

육하원칙 온톨로지 기반의 뉴스 필터링 방법

이석훈⁰¹, 이종현¹, 김장원¹, 정동원², 백두권^{†1}

고려대학교 컴퓨터·전파통신공학과¹, 군산대학교 정보통계학과²

{leha82, momoline, ikaros1223, baikdk}@korea.ac.kr¹, djeong@kunsan.ac.kr²

A News Filtering Method based on 5W1H Ontology

Sukhoon Lee⁰¹, Chonghyeon Lee¹, Jangwon Kim¹, Dongwon Jeong², Doo-Kwon Baik^{†1}

Department of Computer and Radio Communications Engineering, Korea University, Seoul¹

Department of Informatics and Statistics, Kunsan National University, Gunsan²

요 약

인터넷의 영역이 확대됨에 따라 인터넷에서 다양하고 많은 정보를 제공받게 되었다. 뉴스의 영역도 그 중 하나로 신문사, 방송사 등의 많은 언론사들이 인터넷으로 서비스를 확대함에 따라 뉴스 정보의 과다 현상이 일어나게 되었다. 이 때문에 사용자는 방대한 뉴스들 중에서 원하는 뉴스만 걸러서 보기를 원하게 되었고 이를 위한 뉴스 필터링 방법이 연구되었다. 뉴스 필터링 기술들은 주로 사용자의 관심 사항을 예측하여 제공해 주는 뉴스 추천 시스템을 위한 기술 개발에 초점을 두었다. 그러나 기존의 뉴스 필터링 기술들은 사용자의 관심 있어할 만한 뉴스를 추천할 뿐, 관심 없는 뉴스를 제외시키지는 못한다. 예를 들어, 어떤 특종 사건이 생기면 이 사건을 보도하기 위한 뉴스들이 각 언론사 마다 생성되고, 뉴스 추천 서비스를 사용하는 사용자는 기존의 뉴스 필터링 방법에 의해 사용자가 관심 있다고 예측되는 이 사건에 대한 뉴스를 제공받게 된다. 그러나 사용자가 이미 추천된 뉴스 중 하나의 뉴스 혹은 그와 동일한 내용의 다른 언론사에서 제공되는 뉴스를 읽었다면 추천된 뉴스는 이미 알고 있는 정보이므로 사용자는 이 뉴스에 대하여 관심이 없을 것이다. 기존의 뉴스 필터링 방법은 추천 시 중복된 뉴스를 제거하지 못하는 문제점을 지닌다. 이 논문은 이러한 문제점을 해결하기 위해 육하원칙 기반의 필터링 방법을 제시하고, 실험을 통해 이 논문이 제시한 방법의 장단점을 보인다.

1. 서 론

인터넷 시대가 도래하고 방대한 양의 정보가 생성됨 (Information Overload)에 따라 사용자가 원하는 정보를 찾기 위한 시간의 소비가 늘어나게 되었다. 때문에 사용자가 원하는 정보를 가공하고 걸러주기 위한 정보 필터링이 필요하게 되었다[1]. 한국에는 2009년 3월 집계로 1,399개의 인터넷 신문이 존재하며, 주간 기사 게재 건수는 평균 367.3 건인 것으로 조사되었다[2]. 이 자료는 뉴스의 영역에서도 필터링이 필요하다는 것을 보이며, 이를 위하여 정보 필터링을 뉴스의 영역으로 가져와 사용자가 원하는 뉴스만을 제공하기 위하여 뉴스 필터링 기술이 나타나게 되었다[3]. 뉴스 필터링 기능은 사용자가 원하는 정보만을 찾아내기 위한 Collaborative Filtering, Ontology-Content-based Filtering등과 같은

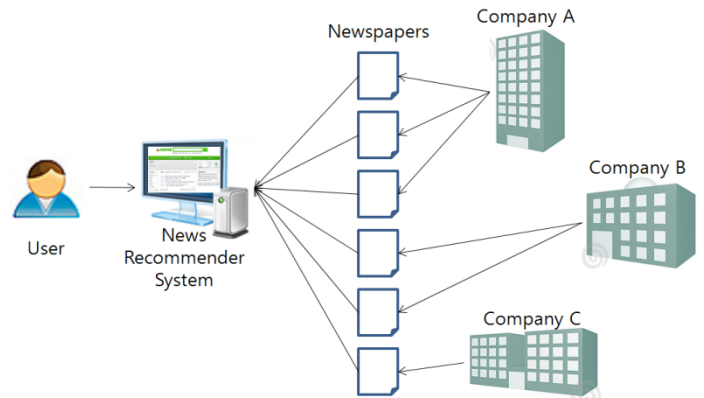


그림 1 뉴스 추천 시스템의 뉴스 제공방식

기술들이 연구되고 있다. 그 예로 구글 뉴스 같은 사이트에서는 뉴스 필터링 기법을 적용시킨 추천 리스트를 제공함으로써 사용자에게 좀 더 사용자의 기호에 맞는 뉴스를 제공해 준다[4].

그림 1은 일반적인 뉴스 추천 시스템의 뉴스 제공 방식을 보여 준다. 여러 언론사는 뉴스 추천 시스템에 각각 뉴스를 제공하고, 뉴스 추천 시스템은 그 뉴스들을 모아 사용자에게 적절한 필터링을 통해 제공해준다. 예

이 연구에 참여한 연구자는 '2단계 BK21 사업'의 지원 받았으며 이 논문은 2009년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(2009-0076863)을 밝힙니다.

† 교신저자 (corresponding author)

를 들어, 특종 사건이 발생했다고 한다면 대부분 언론사들은 그 사건에 대한 뉴스를 보도하게 되고, 사용자는 그 중 어떤 한 뉴스를 읽음으로써 그 사건의 내용을 알게 될 것이다. 하지만 뉴스 추천 시스템에선 여전히 이미 읽은 뉴스와 함께, 다른 언론사에서 제공하는 같은 내용의 뉴스를 추천해 주게 되는 것을 알 수 있다. 뉴스 필터링은 사용자가 원하는 뉴스를 선정하고 걸러서 제공해 주기 위한 기술이므로, 사용자가 관심 있어 할 만한 뉴스의 순위를 매겨 추천 하는 것뿐 만 아니라, 위의 문제와 같이 사용자가 관심 없는 뉴스를 제외 시키는 것도 뉴스 필터링이라 할 수 있다.

이 논문에서는 이러한 새로운 문제를 뉴스의 시맨틱적인 접근 방법으로 해결하기 위하여 뉴스의 육하원칙 비교를 이용한 뉴스 필터링 방법을 제안한다.

이 논문의 구성은 다음과 같다. 2장에서는 뉴스 필터링과 이 논문이 제안하는 방법의 중요 요소인 육하원칙과 뉴스 온톨로지에 대한 관련연구를 기술하고, 3장에서는 제안 모델을 자세히 기술한다. 4장에서는 제안 방법을 실험과 평가를 기술한다. 마지막으로 5장에서는 결론 및 향후 연구 내용에 대하여 기술한다.

2. 관련 연구

이 절에서는 기존의 뉴스 필터링 기법들을 소개하고, 이 논문에서 신문기사의 유형과 제안하는 방법 중 뉴스의 육하원칙을 추출하기 위한 방법, 그리고 뉴스 온톨로지에 대한 연구들을 소개한다.

2.1 뉴스 필터링

뉴스 필터링 기술들은 크게 두 가지 방식이 있다. 첫 번째는 뉴스의 성격에 따라 사용자의 관심 토픽에 초점을 맞춘 Content-based Filtering 방법이고, 두 번째는 사용자의 사회적 환경을 보는 Collaborative Filtering 방법이다.

● Content-based Filtering

사용자의 선호 정보를 포함하는 온톨로지와 뉴스의 온톨로지 선정하여 그 안의 아이템들을 비교하여 가장 관심이 있을 법한 뉴스의 추천 순위를 매기는 방법이다 [5,6]. 이 방법은 사용자와 뉴스의 특징들의 의미를 가지고 뉴스의 추천 순위를 매기는 방법이기 때문에 이미 알고 있는 내용의 뉴스를 판별하지는 못한다. 그러므로 Content-based Filtering은 서론에서 도출한 문제점을 해결하지 못한다.

● Collaborative Filtering

사용자의 사회적 환경을 고려하여 다른 사용자들이 추천하는 뉴스를 추천해 주는 방법이다. 뉴스 필터링을 사

용하는 시스템은 사용자는 뉴스를 본 열람 내역(History Set)을 가지고 있으며, 이를 이용하여 다른 사용자들의 열람 내역과 비교하여 비슷한 주제의 뉴스를 선호하는 사용자들끼리 그룹을 만든다. 이때, 같은 그룹의 다른 사용자들이 본 뉴스를 사용자에게 추천하는 방법이다 [4]. 이 방법 역시, 다른 사용자들이 추천하는 뉴스를 이미 같은 내용의 다른 뉴스를 보았다고 해도 여전히 추천될 것이며, 사용자는 이러한 내용을 또 한번 읽게 되는 수고를 하게 될 것이다. 즉, Collaborative Filtering은 서론에서 도출한 문제점을 해결하지 못한다.

2.2 신문기사

신문기사(Newspaper)의 유형은 그 담긴 틀을 기준으로 해 외형적으로 분류하면 흔히 스트레이트(straight) 기사, 기획(feature) 기사, 인터뷰(interview), 사설이나 칼럼(editorial, column) 등으로 나눌 수 있다. 이 중에서 스트레이트 기사는 주로 사건이나 사고를 있는 그대로 육하원칙(누가, 언제, 어디서, 무엇을, 어떻게, 왜)에 입각하여 독자에게 전달하는 글이다 [7]. 기획, 인터뷰, 사설이나 칼럼 등은 저널리스트나 신문사 별로 사실 전달 보다는 주관적인 생각의 비중이 크거나, 같은 사건을 다루더라도 그 내용이 크게 달라질 수 있다. 따라서 이 논문에서는 육하원칙이 명확히 드러나는 스트레이트 기사에 대해서만 뉴스 필터링의 검증은 하도록 한다.

2.3 육하원칙 자동 추출

2.3절에서 기술한 것과 같이 사실을 정확히 전달해야 하는 뉴스의 성격상 육하원칙은 그 뉴스를 대표할 수 있을 만큼 중요하다. 때문에 이전부터 뉴스에서 육하원칙을 중심으로 정보를 추출하고, 내용을 요약하는 연구가 진행되어 왔는데 [8,9], 이러한 연구들은 육하원칙에 준하는 단어로써 뉴스의 정보를 추출한다. 이 논문에서는 뉴스에서 육하원칙을 추출하는 연구를 이용하여 육하원칙 온톨로지를 구축하고 활용한다.

2.4 뉴스 온톨로지

뉴스의 정보를 온톨로지로 만들기 위한 연구는 존재한다. 뉴스가 포함하는 정보를 사람이 직접 온톨로지로 구축하던 기계가 자동으로 해주던 간에 일반적인, 혹은 특정한 영역의 온톨로지를 사용하여 뉴스를 표현한다 [10]. 이 논문에서 뉴스는 이미 구축되어 있는 일반적인 온톨로지나 특정한 영역의 온톨로지를 사용하여 RDF와 같은 형식으로 표현되어 있다는 전제를 가진다.

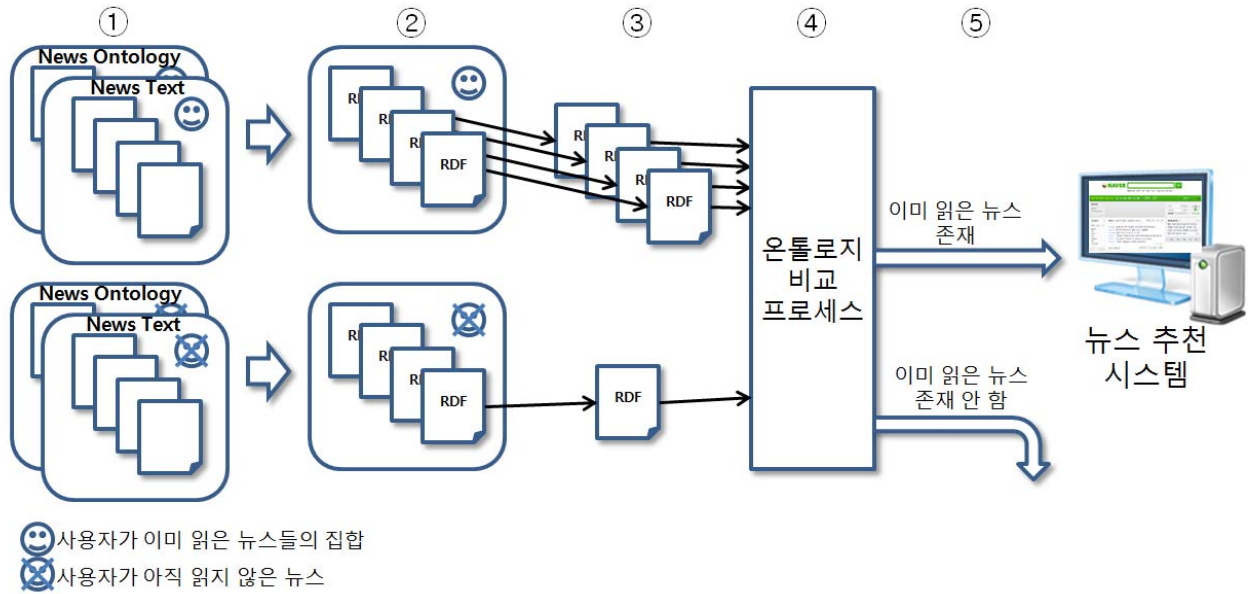


그림 2 제안 방법의 전체 프레임워크

3. 제안 방법

이 절에서는 이 논문에서 제시하는 문제점인 뉴스 추천 시스템에서 이미 읽은 뉴스와 같은 내용의 뉴스가 추천 되는 것을 피하기 위한 뉴스 필터링을 하는 방법을 제안한다. 그림 2는 제안 방법의 전체 프레임워크이며 5단계를 가진다. 이후 프레임워크의 각 단계를 자세히 기술함으로 제안하는 방법을 설명한다.

① 전제 조건과 제한 사항

제안 방법은 시작 시 다음과 같은 전제 조건과 제한 사항을 가진다. 첫 번째는 신문기사가 쓰여질 때, 온톨로지도 함께 구축이 된다. 이것은 Content-based Filtering의 방법에서 이미 온톨로지가 구축되어 사용되고 있는데, 이 Content-based Filtering의 방법이 선행되어 뉴스 온톨로지가 RDF 형태로 표현된다고 전제한다.

두 번째, 시스템은 사용자가 어떤 뉴스를 알고 있다는 것을 알기 위해 사용자의 열람 내역(History Set)과 다른 뉴스 필터링 기술을 사용하여 추천된 뉴스의 리스트를 가진다. Collaborative Filtering은 열람 내역을 가지고 뉴스를 추천해 주는 필터링 기술이므로, 제안 방법이 수행되기 전에 Collaborative Filtering 기법이 선행되어야 한다.

마지막으로 뉴스 추천 시스템에서 처리되는 뉴스들은 스트레이트 기사로 제한한다. 이 논문이 제안하는 방법은 육하원칙에 따른 뉴스의 정보추출이 중요한 부분을 차지한다. 따라서 육하원칙이 가장 명확하게 드러나는 스트레이트 기사만 제공한다는 제한 사항을 가진다.

즉 프레임워크는 위와 같은 전제 조건에 의해 두 가지 뉴스 집합을 가지고 시작한다. 하나는 사용자가 이미 읽은 뉴스들의 집합이고, 다른 하나는 사용자가 아직 읽지 않은 뉴스들의 집합이다. 즉 사용자가 이미 읽은 집합은 사용자의 열람내역으로 볼 수 있으며, 사용자가 아직 읽지 않은 뉴스 집합은 추천된 뉴스 리스트로 볼 수 있다. 그리고, 각각의 집합의 뉴스들은 텍스트로 된 뉴스 자체를 가지고 있을 뿐 만 아니라 추출된 온톨로지를 RDF의 형태로 가지고 있다.

② 육하원칙 온톨로지 추출

두 가지 뉴스 집합이 준비되었다면, 각각의 집합에서 뉴스에서 육하원칙을 추출하는 방법과 구축되어 있는 온톨로지를 사용하여 육하원칙 온톨로지를 생성한다. 이것은 이미 추출된 육하원칙의 요소들을 뉴스의 온톨로지로 연결시키는 작업으로 볼 수 있다. 관련 연구에서 육하원칙을 추출하는 연구는 뉴스의 핵심문장

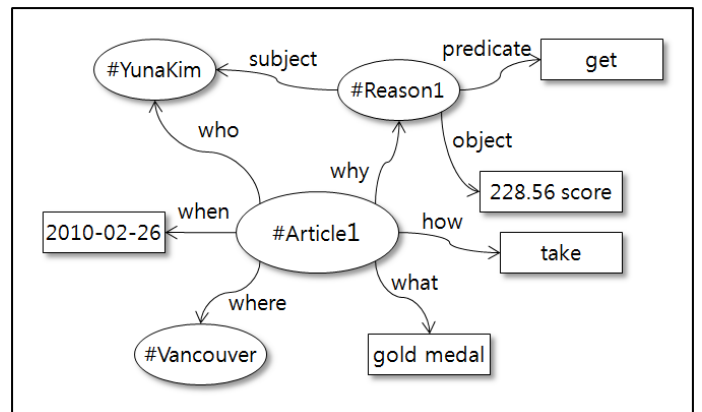


그림 3 육하원칙 온톨로지 예시

에서 중요 단어를 뽑는 작업이기 때문에, 육하원칙에 맞추어 중요 단어를 뽑은 후, 이미 구축되어 있는 뉴스 온톨로지의 URI정보를 가져와 육하원칙 온톨로지를 생성한다. 그림 3은 육하원칙에 맞추어 생성된 온톨로지의 예를 나타낸다. 뉴스의 ‘누가’, ‘언제’, ‘어디서’, ‘무엇을’, ‘어떻게’, ‘왜’ 요소들은 각각 ‘who’, ‘when’, ‘where’, ‘what’, ‘how’, ‘why’의 속성으로 표현되며 URI 혹은 Literal을 가지게 된다. 이때, ‘왜’의 요소는 문장 혹은 절로써 추출 될 수 있다. 이 경우 ‘왜’의 요소를 ‘subject’, ‘predicate’, ‘object’의 Statement를 가지는 온톨로지 표현된다.

③ 비교할 뉴스 선정

생성된 육하원칙 온톨로지를 사용하여 비교할 뉴스를 선정한다. 추천 받은 뉴스가 사용자가 이미 읽었던지, 알고 있는 내용인지를 비교하기 위해서는 사용자가 아직 읽지 않은 뉴스들의 집합 중 하나의 뉴스를 선정하고, 이 뉴스를 사용자가 이미 읽은 뉴스들의 집합의 모든 뉴스와 비교를 해야 한다. 결과적으로 사용자가 아직 읽지 않은 뉴스 집합에서 각각의 뉴스들을 사용자가 이미 읽은 뉴스들의 집합의 모든 뉴스와 비교해야 하므로, 총 두 집합의 뉴스 개수를 곱한 만큼의 비교 횟수를 가지게 된다.

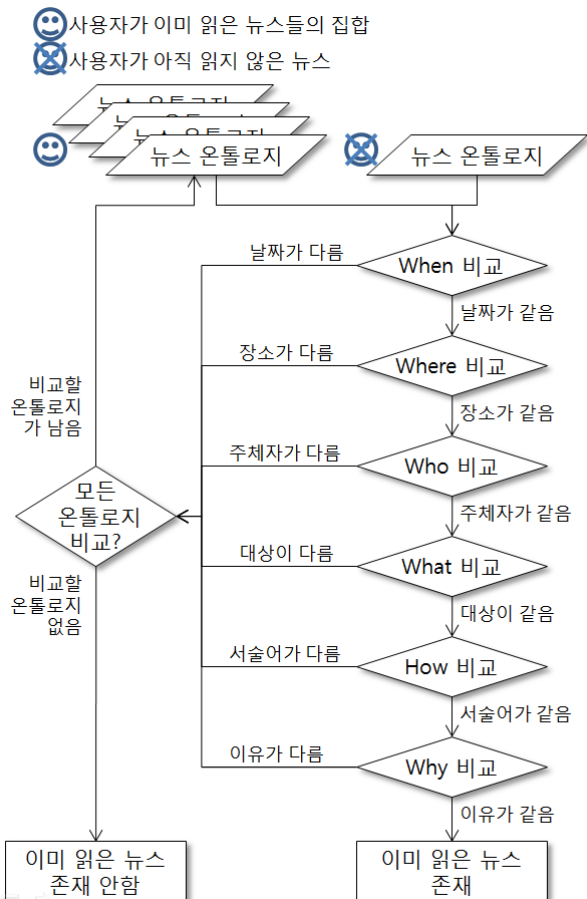


그림 4 온톨로지 비교 프로세스

④ 온톨로지 비교 프로세스

사용자가 아직 읽지 않은 뉴스들의 집합 중 하나의 뉴스가 선정이 되었다면, 사용자가 이미 읽은 뉴스들의 집합 중 하나씩 뉴스를 선정하면서 일대일로 육하원칙 온톨로지를 비교하며 같은 내용의 뉴스가 존재하는지를 평가한다.

두 뉴스가 같은가 다른가를 비교하기 위해서, 그림 4와 같은 육하원칙에 의한 뉴스의 유사도 비교 프로세스를 가진다. 사용자가 아직 읽지 않은 뉴스들의 집합 중 하나의 뉴스가 선정이 되었다면, 1:1로 내용을 비교하기 위해서 사용자가 이미 읽은 뉴스들의 집합 중 하나의 뉴스를 선정한다. 선정된 두 뉴스는 ‘누가’, ‘언제’, ‘어디서’, ‘무엇을’, ‘어떻게’, ‘왜’를 각각 비교해 본다. 그 후 여섯 가지 요소 중 하나의 요소라도 다르면, 사용자가 이미 읽은 뉴스들의 집합의 뉴스를 같은 집합의 아직 평가해보지 않은 뉴스로 선정 후 다시 평가를 시작한다. 만약 여섯 가지 요소 모두 일치하는 뉴스가 하나라도 존재한다면, 이미 읽은 뉴스 존재 판정을 받는다. 하지만 사용자가 아직 읽지 않은 뉴스들의 집합의 모든 뉴스를 평가했는데도 이미 읽은 뉴스 존재 판정이 나오지 않았다면 이미 읽은 뉴스 존재 안 함 판정을 받게 된다.

두 뉴스에서 육하원칙의 각각의 요소가 같은지를 비교할 때는 기본적으로 다음 3가지의 방법으로 확인해 본다.

- 두 요소의 URI가 같은지를 비교
- 두 요소의 Literal이 같은지를 비교
- Literal 혹은 URI의 이름이 동의어 관계에 있는지를 비교

‘누가’, ‘언제’, ‘어디서’, ‘무엇을’, ‘어떻게’, ‘왜’의 6가지 요소들을 비교할 때 위 그림에선 순서를 정했지만 어느 한 요소가 달라도 두 뉴스는 다른 뉴스가 되기 때문에 각각의 요소의 비교는 독립적이며, 순서가 바뀌어도 결과는 변하지 않는다.

⑤ 결과 적용

온톨로지 비교 프로세스의 결과로 ‘이미 읽은 뉴스 존재’라는 결론이 나왔다면, 그 뉴스는 최종적인 뉴스 추천 리스트에서 빠지게 된다. 하지만, ‘이미 읽은 뉴스 존재 안 함’이라는 결론이 나왔다면, 그 뉴스는 최종적인 뉴스 추천 리스트에 포함되어 최후로 사용자에게 보여지게 된다.

4. 실험 및 평가

이 절에서는 제안 방법을 케이스 별로 실험을 하고, 제안 방법에 대한 장점과 단점의 평가를 내려본다.

문제점을 해결하기 위한 방법 중 검증되어야 할 부분은 과연 육하원칙 온톨로지를 사용하여 비교한 것이 타당한가에 대해서이다. 제안하는 프레임워크의 모든 순서는 자동으로 진행 되는 것을 목표로 그 연구를 소개 했지만, 텍스트에서 온톨로지를 자동으로 추출 할 때와, 온톨로지 혹은 언어가 동의어 관계인지를 측정할 때, 한글에 서비스 되는 실제 시스템을 찾기가 어렵다는 문제점이 있었다. 때문에 이 논문은 뉴스에서 추출된 육하원칙 온톨로지를 생성하여 예측될 수 있는 비교 결과를 케이스 별 실험을 통하여 알아보았다.

● 비슷하지만 다른 뉴스인 경우

표 1은 두 뉴스가 비슷한 내용을 다루고 있지만 다른 뉴스인 경우를 나타낸다. 뉴스 1과 뉴스 2는 둘 다 칠레 산티아고에서 열린 월드컵평가전에 대한 내용을 다루고 있다. 하지만 뉴스 1은 3월 3일 월드컵 평가전이 본선 진출을 위해 열릴 예정을 보도한 내용이고 뉴스 2는 2월 28일 월드컵 평가전이 지진 피해로 인해 취소되었다는 내용을 다루고 있어 구체적인 일시와 평가전이 어떻게 된 것인지, 그 이유가 다르다는 것을 알 수 있다.

이런 경우는 두 뉴스가 비슷한 내용을 담고 있기 때문에 단어의 빈도수라던가, 온톨로지의 유사성 등으로 유사성을 검사할 때 높은 수치가 나올 수 있다. 하지만 본질적으로는 다른 사건을 다룬 뉴스이므로 다른 뉴스로 보는 것이 명확하다. 따라서 육하원칙을 사용하여 두 뉴스를 비교하는 것이 좋은 예라 할 수 있다.

표 1 비슷하지만 다른 뉴스인 경우

육하 원칙	뉴스 1		뉴스 2	
	URI or Literal	실제 표현	URI or Literal	실제 표현
Who	#칠레	칠레	#칠레	칠레
When	20100303	3월 3일	20100228	28일
Where	#산티아고	산티아고	#산티아고	산티아고
What	#월드컵 평가전	월드컵 평가전	#월드컵 평가전	월드컵 평가전
How	#예정	예정	#취소	취소
Why	본선 진출	본선 진출	#지진	지진 피해

● 다른 것 같지만 같은 뉴스인 경우

표 2는 두 뉴스가 다른 것 같지만 같은 뉴스인 경우를 나타낸다. 뉴스 3은 김연아가 2월 26일 228.56점을 얻어 밴쿠버 동계 올림픽에서 금메달을 딴 사건을 다루고 있다. 하지만 이 각각의 요소들은 What의 금메달만 빼고 모두 실제 표현 방법이 다르다. 하지만 이것이 가리키는 URI와 Literal은 거의 같고, How의 차지와 획득은 동의어 관계에 있으므로 결과적으로 모든 육하원칙 요소가 같다고 할 수 있다.

이런 경우는 단지 육하원칙의 단어를 추출하여 비교하는 것 보다 육하원칙 온톨로지의 URI와 Literal을 비교하므로 더 정확한 요소 별 비교가 가능함을 알 수 있다. 따라서 육하원칙을 온톨로지로 구축하여 비교하는 것이 타당함을 보인다.

표 2 다른 것 같지만 같은 뉴스인 경우

육하 원칙	뉴스 3		뉴스 4	
	URI or Literal	실제 표현	URI or Literal	실제 표현
Who	#김연아	김연아	#김연아	김연아(고려대20)
When	20100226	2월 26일	20100226	오늘
Where	#Vancouver	밴쿠버 퍼시픽 콜리세움	#Vancouver	밴쿠버 동계 올림픽
What	금메달	금메달	금메달	금메달
How	차지	차지	차지	획득
Why	#김연아, #get, 228.56	228.56점을 얻어	#김연아, #get, 228.56	신기록 228.56점 획득

● 같은 뉴스인데, 다른 뉴스라는 결과가 나올 경우

표 3은 두 뉴스가 같은 뉴스이지만, 다른 뉴스라는 결과가 나오는 경우를 나타낸다. 뉴스 5와 6은 나사의 앤드루 코우츠란 사람이 11월 26일 엔살라두스라는 토성의 위성에 물이 존재한다는 증거를 발견했다고 주장하는 같은 사건의 뉴스이다. 하지만 두 뉴스는 그 표현을 달리하여 거의 다른 육하원칙 온톨로지가 추출되었다.

이런 경우는 뉴스가 주장하는 바를 다르게 강조함으로 발생하는 오류로, 제안 방법으로는 해결하지 못한다는 단점이 있다. 따라서 좀 더 유연한 육하원칙 온톨로지 비교 방법이 필요하다.

표 3 같은 뉴스인데, 다른 뉴스라는 결과가 나올 경우

육하 원칙	뉴스 5		뉴스 6	
	URI or Literal	실제 표현	URI or Literal	실제 표현
Who	#앤드루	앤드루	#엔살라두스	엔살라두스
When	20081126	11월 26일	20081126	지난 26일
Where	#엔살라두스	엔살라두스	#NASA	나사
What	#물 #존재	물 존재	#지구 #가능성	제 2의 지구 가능성
How	#주장	주장	#제기	제기
Why	#증거 #발견	증거 발견	#음전하 #물 분자 #발견	음전하 물 분자 발견

5. 결론 및 향후 연구

이 논문은 신문사, 방송사 등 인터넷에 뉴스를 제공하는 언론사의 수가 증가함에 따라, 같은 사건을 각기 언론사에서 보도함으로써 뉴스 추천 시스템을 사용하는 사용자가 이미 읽은 뉴스 혹은 이미 읽은 뉴스와 동일한 내용의 뉴스를 다시 추천 받게 되는 문제점을 제기하였다. 또한, 사용자가 원하는 뉴스만을 보여주기 위한 뉴스 필터링 방법이, 사용자가 선호하는 뉴스를 뽑아 보여주는 것뿐 만 아니라, 사용자가 관심 없어하는 뉴스를 제외시키는 것도 새로운 뉴스 필터링이 될 수 있다는 점을 제시하였다. 기존의 뉴스 필터링들은 앞서 제기한 문제점을 해결하지 못함에 따라 기존의 뉴스 필터링에 추가적인 필터링 방법을 더함으로 해결 방법을 제시하였다. 기존 연구들과 한글 뉴스의 제한으로 정량적 평가 나타낼 수 있는 케이스 별 육하원칙 온톨로지 비교를 실험함으로써 제안하는 방법의 장, 단점을 알아보았다.

향후 추출된 육하원칙의 요소를 구축된 뉴스 온톨로지의 URI와 매핑 시키는 방법과 구체적인 육하원칙의 비교 방법에 대한 연구가 요구된다. 또한 연구된 방법을 구현하고 이를 통해 정량 평가를 수행함으로써 제안 방법에 대한 분석이 요구된다.

참 고 문 헌

- [1] 최종민, “인터넷 정보추출 에이전트”, 정보과학회지, vol.18, no.5, 2000.
- [2] 황용석, 양승혜, “2009 한국의 인터넷신문”, 한국언론재단, 2009.
- [3] 진승훈, 김종완, 김병만, “코호넨 신경망을 사용한 유즈넷 뉴스 필터링”, 한국정보과학회 가을 학술발표논문집, vol.29, no.2, 2002.
- [4] A. Das, M. Datar, A. Garg, “Google News Personalization: Scalable Online Collaborative Filtering,” Proc. of the 16th international conference on World Wide Web, pp.271-280, 2007.
- [5] P. Shoval, V. Maidel, B. Shapira, “An Ontology-content-based Filtering Method,” International Journal Information Theories & Application, vol.15, pp.303-314, 2008.
- [6] V. Maidel, P. Shoval, B. Shapira, M. Taieb-Maimon, “Evaluation of an Ontology-Content Based Filtering Method for a Personalized Newspaper,” Proc. of the 2008 ACM conference on Recommender systems, pp.91-98, 2008.
- [7] 고혜련, “신문 취재와 기사작성”, 중앙M&B, 2001.
- [8] 이현주, 김계성, 구상욱, 이상조, “신문기사에서 육하원칙 중심의 정보 추출”, 한국정보과학회 봄 학술발표논문집, vol.28, no.1, 2001.
- [9] 윤재민, 정유진, 이종혁, “육하원칙 활성화도를 이용한 신문기사 자동추출요약”, 정보과학회논문지 : 소프트웨어 및 응용, vol.31, no.4, 2004.
- [10] N. Fernandez-Garcia, L. Sanchez-Fernandez, J. M. Blazquez-del-Toro, J. Villamor-Lugo, “The News Ontology for Professional Journalism Applications,” Ontologies : A Handbook of Principles, Concepts and Applications in Information Systems, vol.14, Springer US, 2007.