JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 26 No. 11, pp. 33-40, November 2021
https://doi.org/10.9708/jksci.2021.26.11.033

# A study on the Extraction of Similar Information using Knowledge Base Embedding for Battlefield Awareness

Sang-Min Kim*, So-Yeon Jin*, Woo-Sin Lee*

*Senior Engineer, Intelligent C4I Team, Hanwha Systems Co., Seongnam, Korea
*Chief Engineer, Intelligent C4I Team, Hanwha Systems Co., Seongnam, Korea
*Chief Engineer, Intelligent C4I Team, Hanwha Systems Co., Seongnam, Korea

## [Abstract]

Due to advanced complex strategies, the complexity of information that a commander must analyze is increasing. An intelligent service that can analyze battlefield is needed for the commander's timely judgment. This service consists of extracting knowledge from battlefield information, building a knowledge base, and analyzing the battlefield information from the knowledge base. This paper extract information similar to an input query by embedding the knowledge base built in the 2nd step. The transformation model is needed to generate the embedded knowledge base and uses the random-walk algorithm. The transformed information is embedding using Word2Vec, and Similar information is extracted through cosine similarity. In this paper, 980 sentences are generated from the open knowledge base and embedded as a 100-dimensional vector and it was confirmed that similar entities were extracted through cosine similarity.

▸ **Key words**: Knowledge Base, Embedding, Battlefield Awareness, Natural Language Processing, Artificial Intelligence

## [요 약]

고도화된 무기체계와 복잡한 전략으로 인하여 지휘관이 분석하고 판단해야 할 정보의 복잡도가 증가하고 있다. 지휘관의 적시적 판단을 위해서 전장의 정보를 지식화하고 분석할 수 있는 지능형 서비스가 필요하다. 지능형 서비스는 전장상황 정보로부터 지식을 추출하는 단계와 지식베이스를 구축하는 단계, 지식베이스로부터 전장상황을 분석하는 단계로 구성된다. 본 논문은 두 번째 단계에서 구축 완료된 지식베이스를 임베딩함으로써 입력 쿼리와 유사한 정보를 추출하는 방안을 연구한다. 지식베이스 임베딩을 위해 문장화 과정이 필요하며 random-walk 알고리즘을 적용한다. 문장화된 정보는 Word2Vec을 활용하여 벡터화되고 코사인 유사도를 통해 입력 쿼리와 유사한 정보를 찾는다. 본 논문에서는 오픈 지식베이스로부터 98개 개체를 기준으로 980개의 문장을 생성하고 100차원의 벡터로 임베딩함으로써 코사인 유사도 기반 유사 개체가 추출됨을 확인했다.

▸ **주제어**: 지식베이스, 임베딩, 전장 상황인식, 자연어처리, 인공지능

• First Author: Sang-Min Kim, Corresponding Author: Woo-Sin Lee
*Sang-Min Kim (smkim0153@hanwha.com), Intelligent C4I Team, Hanwha Systems Co.
*So-Yeon Jin (soyeon.jin@hanwha.com), Intelligent C4I Team, Hanwha Systems Co.
*Woo-Sin Lee (woosin.lee@hanwha.com), Intelligent C4I Team, Hanwha Systems Co.
• Received: 2021. 10. 21, Revised: 2021. 11. 24, Accepted: 2021. 11. 24.


http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

# I. Introduction

국방 지휘통제체계 성능이 고도화됨에 따라 지휘관은 전장상황에 대한 다양한 정보 수집 및 분석이 가능해졌다. 하지만 고도화된 무기체계를 기반으로 복잡한 전략이 이뤄지는 미래전장 양상을 미루어 볼 때 분석 대상 정보의 양이 폭증하여 지휘관의 적시적 판단의 어려움은 증가되었다[1]. 지휘관의 지휘결심을 지원하는 기술에 대한 필요성이 대두되었고, 선진국을 중심으로 관련연구가 진행중이다. 미국 국방부 산하 방위고등연구계획국(DARPA : Defense Advanced Research Projects Agency)은 다양하고 이질적인 비정형 정보로부터 문맥을 파악하고 명시적인 대안 해석을 도출하는 AIDA(Active Interpretation of Disparate Alternatives) 프로젝트를 진행 중 이다[2]. 이 프로젝트는 크게 3가지 분야로 나뉘는데 자연어 형태의 상황정보에서 지식을 추출하는 TA1(Task Area1)과 추출된 지식을 기반으로 지식베이스를 구축하는 TA2 그리고 지식베이스로부터 검색된 정보를 기반으로 입력 상황에 대한 다양한 해석을 도출하는 TA3로 구성된다. TA3의 목표를 이루기 위해서는 TA2에서 구축되는 지식베이스에 입력되는 질문과 유사성이 높은 정보를 찾아내는 것이 중요하다. AIDA에 참여한 SAMSON(Semantic Abduction over Multi-Source ObservatioNs)은 계층적 클러스터 공간을 정의하고 Relaxed Query를 제안하였다[3]. Relaxed Query는 최초의 Query 내 검색 정보들을 정의된 계층적 클러스터 공간으로 대체함으로써 지식베이스로부터 다양한 정보가 추출된다. [4] 연구는 지식베이스 온톨로지 스키마를 활용한 하이퍼그래프를 기반으로 연관질의를 제안하였다. 이 연구를 통해 최초 질의와 관련된 개체와 사건이 포함된 연관질의가 생성된다. 하지만 온톨로지 스키마를 기반으로 생성된 연관질의는 지식베이스의 원소스인 상황정보 문서에 담겨있는 의미정보를 반영하는 것에 한계를 갖는다. 지식베이스에 축적된 정보가 온톨로지 스키마 범주에 비해 부족하다면 충분한 정보를 추출할 수 없기 때문이다. 또한 지식베이스 구축 과정에서 발생 가능한 오타 또는 지식의 미추출 등의 문제에서 한계점을 갖는다. 지식베이스 임베딩 정보를 활용하면 원소스 문서에서 파생된 의미관계가 반영된 지식베이스로부터 유사한 정보 검색이 가능하다. 본 논문은 지식베이스 임베딩 정보를 기반으로 지식베이스로부터 유사한 답변 추출이 가능한 시스템을 제안한다. 2장에서 관련 배경지식을 살펴보고 3장에서는 제안하는 시스템을 설명한다. 4장에서는 구현을 통해 예시를 살펴보고, 5장에서 결론을 맺는다.

# II. Background

## 1. Knowledge Base Construction

본 절에서는 상황정보 문서로부터 지식을 추출하고 지식베이스를 구축하는 과정을 간략하게 살펴본다. 여기서 상황정보 문서는 비정형의 자연어로 구성되어 있음을 가정한다.

### 1.1 Ontology

지식베이스를 구축하기 위해서는 지식베이스의 기반이 되는 온톨로지가 정의되어야 한다. 온톨로지의 구성은 클래스, 인스턴스, 속성, 관계 등이 있으며 지식베이스에 저장될 정보들의 개념 계층도 확인이 가능하다. 따라서 지식베이스 도메인에 따라 온톨로지가 정의될 필요가 있다. 즉, 도메인에 따라서 특정 클래스에 포함이 가능한 인스턴스는 동일하지 않다. 온톨로지를 표현하는 언어는 일반적으로 RDF, OWL 등이 있으며 앞서 살펴본 SAMSON은 OWL 기반의 AIF(AIDA Interchange Format) 온톨로지를 활용하여 지식베이스를 구축하였다. Fig. 1.은 AIF 온톨로지의 예시를 보인다[5].

```
<https://tac.nist.gov/tracks/SM-KBP/2019/ontologies/InterchangeOntology> rdf:type
owl:Ontology ;
    rdfs:label "AIDA Interchange Format Ontology" ;
    rdfs:comment "This ontology is designed to represent information exchanged between
TA1, TA2 and TA3 in the AIDA program" ;
    owl:versionInfo "Version 1.0.2" .


################################################################
#    Classes
################################################################

########## Data Representation Classes ##########

:Entity a owl:Class ;
   rdfs:label "Entity" ;
   rdfs:comment "An entity in an AIDA KB, such as a person, organization, etc." .

:Event a owl:Class ;
   rdfs:label "Event" ;
   rdfs:comment "An event in an AIDA KB" .

:Relation a owl:Class ;
   rdfs:label "Relation" ;
   rdfs:comment "An relation in an AIDA KB" .

:SameAsCluster a owl:Class ;
   rdfs:label "Same-As Cluster";
   rdfs:comment "Represents a collection of events or entities which may in fact be the
same".

:Subgraph a owl:Class ;
   rdfs:label "Sub-graph" ;
   rdfs:comment "A collection of entities, events, relations, sentiment assertions, and
RDF statements about these".

:Hypothesis a owl:Class ;
    rdfs:label "Hypothesis" ;
    skos:definition "Represents a hypothesis, which consists of a collection of
entities, events, relations, sentiment assertions, and RDF statements about these." ;
    skos:note "All elements linked to a hypothesis are assumed to be conjoined" .
```

Fig. 1. AIDA Interchange Format Ontology

### 1.2 Knowledge Extraction

지식 추출은 상황정보 문서로부터 지식베이스에 저장할 지식을 추출하는 과정이다. 지식은 문장 속의 개체와 관계 등을 의미하며 자연어처리 분야의 인공지능 기술들이 이용되고 있다. LSTM(Long Short Term Memory), CRF(Conditional Random Field), BERT(Bidirectional Encoder Representaions from Transformer) 등은 그 예이다[6-8]. Fig. 2.는 ‘A sniper shot the protesters.’

라는 문자에서 지식이 추출된 예를 나타낸다. 지식베이스 도메인에 적합한 온톨로지와 지식 추출용 학습모델의 설계사항에 따라 지식 추출의 결과는 달라진다. 추출된 정보들은 지식베이스에 저장하기 위하여 RDF(Resource Description Framework) 형식을 따라 지식그래프로 표현된다. RDF 형식은 문장 속에서 추출된 지식을 'subject-predicate-object'로 구분하여 표현한다. RDF를 통해 지식베이스를 구성하면 입력 문장 내 지식들의 관계 정보가 내포된다. Fig. 2.의 결과를 RDF 형식의 지식그래프로 형상화하면 Fig. 3.과 같다.

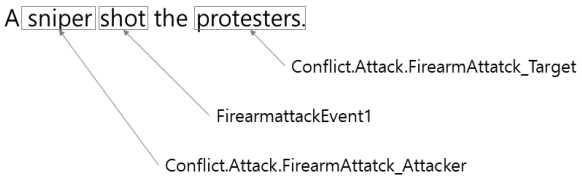


Fig. 2. Example of Knowledge Extraction

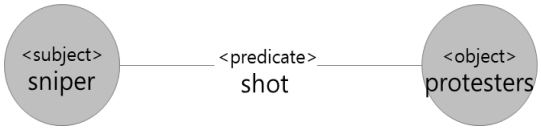


Fig. 3. Example of Knowledge graph using RDF

지금까지 지식베이스 구축 과정을 살펴보았다. 상황정보가 저장된 지식베이스는 전장상황에 대한 다중 해석을 도출하기 위한 기반 자료가 된다.

### 2. Embedding Method

지식베이스로부터 다양한 유사 정보를 검색하기 위해서 지식베이스의 임베딩 정보를 활용한다. 임베딩이란 문장 속 단어들을 벡터로의 표현을 의미하며 관련있는 단어들을 유사한 벡터로 표현한다. 본 절에서는 Word2Vec[9], Fasttext[10], GloVe[11]에 대해 소개한다.

Word2Vec은 Skip-gram과 CBOW(Continuous Bag of Words) 두 가지 방법론으로 구분된다. 두 방법론 모두 학습 단어 집합(window) 내에 단어들이 동시에 등장할 확률을 최대화하는데 학습의 목적이 있다. 차이점을 살펴보자면 Skip-gram은 학습 단어 집합(window)에서 중심 단어로부터 주변 단어를 학습하는 방법이며, 반대로 CBOW는 주변에 있는 단어로 중심단어를 예측하는 방법이다. 예를 들어, 'Struggle against fake information about events in Ukraine'이라는 예문을 학습하기 위하여 window 크기를 2로 설정하고 중심단어를 'information'이라고 가정하면 다음과 같다. Skip-gram은 주변 단어들로부터 'information' 예측하도록 학습하며, CBOW는 중심단어 주변 단어들을 입력으로 사용하여 'information'을 예측하도록 학습한다. 이를 표현하면 Fig. 4.와 Fig. 5.와 같다.

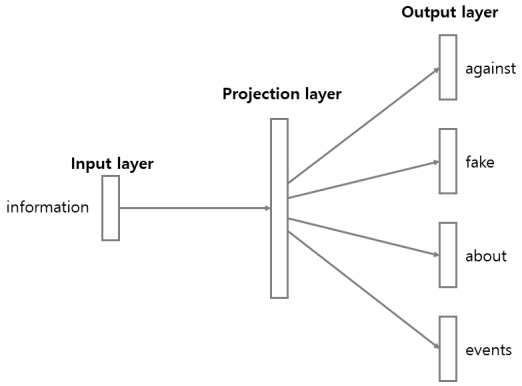


Fig. 4. Skip-gram architecture

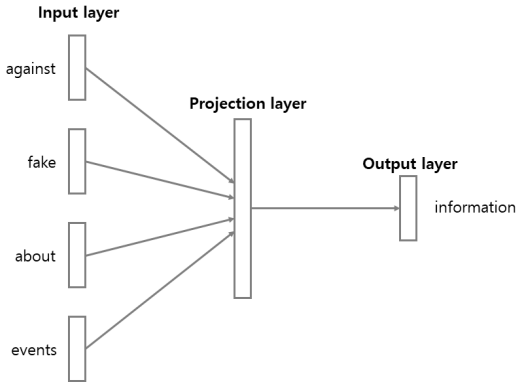


Fig. 5. CBOW architecture

GloVe는 Word2Vec의 관심영역인 window를 확장하여 대상 말뭉치내 모든 단어간의 영향도를 측정하도록 개선된 알고리즘이다. 이 접근법은 Word2Vec 대비 말뭉치 전체 정보를 임베딩에 반영할 수 있는 특징을 갖는다.

Fasttext는 Word2Vec의 Skip-gram으로부터 파생된 임베딩 알고리즘으로 단어를 문자의 ngram 조합으로 변환하여 임베딩하는 방법을 제안하였다. 예를 들어, ngram의 범위를 4로 설정하면 "fake"는 "<fak", "fake", "ake>"와 같은 subword로 구성된다. 여기서 <,>는 단어의 시작과 끝을 의미한다. 이렇게 단어를 문자단위의 subword로 표현하면 대상 말뭉치에 포함되지 않아 학습이 불가능했던 신조어에 대해서도 유의미한 임베딩이 가능하다. 하지만 최적의 임베딩 알고리즘은 입력 데이터의 특성에 좌우될 수 있으며 이는 관련연구를 통해서도 확인할 수 있다[12].

## III. The Proposed Scheme

본 논문은 지휘관의 적시적 판단을 지원하기 위해 지식

베이스 임베딩 정보를 기반으로 지식베이스로부터 유사한 답변 추출이 가능한 시스템을 제안한다. Fig. 6.은 제안하는 시스템이 적용된 개념도를 나타낸다. 입력된 전장상황 문서로부터 추출된 지식들로 지식베이스가 구축되고 제안하는 시스템을 적용하면 지휘관의 검색 요청으로부터 관련된 유사 정보를 다중으로 추출된다. 다양하게 추출된 정보들은 상황분석 모델의 입력으로 들어가 전장상황에 대하여 여러방면으로 해석할 수 있도록 돕는다.

이 과정은 1. 지식그래프 문장 변환 2. 임베딩 3. 쿼리로 구분되며 본 장에서는 단계 별 적용가능한 알고리즘 및 모델을 설명한다.

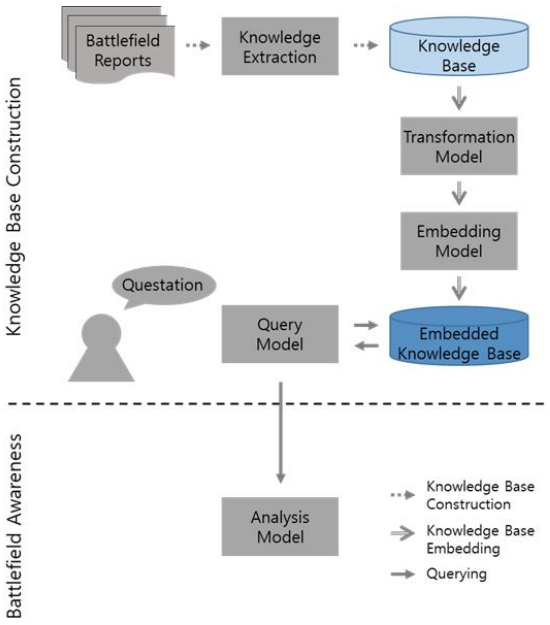


Fig. 6. Architecture of the Proposed System

Table 1. Description of the Proposed System

| Item | Contents |
|---|---|
| Battlefield Reports | A report that summarizes the battlefield |
| Knowledge Extraction | Natural language processing technology that extracts information (knowledge) of interest from battlefield reports |
| Knowledge Base | Knowledge base built on the basis of the extracted information |
| Transformation Model | Technology to convert knowledge graphs into sentences |
| Embedding Model | Technology to vectorize the transformed knowledge graph |
| Embedded Knowledge Base | Knowledge base embedded through Transformation Model and Embedding Model |
| Query Model | Technology to find query-like information in knowledge base |
| Analysis Model | Natural language processing technology to analyze the battlefield using the extracted similar information |

### 3.1 Transformation Model

지식그래프가 갖는 구조와 형식은 데이터의 의미를 잘 표현한다. 하지만 인공지능 기술을 직접 적용하기에는 한계가 있다. 따라서 인공지능 모델의 입력으로 사용되기 적절한 형태로의 변경이 필요하다. 그 첫 번째 과정이 그래프의 문장화 과정이다. Word2Vec 등 임베딩 모델의 학습 관점에서 적합한 입력의 형태는 단어로 이루어진 문장들이다. 따라서 지식베이스를 임베딩하기 위해서는 지식베이스의 정보들을 문장으로 변환할 필요가 있다. 즉 Fig. 3.과 같은 그래프 형태를 문장화한 후 임베딩 모델의 활용이 가능하다. 그래프 문장 변환을 위해 적용가능한 알고리즘들에 대해서 알아본다.

DeepWalk은 Random-walk 개념을 기반으로 그래프의 관심 노드와 인접 노드를 무작위로 선택하여 경로를(walk) 생성하고 이를 그래프의 문장으로 제안했다[13]. Fig. 7.은 전장 수집자산으로부터 수집된 정보를 지식화한 예시로써 노드를 탐지/보고된 객체로 표현하고, 동일 지역을 경유한 객체들을 엣지로 연결시킨 형태이다. 이러한 지식베이스에 DeepWalk 기법을 적용하면 Fig. 8.과 같은 문장이 선정된다. (walk length : 3) 선정된 문장들로 임베딩 알고리즘을 통해 학습을 하면 임베딩 벡터 상 Helicopter#AF3과 Warship#N2 등은 유사한 위치함을 확인할 수 있을 것이다.

DeepWalk는 무작위로 walk을 선택한다. 따라서 Fig. 8에 표현된 것과 같이 Warship#N2을 중심으로만 문장이 형성될 가능성이 있다. 이 경우 Warship#N1과 Tank#A2의 연결성을 확인하기 어렵다.

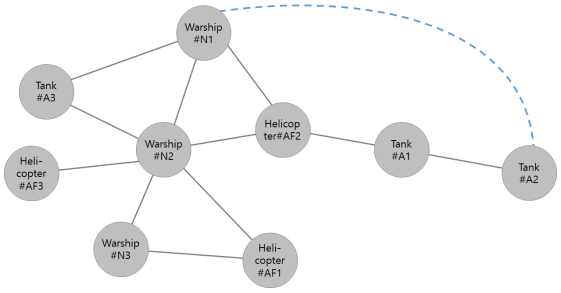


Fig. 7. Example of Knowledge Base

이러한 DeepWalk의 한계점을 개선한 알고리즘이 Node2Vec을 통해 제안되었다[14]. [14]는 'Return parameter'($p$)와 'In-out parameter'($q$)를 정의하고 walk 경로의 방향성을 조절하였다. '$p$'는 시작 노드로 되돌아오는 정도를 조절하고, '$q$'는 시작 노드로부터 얼마나

멀어지게 경로를 취할지를 조절한다. 따라서 walk length : 3 조건을 동일하게 적용하더라도 [14]는 '$q$' 조절을

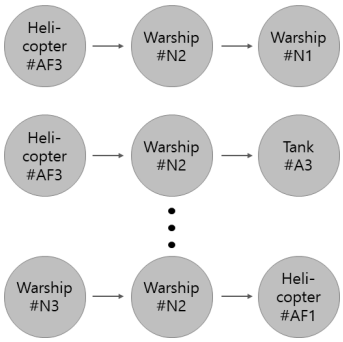


Fig. 8. Representaion of graph to walk using DeepWalk

통해 Fig. 9.와 같이 Warship#N1과 Tank#A2의 관계를 문장에 포함한다. [14]는 지식베이스의 특성에 따라 혹은 설계자의 의도에 따라 walk 선정 경로의 특성을 조절할 수 있으며 두 parameter를 값은 값으로($p$=$q$) 설정할 경우 [13]은 [14]와 동일하게 무작위 경로 설정을 진행한다.

전장상황을 담은 지식베이스는 Fig. 7.과 같은 단순 관계 그래프만으로 표현되기는 어려울 것이다. RDF 형식을 따라 지식그래프를 생성하면 복잡하고 다양한 전장상황을 담기에 적합할 것이다. 하지만 [13]과 [14]의 알고리즘을 RDF 지식그래프 변환에 직접 사용하기에는 무리가 있다. RDF 지식그래프는 Fig. 3.과 같이 엣지마다 새로운 관계가 정의되므로 기존의 방법을 사용할 경우 엣지가 갖는 관계 정보가 사라질 가능성이 있다.

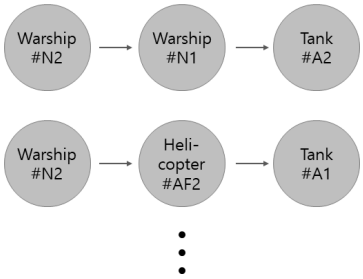


Fig. 9. Representaion of graph to walk using Node2Vec

제안하는 시스템은 지식베이스의 정보를 기반으로 검색과 추론이 이뤄지므로 지식그래프의 문장화 알고리즘은 지식그래프 내의 모든 지식요소를 문장화 할 수 있어야 한다. [15]의 접근방법은 RDF 형식 지식그래프의 변환에 특화된 알고리즘이다. [15]는 RDF 지식그래프가 가지고 있는 노드와 엣지 정보 모두를 사용하여 문장화할 수 있는 RDF2Vec을 제안하였다. Fig. 11. 은 RDF 형식의 지식그래프를 [15]에서 제안하는 방법을 이용하여 Fig.10.을 문장화한 예시이다. 노드에 연결된 엣지 정보를 이용하여 문장화를 진행하므로 노드 기반으로 문장화를 진행하는 기존 알고리즘과 달리 노드와 노드간의 관계 정보를 반영하여 문장화가 가능하다.

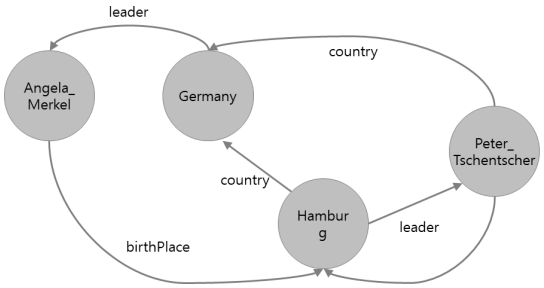


Fig. 10. Example of Knowledge Base in RDF

Hamburg -> country -> Germany -> leader -> Angela_Merkel

Germany -> leader -> Angela_Merkel -> birthplace -> Hamburg

Hamburg -> leader -> Peter_Tschentscher -> residence -> Hamburg

Fig. 11. Representaion of graph to walk using RDF2Vec

### 3.2 Embedding Model

지식그래프의 문장 변환이 완료되면 임베딩 모델을 이용하여 입력 정보가 벡터화된다. 앞서 살펴본 임베딩 기법 DeepWalk 및 Node2Vec, RDF2Vec 알고리즘은 모두 문장화된 지식그래프에 Word2Vec을 적용하여 임베딩하였다. II장 2절에서 살펴본 바와 같이 Word2Vec은 사용자가 설정한 window 안에 포함된 단어들이 동시에 등장할 확률을 극대화하는 방향으로 학습이 진행된다. 따라서 window 내에 포함되지 않는 단어들에 대해서는 유효한 임베딩 결과를 얻기 어렵다. 또한 지식베이스 구축 과정에서 발생할 수 있는 지식추출의 오류나 오타 등은 임베딩 결과에 문제를 영향을 준다. 이를 해결하기 위해 단어내의 문자 기준으로 임베딩을 학습하는 Fasttext 등과 같이 확장된 임베딩 알고리즘 적용을 고려할 필요가 있다.

### 3.3 Query Model

지식베이스에 대한 임베딩을 얻어내면 지식베이스로부터 지휘관이 관심을 갖는 키워드와 유사한 정보들을 추출한다. 이는 cosine similarity 등의 기법을 활용할 수 있으며 관심 키워드의 벡터값과 유사도가 높은 단어를 추출하면 된다. 추출된 단어들은 임베딩 되기 전 지식베이스에 쿼리하여 각 단어별로 연결된 지식들을 추출한다. Fig. 12.는 입력되는 키워드를 기반으로 지식이 추출되는 과정을 보인다.

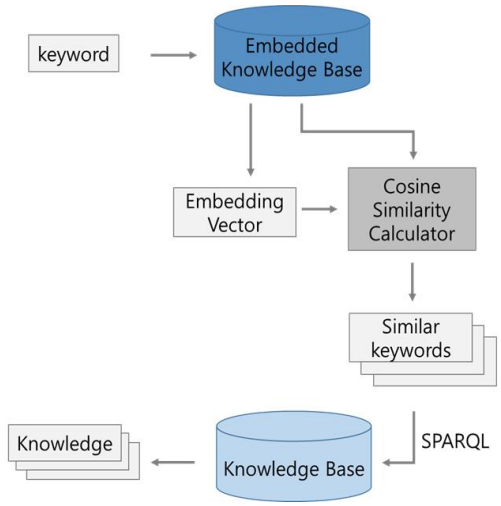


Fig. 12. Architecture of the Query Model

## IV. Implementation

본 장에서는 제안하는 시스템 기능 구현예시를 보인다. 구현의 구성은 지식베이스로부터의 문장 생성, 생성된 문장의 임베딩, 임베딩 결과의 활용으로 구분된다. 이 기능은 [16]에서 제공하는 소스코드를 기반으로 python3.9 버전에서 구현되었다.

**Data**:

KG = RDF Knowledge Graph
Embedder = word2vec
Walker = strategy of walks
Entities = Interested entities in KG

**Flow:**

```
1  RDF2VecTransformer(Embedder, Walkers)
2  RDF2VecTransformer.get_walks(KG, Entities)
3  RDF2VecTransformer.fit(Walks, Entities)
4  RDF2VecTransformer.transform(KG, Entities)
```

Fig. 13. Description of the RDF2Vec Function

국방 도메인의 지식베이스로부터 기능 구현을 보이기 위해서는 전장상황 문서 기반의 지식베이스 구축이 선행되어야 한다. 하지만 국방 특성상 해당 데이터에 대한 접근이 어려우므로 오픈 지식베이스를 통해 제안 시스템에 대한 가능성을 확인하고자 한다. 국방 도메인과 오픈 지식베이스의 도메인은 상이하지만, 지식그래프로부터 문장을 추출하고 임베딩을 통해 유사 정보를 추출하는 매커니즘은 동일하게 적용이 가능하다. 단, 국방 도메인에 대한 임베딩 모델의 재학습은 필요하다.

지식베이스로부터 문장을 추출하고 추출된 문장을 임베딩하는 각 모듈은 Fig. 13.과 같으며 모듈별 인자 값에 대한 설명은 다음과 같다. 'KG'는 RDF 형식의 지식베이스이며 본 구현에서는 'dbpedia.org'에서 제공하는 오픈 지식베이스를 활용한다. 임베딩 모델에 대한 정보는 'Embedder'에 설정한다. Walker는 지식그래프로부터 문장을 추출하기 위한 조건을 설정한다. 노드당 출력할 문장의 개수와 문장의 길이 등이 이에 해당한다.

**Data**:

KG = dbpedia/org
Walker = depth 10, length 4
Entities = Germany …

**Walks:**

| | |
|---|---|
| 1 | ('http://dbpedia.org/resource/Germany', 'http://dbpedia.org/ontology/wikiPageWikiLink', "b']\\xa0\\x00\\xeb*\\xf4\\xact'", 'http://dbpedia.org/ontology/wikiPageWikiLink', "b'\\xfbQ\\xd1_\\xd2\\x9b$z'", 'http://purl.org/dc/terms/subject', "b'\\xfbA\\xc6\\xe2\\xc4v\\xbe%'", 'http://www.w3.org/2004/02/skos/core#broader', "b'\\xc1\\xa3\\xce\\xc3\\xa8\\x0bK\\x8d'"), |
| 2 | ('http://dbpedia.org/resource/Germany', 'http://dbpedia.org/ontology/wikiPageWikiLink', "b']\\xa0\\x00\\xeb*\\xf4\\xact'", 'http://dbpedia.org/ontology/wikiPageWikiLink', "b'\\xfbQ\\xd1_\\xd2\\x9b$z'", 'http://xmlns.com/foaf/0.1/isPrimaryTopicOf', "b'\\xd3\\xe2\\x88Yz^\\x95F'", 'http://purl.org/dc/elements/1.1/language', "b'\\x9c\\xfe\\xfe\\xd8\\xfb\\x94\\x97\\xba'"), |
| 3 | ('http://dbpedia.org/resource/Germany', 'http://dbpedia.org/ontology/wikiPageWikiLink', "b']\\xa0\\x00\\xeb*\\xf4\\xact'", 'http://dbpedia.org/ontology/wikiPageWikiLink', "b'\\xfbQ\\xd1_\\xd2\\x9b$z'", 'http://xmlns.com/foaf/0.1/isPrimaryTopicOf', "b'\\xd3\\xe2\\x88Yz^\\x95F'"), |
| … | … |

Fig. 14. Example of the extracted walks

'Entities'는 지식베이스 내에서 임베딩을 진행할 그래프들의 기준이 되는 개체들을 의미한다. Entities에 입력되는 개체들을 기준으로 문장이 생성된다. Fig. 14.는 KG에서 특정 개체(국가 및 수도)와 관련된 문장을 노드 당 10개의 문장을 4의 길이 조건으로 추출한 예시이다. 'Germany' 노드로부터 RDF 속성인 predicate를 반영하여 문장화가 진행되었다. 이 결과를 시각화하면 Fig. 15.와 같다.

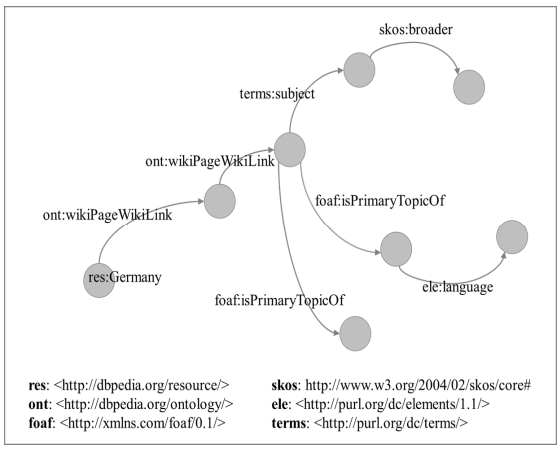


Fig. 15. Representation of the graph based on extracted walks

이와 같이 문장화된 지식그래프는 Word2Vec등의 임베딩 모델을 통하여 벡터화 된다. 본 구현에서는 98개의 개체를 기준으로 추출된 980개의 문장을 Word2Vec의 Skip-gram 알고리즘을 활용하여 학습한 후 100차원의 임베딩 결과를 얻었다. 그 결과를 Fig. 16.에 시각화하였다.

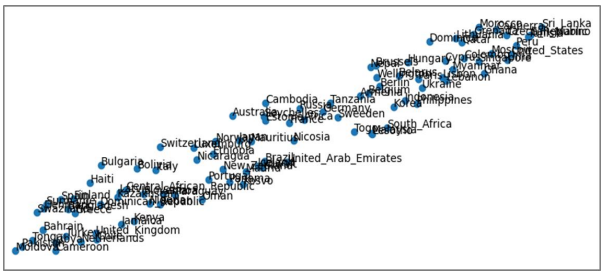


Fig. 16. Representaion of graph to walk using Dee

python API ‘gensim.similarities’ 클래스를 이용하여 ‘Cameroon’과 유사도가 높은 개체 ‘Canberra’, ‘Brussels’를 찾을 수 있었다. ‘Cameroon’을 지휘관의 관심 키워드라고 가정할 경우 제안하는 시스템은 지휘관에게 ‘Cameroon’ 뿐만 아니라 유사도가 높은 ‘Canberra’와‘Brussels’라는 키워드를 제공할 수 있을 뿐만 아니라 KG 쿼리를 통해 유사 단어들과 연결된 지식까지 제공이 가능하다.

현재 우리군은 분석관의 경험과 판단을 기반으로 전장상황을 판단하고 있는 실정이다. 향후 우리군에 제안하는 시스템과 같이 지휘관의 지휘결심을 지원하는 기술이 적용된다면 분석관 개인이 보유한 경험, 역량 등의 제약에서 벗어나 수많은 양의 정보를 분석하여 지휘관의 적시적 판단을 도울 수 있다.

## V. Conclusions

본 논문은 고도화되어 가는 미래전에서 지휘관의 적시적 판단을 지원하기 위한 임베딩 기반의 지식베이스 활용 시스템을 제안하였다. 해당 시스템은 지식그래프 기반의 정보들을 문장화한 후 임베딩함으로써 전장상황을 해석하기 위한 인공지능 모델들에 적합한 형태를 생성한다.

임베딩된 지식베이스로부터 전장상황을 분석하기 위한 지휘관의 관심 개체와 유사한 개체들을 추출함으로써 확장된 키워드 검색이 가능하며 이를 기반으로 지식베이스로부터 폭넓은 정보가 추출된다. 추출된 정보는 전장상황 해석을 위한 인공지능 모델의 입력으로 사용되어 다양한 해석과 판단이 수행된다.

향후 자연어 문맥에 대한 이해와 학습효율이 향상된 사전학습모델 Electra[17], 한국어에 특화된 학습모델[18-19] 등을 활용하여 유사정보를 추출하는 방안을 고도화할 예정이다.

## REFERENCES

[1] Se-Hyeon Jo and Hack-Jun Kim, "A Study on Building Knowledge Base for Intelligent Battlefield Awareness Service," Journal of the Korea Society of Computer and Information 25(4), 2020.4, 11-17(7 pages)

[2] M. Li, et al., “GAIA at SM-KBP 2019 - Amulti-media multi-lingual knowledge extractionand hypothesis generation system,” TACSM-KMP, 2019

[3] Andersen, Carl F. et al. “KB Construction and Hypothesis Generation Using SAMSON.” TAC (2019).

[4] So-yeon Jin and Woo-sin Lee. ”A Study on Multiple Reasoning Technology for Intelligent Battlefield Situational Awareness” The Journal of Korean Institute of Communications and Information Sciences 45(6), 2020.6,1046-1055 (10 pages)

[5] TAC ( Text Analysis Conference) StreamingMultimedia Knowledge Base Population(SM-KBP) 2019 guidelines, Retrieved May 8,2020, from https://tac.nist.gov/tracks/SM-KBP/2019/ontologies/LDCOwlOntology.

[6] Hasim Sak, Andrew Senior, Franscois Beaufays "Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling," In proc. Interspeech.

[7] Zhiheng Huang, Wei Xu, Kai Yu, “Bidirectional LSTM-CRF Models for Sequence Tagging, ” arXiv preprint arXiv:1508.01991, 2015.

[8] Devlin, Jacob, et al. “Bert: Pre-training of deep bidirectional transformers for language understanding.” arXiv preprint arXiv:1810.04805, 2018.

[9] T. Mikolov, K. Chen, G. Corrado, and J. Dean. “Efficient estimation of word representations in vector space.”, In Proceedings of ICLR Workshops Track, 2013. arxiv.org/abs/1301.3781.

[10] P. Bojanowski et al., “Enriching word vectors with subword information,” Transactions of the Asociation for Computational Linguistics, No. 5, p. 135-146, 2017

[11] J. Penington, R. Socher, & C. D. Maning, “Glove: Global vectors for word representation,” In Procedings of the 2014 conference on empirical methods in naturalanguage procesing (EMNLP), p. 1532-1543, Oct. 2014.

[12] Hyungsuc Kang, Janghoon Yang, "Performance Comparison of Word2vec and fastText Embedding Models", Journal of Digital Contents Society 21(7), 2020.7, 1335-1343 (9 pages)

[13] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 701-710. ACM, 2014.

[14] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016.

[15] Petar Ristoski, Jessica Rosati, Tommaso Di Noia, Renato De Leone, and Heiko Paulheim. 2018. RDF2Vec: RDF Graph Embeddings and Their Applications. Semantic Web, 10(4):721 -752.

[16] @inproceedings{pyrdf2vec, author = {Gilles Vandewiele and Bram Steenwinckel and Terencio Agozzino and Michael Weyns and Pieter Bonte and Femke Ongenae and Filip De Turck}, title = {{pyRDF2Vec: Python Implementation and Extension of RDF2Vec}}, organization = {IDLab}, year = {2020}, url = {https://github.com/IBCNServices/pyRDF2Vec}}

[17] K. Clark et al., “ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators.” in Int. Conf. Learning Representations, Addis Ababa, Ethiopia, May 2020.

[18] KoBERT, https://github.com/SKTBrain/KoBERT

[19] KoELECTRA, https://github.com/monologg/KoELECTRA

## Authors

Sang-Min Kim received the B.S., M.S. degrees in Electronic Engineering from Kwangwoon University, Korea, in 2013, 2015, respectively. Sang-Min Kim is currently a senior engineering in Hanwha systems.

He is interested in natural language processing and deep learning.

So-Yeon Jin received the B.S. degree in Computer Engineering from Chonbuk National University, Korea, in 2003. So Yeon Jin is currently a chief engineer in Hanwha systems.

She is interested in data links, machine learning, military communications, unmanned systems.

Woo-Sin Lee received the B.S., M.S. and Ph.D. degrees in Computer Engineering from Kwangwoon University, Korea, in 2001, 2003 and 2007, respectively. Dr. Lee is currently a chief engineer in Hanwha Systems.

He is interested in data links, tactical networks.