 작성된 뉴스 기사를



<그림 2> 다양한 유형의 인포그래픽 예시

4) News Clustering, <<http://lab.naver.com/newscluster/>>

5) News Library, <<http://newslibrary.naver.com/>>

6) <<http://insight.some.co.kr/campaign.html>>

7) 인포그래피이란?, <<http://v-vdesign.com>>

정리, 요약하여 이미지로 표현한 인포그래픽 기사를 지속적으로 구축하여 서비스하고 있다. 또한 웹 기반 혹은 클라이언트 기반으로 인포그래픽을 저작할 수 있는 다양한 저작도구들도 개발되어 제공되고 있다.<sup>8)</sup> 그러나 대부분의 인포그래픽 기사들이 텍스트의 내용을 수동으로 가시화하는 작업을 거쳐서 구축되고 있으며 이로 인해 기사 내용의 심층 분석적 가시화나 최신 기사들의 신속한 인포그래픽스화는 거의 이루어지지 않고 있다. 만일 앞에서 언급한 인포그래픽 생성 도구에 개체명 인식, 용어 추출, 통계 분석, 관계 추출, 분류, 요약 등의 지식화 지원 기술이 접목된다면 인포그래픽 뉴스 기사 구축 과정을 획기적으로 효율화할 수 있을 것이다.

KISTI의 과학향기 콘텐트의 특성 상 이용자의 편의와 관심도 유발 극대화를 위한 과학기술 인포그래픽 칼럼 구축 및 서비스는 가장 효과적인 차세대 과학 저널리즘 서비스 대안이 될 수 있을 것이다.

#### 4. KISTI의 과학향기 콘텐트 분석

이 장에서는 본 논문에서 연구 대상으로 채택한 KISTI의 과학향기 콘텐트에 대한 세부적인 분석 결과를 설명한다. 우선 2014년까지 구축된 KISTI의 과학향기 데이터베이스의 규모는 <표 2>와 같다.

<표 2> 과학향기 콘텐트 구축 현황(2014년)

구분	합계
칼럼	FOCUS 과학
	FUSION 과학
	FUNNY 과학
	소계
	1,535
	과학상식
	방송 동영상
	팟 캐스트 음성콘텐츠
	85

이 중에서 현재 서비스 중인 칼럼 중 최근에 구축된 칼럼 기사 약 100건을 중심으로 세부 내용을 분석하여 과학향기 콘텐트의 특징 및 장점과 함께 반자동 지식화 관점에서의 몇 가지 문제점 및 난점을 제시한다.

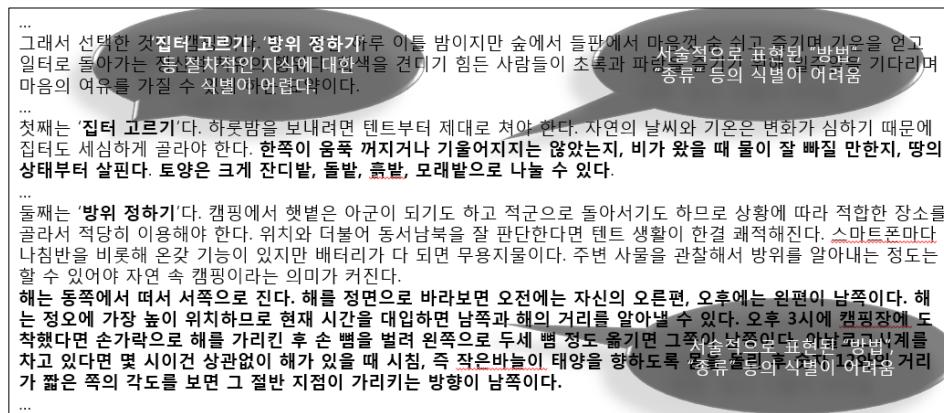
... 혈중 칼륨 농도가 높아지면서 심장 박동이 불규칙해지고 이는 돌연사의 주범인 부정맥의 원인이 된다. 칼슘과 나트륨이 피 속에 많아지면서 전해질 불균형으로 부종과 통증도 유발한다. 또 근육이 빠르게 근무력감과 근육통, 피로감도 느끼게 된다.

... 흰문근육해증은 장시간 강도 높은 운동을 하거나 평소 잘 사용하지 않는 근육을 자주 사용할 때 많이 발생하는 것으로 알려졌다. 이 외에도 음주나 간질, 약물 부작용 등이 원인으로 꼽힌다. 바이러스 질환도 원인이 된다. 지난 2009년 전국적으로 유행했던 '신종플루'에 감염돼 '타미플루'를 처방 받아 복용했던 여고생이 사망했는데 원인이 흰문근육해증이었다. 흰문근육해증의 정확한 진단은 혈중 크레아틴 키나아제(creatinine kinase, 근육 활동시에 ATP 생성 반응을 촉매하는 효소, 이하 CK) 수치로 확인한다. 근육 세포 속 물질인 CK의 정상 범위는 22~198 U/L인데 흰문근육해증인 경우 정상치의 10배에서 200배 이상 증가한다.

... 치료는 수액을 지속적으로 투여해 마이오플로비를 소변으로 배출시키는 것이 일반적이다. 일상 활동 정도가 높게 나올 때는 부정맥 예방을 위해 심전도 검사도 받아 보는 것이 좋다.

<그림 3> 과학향기 칼럼 내에서의 전문용어 활용 예시

8) <<http://trendw.kr/design/201405/12427.t1m>>



〈그림 4〉 과학향기 칼럼 내에서의 절차적 지식, 서술적 개념 표현 예시

#### 4.1 KISTI의 과학향기 콘텐트의 특징 및 장점

과학향기 콘텐트의 특징 및 장점은 크게 3가지로 구분되어 설명될 수 있다. 우선 서두에서도 언급하였듯이 이 데이터베이스는 과학기술 분야에 특화된 저널리즘 정보(Science Journalism Contents)를 담고 있다. 다시 말해서, 과학기술 분야의 다양한 이벤트(event)에 대한 직관적 보도, 논평해설 등을 중심으로 내용이 구성되어 있다. 여기서 이벤트란 단순히 “사건” 만을 의미하는 것이 아니라 “행사”, “발견”, “발명”, “인물의 행위”, “현상” 등을 아우르는 개념으로서 만일 과학기술 분야에 국한한다면 “신기술”, “발명”, “과학자”, “자연현상”, “실험” 등과 같은 일련의 동적인 수행의 과정 혹은 결과를 통칭할 수 있다(Choi et al. 2012b). 이러한 특징을 기반으로 앞에서 언급한 사이언스 저널리즘의 정의 및 역할에 충실하게 구성되어 사용자가 비교적 쉽게 이해하고 활용할 수 있는 과학 상식 서비스를 지향하고 있다. 두 번째로 이야기 전개 방식의 멀티미디어 자료라는 특성을 가진다. 다양한 매체(텍스트, 음성, 이미지,

동영상 등)를 활용하여 자연스러운 이야기 및 담화 수준의 콘텐트를 추구함으로써 사용자의 흥미를 유도하는 상호작용적인 서비스 구현이 가능하다. 마지막으로 소규모, 고품질의 콘텐트 구축 및 서비스를 추구하고 있다. 이를 위해, 과학 분야 칼럼에 전문성을 가진 검증된 소규모 칼럼니스트 그룹이 지속적으로 콘텐트를 공급하고 있다. 규모가 작다는 부분이 다양한 과학 분야를 다룰 수 없다는 점에서는 단점이 될 수도 있으나 그 만큼 양질의 콘텐트가 구축되고 있다는 측면에서는 시스템을 활용한 반자동 지식화에 이득이 될 수 있다.

#### 4.2 콘텐트 가공 및 분석의 어려움

2장에서 언급한 바와 같이 텍스트 내에는 다양한 유형의 지식이 내재되어 있으며 이를 식별하고 추출하기 위해서는 개별 유형의 지식이 실제로 글로서 표현되는 형태를 분석해야 한다. 이러한 관점에서 과학향기 콘텐트는 서로 상반되는 두 가지 특징을 가지고 있다. 우선 아무리 일반인을 대상으로 한 이해하기 쉬운 콘텐트를

표방한다고 하더라도 그 분야가 과학기술분야라면 아래 그림과 같이 콘텐트 내부에 다양한 전문용어가 활용될 수밖에 없다.

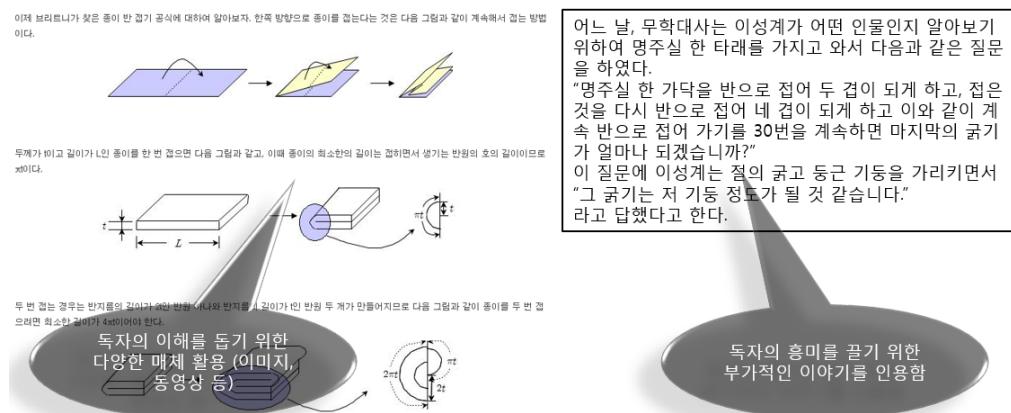
〈그림 3〉에서 보는 바와 같이 칼럼니스트 혹은 분야 전문가가 아무리 쉽게 기사를 작성하려고 해도 제한된 지면 내에서 직관적인 설명이 불가능한 부분이 있으며 이 경우에는 전문 용어를 적극 활용하고 있다. 또한 어려운 과학 지식에 대한 전문적인 묘사를 그대로 인용하고 있으며 (“~전해질 불균형으로 부종과 통증도 유발한다~”), 분야별로 논문 등과 같은 전문 정보에서 암묵적으로 자주 활용되는 표현 문구 (“~생성 반응을 촉매하는 효소~”)도 자주 등장한다.

이에 반하여, 〈그림 4〉와 같이 독자에게 어려운 과학지식을 쉽게 설명하기 위해서, 전문 용어의 사용을 자제하고 가급적 쉽고 길게 풀어서 기술하려는 특징도 보인다. 논문이나 특허 등과 같은 전문정보와는 다르게 설명의 용이함을 위해서 간접적이고 서술적인 문체를 활용하고 있으며, 이로 인해서 텍스트 내에서 핵

심적인 개념 요소(Information Nuggets, 전문 용어 등)를 식별하기가 쉽지 않다. 또한 현재까지 가장 연구가 활발히 진행되어 온 관계형 지식(Relational Knowledge) 추출 방법론으로는 추출하기가 쉽지 않은 절차적 지식이 상당수 내포되어 있다.

추가적으로 〈그림 5〉와 같이 독자의 이해에 도움을 주는 이미지, 동영상, 참고자료 등을 자동화된 콘텐트 가공 및 분석 프로세스 상에서 부작용을 초래할 수 있다.

그림에서 보는 바와 같이 스토리텔링(story-telling)에 있어서 이미지는 매우 중요한 역할을 하고 있으나, “~그림에서 보는 바와 같이~” 등과 같은 표현을 통해서 텍스트 설명을 생략하는 경향 때문에 핵심적인 정보를 텍스트 분석만을 통해서는 확보하기 힘든 상황이 초래된다. 또한 저자가 전달하고자 하는 핵심적인 내용과는 상관 없는, 단지 이해를 돋기 위한 이야기 등은 텍스트 분석에서 불필요한 자료로 분석될 가능성이 매우 높다(“무학대사”, “이성계” 등).



〈그림 5〉 과학향기 칼럼 내에서의 다양한 매체 및 참고자료 활용 예시

결론적으로 KISTI의 과학향기 콘텐트의 지식화 및 구조화 작업은 완전 자동화가 어려우며, 명확한 시나리오와 프로세스에 기반한 반자동 플랫폼이 적합하다. 앞에서도 잠시 언급 하였지만 현재까지 활용할 수 있는 수준의 지식 추출 모델은 관계형 지식 추출 모델들이 대부분이며 과학향기 콘텐트는 이보다 훨씬 다양한 형태의 지식을 포함하고 있기 때문이다. 이러한 유형의 지식은 지식 구축자가 지속적으로 식별하여 규칙화함으로써 향후 자동 추출의 기반을 다질 수 있다. 다음 장에서 과학향기 콘텐트에 대한 지식화 및 구조화 작업에 대한 상세한 내용을 다룬다.

## 5. 과학향기 지식베이스 구축 방안

이번 장에서는 과학향기 칼럼 정보에 존재하는 지식의 유형에 대한 분석과 함께 추출될 지식의 표현 방안 및 추출 방안에 대해서 세부적으로 살펴본다.

### 5.1 추출 대상 지식 유형

#### 5.1.1 관계형 지식

관계형 지식이란 텍스트 내에 존재하는 두 용어 혹은 개체명 간의 의미적 연관 관계를 식별하고 이를 트리플 형태(개체-관계-개체)로 추출된 지식으로서 현재까지 가장 일반적인 지식의 유형이다. 여기서 관계를 술어(predicate)라고 표현하고 술어의 대상이 되는 한 쌍의 개체를 논항(argument)이라고 표현하여 하나의 트리플을 “관계(논항, 논항)” 형태로 나타내기

도 한다(Choi et al. 2012b). 과학향기 내에는 다양한 전문용어, 인명, 지명 및 기관명 등이 존재하며 이들을 식별함과 동시에 이들 간의 관계까지도 식별하여 체계화시켜야 한다. 예를 들어, “자동차 사고를 막을 수 있는 기술은 자율 응급 제동 시스템(AEB)이다.”라는 문장에서 추출될 수 있는 관계형 지식은 “방지(자동차 사고, 자율 응급 제동 시스템)”와 “동일(자율 응급 제동 시스템, AEB)”이 추출될 수 있다. 또한 “볼보자동차가 이런 기술을 세계 최초로 적용한 ‘시티 세이프티’ 기능을 실제로 선보였다.”라는 문장에서는 “출시(볼보자동차, 시티 세이프티)”가 추출될 수 있다.

#### 5.1.2 절차적 지식

절차적 지식이란 특정 목표나 해결해야 할 과제를 풀어 나가기 위한 방법들을 순차적으로 나열한 지식을 의미한다. 예를 들어, KISTI 과학향기 기사 중의 하나인 “과학캠핑 방법”에 대한 칼럼 내에는 “(1) 집터 고르기 → (2) 방위 정하기 → (3) 불 피우기 → (4) 주변 관찰하기” 등과 같은 절차적 지식이 존재하고, “즉석 해시계 만드는 방법”을 설명한 칼럼에서는 “(1) 평평한 땅에 짤막한 나무 막대를 세우거나 꽂은 후 그림자의 끝부분에 표시 → (2) 10분 후 다시 표시를 한 후 두 점을 연결해서 직선을 그려서 동-서 축 표시 → (3) 동-서 축에 수직선을 그어서 남-북을 표시”와 같은 절차적 지식을 발견할 수 있다.

#### 5.1.3 정의형 지식

정의형 지식은 특정 용어나 개념에 대한 명확한 정의를 내린 문장이 포함하고 있는 지식이다. 예를 들어, “칼데라(caldera)란 화산 폭발

이 끝난 후 마그마가 빠져나와 생긴 지하의 빈 공간이 산 정상부의 무게를 이기지 못해 꺼져 생긴 넓은 함몰분화구를 말한다.” 혹은 “법률적으로는 인공조명의 부적절한 사용으로 인해 과도한 빛이 생기거나 정해진 영역 밖으로 누출되는 빛이 건강하고 폐적한 생활을 방해하는 상태를 빛공해로 규정한다.” 등과 같은 문장은 정의형 지식을 내포하고 있다.

#### 5.1.4 수치형 지식

수치형 지식은 각종 지표, 통계, 수치 등이 포함된 문헌에서 추출될 수 있는 정량화된 지식이다. 예를 들면 다음 문장들에서 수치형 지식을 추출할 수 있다.

- “영국 방위 산업체 키네틱 사의 무인기 ‘제페(Zephyr)’가 지난 2010년 7월 14일 21분 연속 비행한 기록을 가지고 있다.”
- “가장 앞선 기술력을 자랑하는 미국이 운용하는 군용 무인기 비율은 2005년 5%에서 2010년 41%로 크게 늘어났다.”
- “항공 우주 관련 연구 기관인 틸 그룹(Teal Group Corp.)은 2023년 전 세계 무인기 시장 규모가 890억 달러(약 93조 원) 수준으로 성장할 것이라는 전망을 내놨다.”
- “고고도(高高度, 지상 7,000에서 10,000m 쯤의 높이)”

#### 5.1.5 기타 유형 지식

그 외에도 인물의 신상 명세, 특정 군집의 종류(모기의 종류: 작은빨간집모기, 중국얼룩날개모기, 아에데스 알보펙투스), 사건의 진행 과

정(임진왜란 진행 과정 등)과 같은 지식들이 과학향기 칼럼 내에 존재한다.

### 5.2 지식베이스 구축 및 표현 방안

KISTI의 과학향기 콘텐트를 기반으로 한 구조화된 저널리즘 서비스를 실현하기 위해서는 체계적이고 유용한 지식베이스 구축 전략이 필요하다. 이를 위해서는 목표지향적인 지식 표현 방법론과 반자동 지식 구축 플랫폼 개발이 가장 핵심적이다. 본 절에서는 이 두 가지 방법론에 대해서 세부적으로 설명한다.

#### 5.2.1 지식 표현 방안

지금까지 인공지능, 시맨틱 웹, 언어 공학 등의 학문 분야에서 다양한 지식 표현 방법론이 제안되어 왔다. 전통적으로 지식 표현 방법론 연구는 주로 인공지능 분야에서 이루어져 왔으나 상당한 수준의 연구 성과에도 불구하고 실제 사용자가 그 효과를 체감할 수 있는 지식 표현 방법론 및 그에 따른 지식 서비스 모델은 아직까지도 요원한 상태다. 특히 시맨틱 웹 분야 연구의 활성화는 다양한 형태의 지식 표현 메커니즘(RDF, DAML, OWL 등)을 양산하였으나 실제 이러한 표준들을 바탕으로 실용 지식을 구축하고 활용하는 수준은 아직까지도 초기 단계를 벗어나지 못하고 있는 실정이다. 그 중 RDF(Resource Description Framework)<sup>9)</sup> 정도가 어느 정도 활용되고 있는 상황이며 이마저도 심층적인 지식을 표현하는 다양한 기능을 갖추었음에도 불구하고 지식 구축의 비효율

9) <<http://www.w3.org/RDF/>>

성으로 인해 제한된 정형 데이터에 국한되어 활용되고 있다.

이러한 상황에서 과학향기 콘텐트를 활용한 구조화된 저널리즘 서비스 및 6장에서 다룬 과학기술 인포그래픽 서비스를 위한 지식 표현 방법은 목표 지향적이어야 한다. 앞에서 언급한 대부분의 지식 표현 모델은 신규 지식 발견이나 추론 등과 같은 기능을 전제로 하고 있으므로 이러한 고수준의 지식 처리가 필수적이지 않은 상황에서 굳이 복잡한 지식 표현 메커니즘을 채택할 필요가 없다. 다시 말해서, 구현하고자 하는 목표 서비스에 최적화된 단순한 형태의 지식 표현 방법을 고안하는 것이 가장 중요하다. 최종적으로는 앞에서 설명한 지식 유형을 저장하고 표현하기 위해서 단지 관계형 데이터베이스를 활용한 튜플(tuple) 형태의 테이블 기반 저장 방법을 채택하면 된다. 관계형 지식은 일반적으로 RDF로의 표현이 일반적이었으나, 관계형 데이터베이스로 직접 저장하여 관리가 가능하다. 절차적 지식도 마찬가지로 관계형 DB를 사용하여 저장하되 구성 세부 단계에 대한 필터링 및 정규화 과정을 통해서 가급적 명확하게 식별될 수 있는 정보가 저장되어야 한다. 정의형 지식은 특정 용어 혹은 개념과 그에 대한 정의문을 그대로 저장하여 활용할 수 있으며, 수치형 지식은 우선적으로 처리할 수 있는 정량 지식의 형태를 정의하여 이를 테이블로 구성하고 향후에 지속적으로 이를 확장해 나가면 된다.

위에서 기술한 지식 유형 외에도 다양한 형태의 지식이 존재할 수 있으므로 이를 직관적이고 수월하게 저장, 관리할 수 있는 통합 스키마를 설계해야 한다. 단, 이 과정에서 관습적인

관점에서의 지식 표현이라는 학술적 개념에 너무 고착되어 비실용적인 스키마 및 표현 구조가 도출되지 않도록 주의해야 한다.

### 5.2.2 지식 구축 방안

앞에서 범용적인 측면에서 과학향기 콘텐트에 존재하는 다양한 지식 유형을 살펴보았으나, 본 논문에서 구축하는 지식은 완벽하게 대상 서비스에 의존적이어야 한다. 이는 현재의 기술적 한계점으로 인해서 완전 자동화된 지식 추출이 불가능한 것과 목표 중심적이지 않은 범용 지식 베이스 구축에 대한 시도가 대부분 실패로 돌아갔다는 사실에 기인한다. 본 논문의 연구 목표가 구조화된 저널리즘 서비스를 위한 과학기술 인포그래픽 기반 서비스 체계 구축이므로 이를 위한 지식 구축 방안이 마련되어야 한다.

우선적으로, 초기 단계에서는 과학향기 내의 콘텐트를 수작업으로 분석하면서 앞에서 언급한 유형들의 지식을 파악하고 저장함으로써 인포그래픽을 직접 구축하는 과정이 필요하다. 이를 위해서는 추출 대상 지식의 식별과 분류 작업이 수반되어야 하며 분석 결과에 대한 지식 표현 스키마를 단계적이고 반복적으로 수정, 확장해 나가는 과정이 필요하다. 이런 과정을 거쳐서 구축되는 인포그래픽 기사는 기존 대부분의 언론사에서 구축하고 있는 인포그래픽과는 달리 지식(데이터) 부분과 가시화(그림) 부분이 내부적으로 완벽하게 분리되어 있다. 이는 <그림 6>에서 명확하게 표현되어 있다.

또한 완전 자동화를 기반으로 구축된 지식베이스는 품질 저하 문제가 발생하므로 중장기적인 전략을 수립하고 수동 구축에서 출발하여 점차 자동화 비율을 늘려 나가야 한다. 초기에

는 견고한 수동 지식화 프로세스를 수립하고 구동함으로써 양질의 지식베이스 생산과 동시에 이를 통해 고수준의 과학지식 구축 인력 양성을 도모해야 한다. 이를 위해서 본 논문에서는 지식 구축인력이 손쉽게 지식을 구축할 수 있도록 지원하는 통합 지식 저작도구를 제안한다. <그림 7>은 이 도구의 전체적인 구성도를 보여주고 있다.

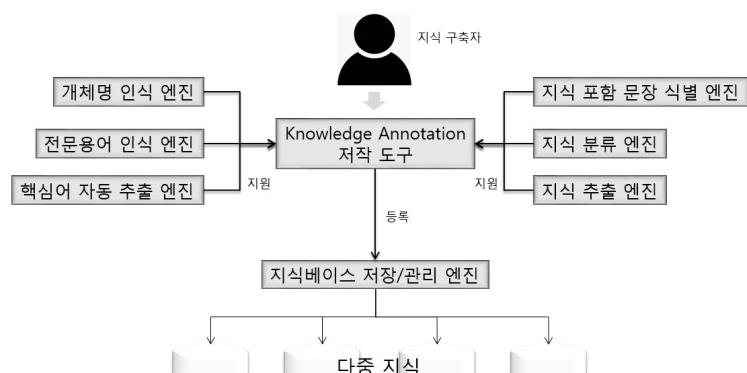
이 저작도구는 지식 구축자의 편의성을 극대

화하기 위한 사용자 인터페이스를 갖추어야 하며 <표 3>과 같이 총 6 가지의 기반 엔진이 탑재되어 구동된다.

위의 엔진들은 지식이 실제로 저작되는 과정에서 실시간으로 수행됨으로써 지식 구축자가 신규 지식을 빨리하고 처리하는데 지속적인 도움을 준다. 최종적으로 추출된 지식은 검증을 통해 지식베이스 저장/관리 모듈로 입력되며 저장된다.



<그림 6> 지식베이스 영역과 가시화 영역의 분리



<그림 7> 통합 지식 저작 도구 구성도

〈표 3〉 지식저작도구에 포함되는 기반 엔진 종류

기반 엔진	기능 설명
개체명 인식엔진	텍스트 내에 존재하는 인명, 지명, 기관명을 자동으로 식별
전문용어 인식엔진	과학기술 분야 텍스트에 자주 출현하는 기술용어를 식별
핵심어 추출엔진	전체 기사의 내용을 대표하거나 중요도가 높은 핵심 단어 및 구절을 자동으로 식별
지식포함문장 식별엔진	텍스트를 구성하는 개별 문장의 중요도, 패턴, 정보량 등을 전체적으로 계산하여 해당 지식이 포함되어 있을 가능성이 높은 문장만을 선별
지식분류엔진	지식 포함 문장 식별 엔진에서 식별된 문장이 포함하고 있는 지식의 종류를 판단 (관계형, 정의형, 통계적, 절차적 지식 등)
지식추출엔진	지식의 종류별로 해당 지식의 실제 값 혹은 인스턴스를 추출하여 구조화



〈그림 8〉 과학향기 지식베이스 구축 프로세스

〈그림 8〉은 본 논문에서 제안하는 지식 구축 프로세스를 나타내고 있다. 총 9 단계로 구성된 이 프로세스는 초기 버전으로서 향후 실질적인 구축 작업이 진행됨에 따라 다양한 형태로 변형/확장될 수 있다. 우선 지식화 대상 과학향기 칼럼 선택 과정에서는 지식 저작 도구에서 지식화 대상 문서를 선택하여 화면에 내용을 표시한다. 이 때 저작 도구는 개별 처리 문헌에 대한 버전 관리 기능을 제공하여 구축자가 자신의 문헌을 시간과 장소에 구애받지 않고 처리할 수 있다. 그 다음으로 칼럼 문헌 내에 존재하는 전문용어, 핵심어, 개체명 등을 반자동으

로 식별하는 과정을 거친다. 이 때, 〈표 3〉에서 설명한 기반 엔진들이 백그라운드 형태로 동작하여 우선적으로 자동 식별된 개체들을 화면에 표시하게 되고 그 결과를 구축자가 검토함으로써 오류를 수정하거나 누락된 부분을 후처리 정제하게 된다. 이 과정에서 저작 도구는 구축자가 올바르게 정제한 결과를 내부 통합 사전에 저장/관리함으로써 향후 동일한 용어 혹은 개체명에 대해서는 사전탐색 만을 통해서 표지(annotation)가 될 수 있도록 한다. 또한 저장된 개체 및 해당 문헌은 기반 엔진의 성능 향상을 위한 학습 집합으로 활용이 가능하다. 위와

같이 반자동 처리된 문헌을 바탕으로 구축자는 지식화 대상 핵심 문장(관계형, 절차적, 정의형, 수치적 지식 포함 문장)을 선별하게 된다. 이때도 <표 3>의 지식포함문장 식별엔진이 동작하게 되며 구축자는 선별 결과를 후처리 정제하는 작업만을 수행하면 된다. 문장이 선별되면 현재 문장이 포함하고 있는 지식의 종류를 분류하게 되는데 이러한 과정도 지식분류엔진으로 1차 수행되고 구축자가 후처리 과정을 통해서 정제하게 된다. 최종적으로 식별되고 정제된 지식은 그 유형에 따라 최종 검증 및 등록 과정을 거쳐서 지식화된다.

### 5.2.3 지식 구축자

지식 구축자의 역할은 통합 지식 저작 도구를 활용하여 실제로 과학향기 칼럼에서 지식을 추출하고 저장/관리하는 업무와 구축 완료된 지식에 대한 품질 검증 업무로 나뉘어 있다. 이러한 지식 구축자는 비록 단순 데이터 입력자의 업무보다는 높은 수준의 업무를 수행할 수 있어야 한다. 그러나 초기 단계에서의 기본적인 소양 및 자질은 (1) 데이터베이스 및 스키마 개념에 대한 일반적인 이해능력 보유, (2) 국어 국문학 및 문헌정보학 관련 전공, (3) 언어 및 콘텐트 처리 감각 보유, (4) 구축 도구를 활용하여 데이터베이스를 실제 구축하고 배포 경험 소유 등으로 크게 높지 않다.

업무 초기에는 전문가의 지도 및 감수 하에 기본적인 교육이 필요하며 칼럼 내에서 핵심적인 문장 및 문구를 발견하고 이를 구조화하는 업무에 대한 노하우 및 경험을 점차적으로 쌓게 된다. 많은 부분이 자동화된 기반 엔진을 통해서 수행되므로 구축자는 수행 결과에 대한

검증 및 누락 부분에 대한 보강에 집중화된 업무를 수행하게 된다.

위와 같은 능력과 경험을 가진 지식 공학 인력을 지속적으로 교육하고 확보한다면 향후 과학향기 칼럼 외의 다양한 콘텐트에 대한 지식화에 활용될 수 있다. 또한 지식 공학자 육성 관점에서 지금까지 자동화 소프트웨어 기술에 과도하게 집중된 상황을 탈피하여 진정한 콘텐트 전문가 혹은 데이터 전문가를 양성하여 새로운 일자리 창출 기회까지 엿볼 수 있다. 추가적으로 앞에서 언급한 통합 지식 구축 도구의 자동화 성능 개선을 위한 다양한 기반 자원 창출 및 피드백 수렴 창구의 역할까지도 할 수 있다.

## 6. 차세대 과학향기 서비스 모델

비록 5장에서 설명한 지식베이스 구축 결과물을 바탕으로 다양한 형태의 서비스 모형을 생각해 볼 수 있으나, 차세대 과학향기 서비스 모델 발굴에 있어서 본 논문에서 세운 기반 전제 조건은 (1) 과학향기 콘텐트의 재미, 특성, 흥미 및 이용자 친화성을 유지해야 하고 (2) 현재 과학향기를 활용하고 있는 주요 이용자 층에 대한 특성을 파악해야 하며, (3) 과학기술 분석 서비스를 단순하게 모방하여, 무의미하고 복잡한 도표 및 그래프의 남용 및 사용자 인터페이스의 과도한 사용 등은 기존 사용자들의 이탈을 초래할 수 있다는 것이다.

이를 바탕으로 본 논문에서 제안하는 새로운 형태의 지식베이스 기반 서비스는 “과학기술 인포그래픽(Scientific Infographic) 기반 과학향기 서비스”이다. 과학기술 인포그래픽이란 서두

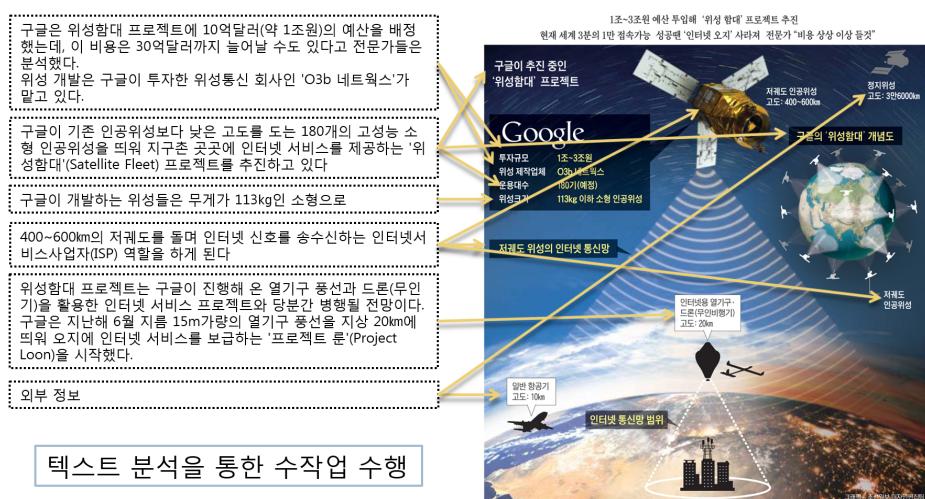
에서 언급한 것처럼 새로운 가시화 기반의 저널리즘 정보로서 본 논문에서는 과학향기 기사 각각을 분석하여 이를 차트, 지도, 다이어그램, 그리고 일러스트레이션 등을 활용하여 한눈에 파악할 수 있도록 구성된 서비스 기반 콘텐트라 정의한다. 과학향기 칼럼은 특정 분야기술, 트랜드, 과학 개념 등을 총체적으로 설명하고 있고 다양한 유형의 지식 요소를 포함하고 있으므로 이를 가시화하여 직관적으로 보여준다면 서비스의 만족도가 높아질 것으로 기대한다. 이는 조선일보나 동아일보 등과 같은 기존 언론기관들이 자신들의 저널리즘 정보를 지속적으로 인포그래픽화 함으로써 서비스 만족도를 제고하고 있는 사실에 기인한다.<sup>10)</sup>

## 6.1 과학기술 인포그래픽 구축 방안

저널리즘 정보를 바탕으로 한 인포그래픽 구

축 방법은 현재까지도 수작업 기반의 단순 그 래픽 작업을 벗어나지 못하고 있는 실정이다. <그림 9>에서 보는 바와 같이, 전체 기사에서 핵심적인 내용만을 발췌하여 이를 정리하고 각각을 이미지에 추가함으로써 대부분의 인포그래픽 기사가 구축되고 있다. 그러나 인포그래픽이 포함하는 내용이 기사나 칼럼의 핵심적인 지식을 가시화한 것이라면 이는 앞에서 설명한 지식화를 통한 가시화 방법을 적용할 수 있는 여지가 있다.

과학향기 콘텐트를 이용한 인포그래픽 구축 방안을 마련함에 있어서 5장에서 설명한 지식 베이스 구축 결과는 핵심적인 역할을 수행하게 된다. 우선적으로 과학향기 내의 콘텐트를 직접 수작업으로 분석하면서 인포그래픽을 생성하는 과정이 초기 단계에서 필요하다. 이는 4장의 지식 구축 프로세스와 연동되어 수행된다. 5장에서는 지식베이스 구축 중심으로 설명하였



〈그림 9〉 수작업을 통한 인포그래픽 구축 과정 (조선일보)

10) 조선일보 인포그래픽, <http://inside.chosun.com/>

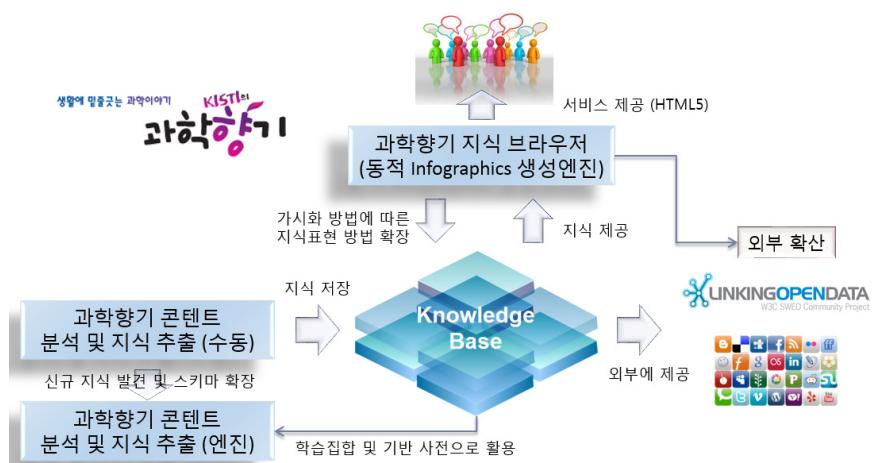
으나 최종적으로 과학기술 인포그래픽 서비스를 구축하기 위해서는 지식베이스 구축과 함께 인포그래픽 생성이 동시에 진행되어야 한다. 최종적으로 추출된 지식이 원하는 형태로 표현되기 위해서는 앞의 <그림 6>에서 보인 바와 같이 지식베이스 계층과 가시화 층이 완벽하게 분리되어야 하며, 다양한 인포그래픽 템플릿이 활용되어 추출된 지식의 표현 다양성을 확보해야 한다.

## 6.2 지식기반 인포그래픽 서비스 아키텍처

<그림 10>은 5장에서 제안한 통합 지식 저작 도구와 결합된 지식베이스 기반 인포그래픽 서비스 플랫폼 아키텍처를 보여주고 있다.

우선 중간 부분에 있는 지식베이스는 5장에서 구축되는 다양한 유형의 지식이 저장되는 저장소를 나타낸다. 앞에서도 언급하였듯이 전통적인 지식 표현 방법에서 탈피하여 인포그래픽 서비스 중심의 가시화 요소에 특화되고 단

순한 지식 표현 기법이 활용되어야 한다. 과학 향기 지식 브라우저는 동적 인포그래픽 생성 엔진이라고도 불리며 지식베이스에 저장된 지식을 인포그래픽 형태로 표현하는 기능을 수행한다. 초기에는 미리 구축한 이미지 템플릿에 데이터 혹은 수치 등을 삽입하는 형태로 쉽게 개발될 수 있으며, 향후에는 레이아웃, 색상, 아이콘, 바이럴 팩터(viral factor)들의 결합 등을 지원하는 과학기술 인포그래픽 저작 플랫폼으로 발전될 수 있다. 그림의 좌측에서 보는 바와 같이, 수동으로 추출되는 지식과 자동으로 추출되는 지식이 지속적으로 지식베이스에 축적됨으로써 동적으로 생성되는 인포그래픽의 핵심 지식 자원 공급원 역할을 수행한다. 구축된 과학향기 지식은 인포그래픽화된 형태의 패키지화된 지식 혹은 지식베이스 스키마에 저장된 형태의 원본 지식 형태로 외부에 확산될 수 있으며 다양한 미디어를 통해서 배포되고 유통될 수 있다. 앞에서도 강조하였듯이 본 시스템에서 구축되는 지식베이스는 주로 서비스를 위한



<그림 10> 과학향기 콘텐트를 활용한 지식베이스 기반 과학기술 인포그래픽 서비스 플랫폼

기반 자원으로 활용이 되지만 부가적으로 자동 지식 추출 엔진들의 성능 자가 증강에도 활용 될 수 있다. 이런 이유로 지식 구축 초기에 필요 한 수작업 비중이 시스템이 지속적으로 운영되고 자원이 확장됨에 따라 획기적으로 낮아지는 효과를 볼 수 있다(Riloff and Jones 1999).

위와 같이 구성되는 지식베이스 기반 과학기술 인포그래픽 서비스의 특징은 다음과 같다.

- 콘텐트가 지식으로 변환되어 저장된 지식 베이스를 활용하므로 고품질의 서비스가 신속하게 제공될 수 있다.
- 인포그래픽 템플릿 및 가시화 방법론의 확장을 통해서 다양한 형태의 서비스 레이아웃 실현이 가능하다.
- 개별 기사에 해당하는 인포그래픽은 서로 융합이 가능하여 새로운 가시화 아이템의 동적 생성이 가능하다(타임라인 기반 인포그래픽 + 위치 및 지리 기반 인포그래픽 = 시간에 따른 위치 및 지리 정보의 변화를 나타내는 인포그래픽).

위의 특징들은 모두 정보 및 지식 영역과 가시화 및 서비스 영역이 지식베이스 구축 단계에서부터 완벽하게 독립적으로 구축/운영되는 특성과 지식베이스가 서비스 관점에서 단순하고 직관적으로 표현될 수 있기 때문에 가능하다.

## 7. 결론 및 향후 연구 방향

본 논문에서는 KISTI의 과학향기 칼럼을 지

식화하고 이를 기반으로 새로운 형태의 서비스인 과학기술 인포그래픽 기반의 서비스 체제를 제안하고 이를 세부적으로 설명하였다. 전문가나 특정 분야 과학자를 대상으로 하는 전문 정보와는 달리 과학향기 콘텐트는 일반 대중을 대상으로 하는 과학 저널리즘 범주에 속하며 이를 지식화하기 위한 방법론 즉, 지식 유형, 지식 추출 방법 및 절차 등을 세부적으로 살펴보았다. 또한 현재 그 중요성이 지속적으로 부각되고 있는 인포그래픽 기반 저널리즘 체제를 과학기술 분야 칼럼 정보에 적용함으로써 얻을 수 있는 다양한 긍정적 효과에 대해서도 기술하였다.

본 연구 결과는 향후 연구로 수행될 실질적인 시스템 개발에 있어서 핵심적인 가이드라인 역할을 수행할 수 있다. 또한 지금까지 시도가 거의 전무했던 과학 저널리즘 정보의 지식화 모델을 새롭게 제시했다는 것에 의미를 둘 수 있다. 현재까지 수작업으로만 시도되었던 인포그래픽 저작 방식에 새로운 변화를 줄 수 있는 반자동화 모형을 제안한 것도 주요 연구 성과로 볼 수 있다.

향후 연구로서, 제안된 지식 유형, 지식화 절차, 서비스 방법론, 시스템 구성도 등을 기반으로 실제 개발 및 구현에 차수함으로써 자가 증강형 과학 저널리즘 지식화 및 인포그래픽 기반 서비스를 실현해야 한다. 또한 과학향기 콘텐트에서 도출될 수 있는 추가적인 유형의 지식들도 지속적으로 발굴함으로써 구축되는 지식베이스의 심도 및 수준을 향상시킬 수 있는 방안을 마련해야 한다.

## 참 고 문 헌

- [1] 강정수. 2014. “뉴욕타임스 혁신보고서의 교훈: 멋지게 실패하자!” 『슬로우 뉴스』.
- [2] 김두희. 2015. 동아사이언스. [online] <<http://www.dongascience.com/>>
- [3] 박종인. 2012. 『사이언스 저널리즘』. Communication Books.
- [4] KISTI, 2015. KISTI의 과학향기. [online] <<http://scent.ndsl.kr>>
- [5] Bankoff, J. 2015. VOX MEDIA. [online] <<http://www.voxmedia.com/>>
- [6] Choi, S. P., Chun, H. W., Jeong, C. H., Song, S. K. and Jung, H. 2012. “SINDI-WALKS: Workbench for PLOT-based Technological Information Extraction and Management.” *Green Computing and Communications (GreenCom)*.
- [7] Choi, S. P., Song, S. K., Jung, Geierhos, H., M. and Myaeng, S. H. 2012. “Scientific literature retrieval based on terminological paraphrases using predicate argument tuple.” *SoftTech 2012*.
- [8] Choi, S. P., Lee, S., Jung, H. and Song, S. 2014. “An intensive case study on kernel-based relation extraction.” *Multimed. Tools Appl.*, 71(2): 741-767.
- [9] Huffington, A. 2015. The Huffington Post. [online] <<http://www.huffingtonpost.com/>>
- [10] Kobyliński, Ł. and Przepiórkowski, A. 2008. “Definition Extraction with Balanced Random Forests.” *Advances in Natural Language Processing 2008*, 237-247.
- [11] Peretti, J. 2015. BuzzFeed. [online] <<http://www.buzzfeed.com/>>
- [12] Riloff, E. and Jones, R. 1999. “Learning dictionaries for information extraction by multi-level bootstrapping.” AAAI/IAAI, pp. 474-479. [1] The New York Times. 2014. *NYT Innovation Report 2014*.
- [13] Soderland, S. 1999. “Learning information extraction rules for semi-structured and free text.” *Machine Learning*, 34(1-3).
- [14] Song, S., Oh, H., Myaeng, S. H., Choi, S. P., Chun, H., Choi, Y. and Jeong, C. 2011. “Procedural Knowledge Extraction on MEDLINE Abstracts.” *Active Media Technology 2011*, 345-354.
- [15] Tarling, J. 2013. “Storylines vs object oriented news.” Top Drawer Sausage. [online] <<http://topdrawersausage.net/2013/07/07/storylines-vs-object-oriented-news/>>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- [1] Kang, J. S. 2014. “Lesson Learned from New York Times’ Innovation Report.” *Slow News*,

- [2] Kim, D. H. 2015. Dong-A Science. [online] <<http://www.dongascience.com/>>
- [3] Park, J. I. 2012. *Science Journalism*. Communication Books.
- [4] KISTI, 2015. KISTI's Scent of Science. [online] <<http://scent.ndsl.kr>>