

온톨로지 학습을 이용한 쿨벤더의 미래유망기술 키워드 추출

Extracting keyword of emerging technology using ontology learning in cool vendor

이 태 균, 신 건 철*, 김 수 경**
한밭대학교, 한밭대학교*, 한밭대학교**

Lee tae-kyun, Sin gun-chul*, Kim su-kyeong**
Hanbat National Univ., Hanbat National Univ.*,
Hanbat National Univ.**

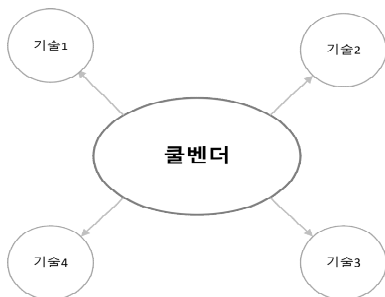
요약

최근 많은 기업 중에서 가트너는 매년 미래유망기술과 쿨벤더를 발표한다. 우리는 쿨벤더에서 제공하는 여러 정보들을 분석하여 미래유망기술에 대한 키워드를 찾고 이것을 실제 기술명과 연관짓고자 한다. 이 모든 과정의 전체적인 그림이 온톨로지 모델에 담긴다. 이 연구는 향후 어떤 집단의 미래를 이끌어갈 핵심 기술을 찾고자 하는 결정권자들에게 도움이 될 것이다.

I. 서론

현재 가트너(Gartner), 구글(Google), IBM 등 많은 기업에서 이미 미래의 유망 기술을 찾기 위해 노력하고 있다. 특히, 가트너에서는 매년 약 2,000개 정도의 유망 기술들을 발표하고, 이러한 유망 기술 분야에서 선도하고 있는 회사들을 ‘쿨벤더(Cool Vendor)’라는 이름으로 발표하고 있다. 쿨벤더 중에는 다른 기업으로부터 투자를 받은 곳도 포함되어 있다. 대표적인 예로, 최근에 알파고로 떠오르고 있는 ‘딥마인드(Deepmind)’는 2012년에 쿨벤더로 선정되어 2014년에 구글에 인수되어 투자를 받았다. 그리고 2016년에 이세돌과의 바둑 대결로 대중들로부터 많은 관심을 받았다.

앞의 사례와 같이, 우리도 쿨벤더로부터 나오는 키워드를 이용해 미래유망기술을 예측하고자 한다. 우리는 예측하는 과정에서 온톨로지를 이용한다. 이에 대한 개념적인 그림을 [그림 1]에 간단히 표현하였다.



▶▶ 그림 1. 쿨벤더-기술 연결 개념도

II. 관련 연구

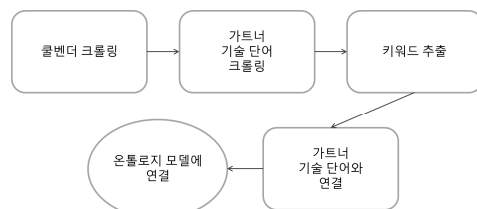
1. 쿨벤더(Cool Vendor)

쿨벤더는 매년 가트너에서 선정하는, 미래 기술을 이용한 혁신적인 경영으로 떠오르고 있는 기업이다. 가트너는 쿨벤더를 매년 선정하고 있는데, 기본적으로 쿨벤더에 대한 모든 것은 가트너 웹페이지에서 볼 수 있다 ([그림 2]). 2015년에 선정된 쿨벤더는 총 99개 분야에서 516개 이다. 이 중에는 다른 분야에서 중복되어 선정된 쿨벤더가 포함된다.

Gartner Cool Vendors Research	
Access to research documents may vary based on your subscription.	
Cool Vendors in Big Data, 2015 21 April 2015 Big data is past our Hype Cycle's Peak of Inflated Expectations and in the Trough of Disillusionment. Vendors that emerge will need greater functionality and consistent performance. Here, we present some Cool Vendors that are attempting to "make big data real."	Cool Vendors in the Internet of Things, 2015 30 April 2015 Innovative approaches for the Internet of Things focus on core building-block technologies in communications, IoT platforms and security. We profile five Cool Vendors that can help CIOs, IT leaders and decision makers build IoT solutions.
Cool Vendors in Mobile and Wireless, 2015 10 April 2015 Looking at innovative companies is important for mobile leaders when creating a digital strategy. This research introduces four cutting-edge companies that mobile leaders should consider in the context of a digital strategy and mobile security. We also review the evolution of two past Cool Vendors.	Cool Vendors in Cloud Security Services, 2015 24 April 2015 Gartner's Cool Vendors in cloud security services offer highly innovative technologies that address an organization's demand for agile, responsive and cost-effective solutions. CISOs should use this research when evaluating technology trends and planning future needs.

▶▶ 그림 2. 가트너 쿨벤더 2015

우리는 여기에서 쿨벤더로부터 키워드를 추출하고, 추출된 키워드로부터 미래유망기술을 예측하고자 한다. 이 과정을 도식화하여 [그림 3]에 보인다.



▶▶ 그림 3. 쿨벤더 연구 과정

웹페이지에 있는 쿨벤더에 대한 정보의 종류에는 분야(Field), 날짜(Date), 이름(Name), 위치(Location), 쿨벤더 선정이유(Why Cool), 벤더가 직면한 과제(Challenges), 해당 벤더를 누가 관심 있게 봐야 하는지(Who Should Care)가 있다. 여기에서 Why Cool, Challenges, Who Should Care를 이 글에서는 벤더에 대한 '정보'라 한다. 크롤링은 이러한 비정형 데이터를 모아 정형 데이터로 만드는 과정이다. 우리가 여기서 하고자 하는 연구를 진행하기 위해서는 쿨벤더에 대한 정보와 이것을 기술과 연결시키기 위해 가트너 웹페이지에 있는 기술 단어(Technology Glossary)가 필요하다. 따라서 이 두 가지 데이터를 크롤링 하는 것으로 시작한다.

그 후에, 쿨벤더 텍스트를 분석하여 키워드를 추출하고, 주요 키워드와 기술 단어를 연결함으로써, 쿨벤더와 기술을 연결한다.

2. 기술 키워드(Technology Keyword) 추출

기술 키워드라는 것은 쿨벤더 정보의 내용에서 기술에 해당하는 단어를 말한다. 여기에서 말하는 기술 단어는 앞에서 언급한 가트너 기술 단어를 기준으로 한다.

기술 키워드를 추출하기 전에, 일단 가트너 정보를 모두 파싱(Parsing)하여 그 중에서 명사구를 찾는 과정을 거친다. 이것의 결과에는 it, this와 같은 불용어(Stopword)가 포함되어 있다. 이 명사구에 대한 카운트를 [그림 4]에 내림차순으로 상위 10개 키워드만을 나열하였다.

KEYWORD	COUNT
it	1481
this	372
they	336
the company	280
users	199
data	198
them	180
there	175
customers	169
example	167

▶▶ 그림 4. 쿨벤더 정보 키워드

사실, [그림 4]의 키워드는 모두 기술 키워드가 아니므로 불용어이다. 우리는 이 키워드들 중에서 기술 단어만을 찾아 기술 키워드를 추출한다. 최종 결과물의 상위 10개 키워드를 [그림 5]에 보인다(* 표시는 하나의 쿨벤더에서만 나오는 키워드임).

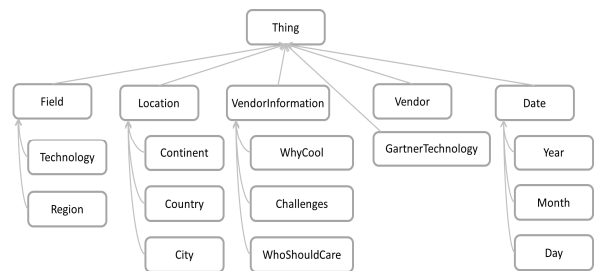
KEYWORD	COUNT
CIOS(Clinics in Orthopedic Surgery)	57
IoT(Internet of Things)	38
CSPs(Communications Service Providers)	35
Docker	29
Internet	27
Big Data	27
Salesforce	19
Cloud Services	17
Automation	17
Datacoup*	17

▶▶ 그림 5. 쿨벤더 정보 기술 키워드

[그림 4]와 [그림 5]의 실험 결과는 스탠포드 파서(Stanford Parser)를 이용한 결과이지만, 파서가 완전하지 않기 때문에 실험 결과에 기대만큼 만족스럽지는 않았다. 예를 들면, Machine Learning은 하나의 기술 키워드이지만 파서는 Machine과 Learning을 분리하여 추출한다.

3. 온톨로지(Ontology)

온톨로지 모델에 정의된 클래스는 크게 Field, Date, Vendor, Location, VendorInformation, GartnerTechnology 이다. 온톨로지 모델에 정의된 주요 클래스들의 상속관계를 그래프구조로 나타낸 것이 [그림 6]에 나와 있다.



▶▶ 그림 6. 쿨벤더 온톨로지 모델

[그림 6]에서 Field 클래스는 하위 클래스로 Technology, Region 클래스를 가지며, Location 클래스는 하위 클래스로 Continent, Country, City 클래스를 가지며, VendorInformation 클래스는 WhyCool, Challenges, WhoShouldCare 클래스를 가진다. 마지막으로, Date 클래스는 Year, Month, Day 클래스를 하위 클래스로 가진다.

III. 결론

본 논문은 온톨로지를 이용하여 쿨벤더의 여러 정보로부터 미래유망기술에 대한 키워드를 추출하는 과정에 대하여 기술하였다. 실험에서 쿨벤더 정보를 파싱(Parsing)함으로써 기술 관련 키워드를 얻을 수 있음을 확인하였다. 앞으로는 더 완벽하게 기술 키워드를 추출할 수 있는 파서를 개발하고, 워드넷(Wordnet)과 같은 데이터베이스를 도입하여 기술 동의어에 대한 처리가 가능하도록 하는 연구가 가능할 것이다.

■ 참고 문헌 ■

- [1] 곽수일, "디지털 비즈니스 모델", 학술세미나, pp.1-2, 2005.
- [2] 최장택 외, "2015년 KISTEP 10대 미래유망기술 선정에 관한 연구", 연구보고, pp. 1-4, 2015.
- [3] 김수경 외, "온톨로지 기반의 사용자 의도를 고려한 맞춤형 검색 서비스", 지능정보연구, pp. 132-134, 2012.