Expert Systems With Applications 141 (2020) 112965

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Automating the expansion of a knowledge graph

SoYeop Yoo, OkRan Jeong*

Gachon University, 1342, Seongnam-daero, Sujeong-gu, Seongnam-si, Gyeonggi-do, Republic of Korea

A R T I C L E I N F O

Article history:
Received 5 October 2018
Revised 18 September 2019
Accepted 18 September 2019
Available online 19 September 2019

Keywords:
Auto expansion
Knowledge graph
Neologisms
Semantic analysis
Multilingual BERT model

A B S T R A C T

In order to make computers understand human languages and to reason, human knowledge needs to be represented and stored in a form that can be processed by computers. Knowledge graphs have been developed for use as a form of the knowledge base for words and general relationships among words. However, they have two limitations. One is that the knowledge graph is limited in size and scope for most of the human languages. Another is that they are not able to deal with neologisms that form a part of the human common sense. Addressing these problems, we have developed and validated PolarisX which can automatically expand a knowledge graph, by crawling and analyzing the news sites and social media in real-time. We utilize and fine-tune the pre-trained multilingual BERT model for the construction of knowledge graphs without language dependencies. We extract new relationships using the BERT-based relation extraction model and integrate them into the knowledge graph. We verify the novelty and accuracy of PolarisX. It deals with neologisms and does not have language dependencies.

1. Introduction

The fundamental problem in making computers understand human languages is to store human knowledge in computers. During the past fifty years, researchers have developed various ways to represent knowledge. One of them is knowledge graphs. Knowledge graphs are a means of storing and using data, which allows people and machines to better tap into the connections in their datasets. They represent words as nodes and the relationships as edges between words (Paulheim, 2016; Singh et al., 2002; Singhal, 2012; Zhang, 2002).

There are many definitions in the knowledge graph. In Lisa and Wolfram (2016), the knowledge graph is defined as follows: “A knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge.” As this definition describes, ontology contains all the information collected like a knowledge base. On the other hand, the knowledge graph contains knowledge extracted from the collected information. In other words, the knowledge graph can be regarded as a superordinate concept of the knowledge base.

Currently, most knowledge graphs have two important limitations. One is the size and scope of the knowledge base which most knowledge graphs have. The knowledge base is rather large for a small number of core languages, such as English, French, Chinese, and so on. However, it is rather small, and so not very useful, for most other languages, such as Korean, Turkish, and so on.

Another limitation of the knowledge graphs is that they are not able to deal with neologisms, that is, newly coined words or new meanings for established words. Neologism[1] means a new word or expression, or a new meaning for an existing word. Neologisms reflect the cultural and social trends that in turn become part of the common sense of a large group of people using a particular human language. Knowledge graphs can support the search for established words and other words connected to them because there is a sufficient amount of stored data. However, they are not able to support the search for neologisms until the amount of stored data crosses a certain threshold (Pujara, Miao, Getoor & Cohen, 2013).

Neologisms continue to become meaningful in people’s lives, but existing studies are only focusing on neologisms, especially when existing words have new meanings. In a sentence, new words that are not already in place are usually out-of-bag and are not recognized as new words. It is a very big deal to figure out new words in the text. In sources where people are free to write text, such as social media, the role of these neologisms is so big that it is an important challenge to deal with neologisms on knowledge graphs.

Table 1 shows some examples of neologisms. There are many words that are newly coined or have new definitions. The meanings of the neologisms are from Urban Dictionary.[2] For example,

* Corresponding author.
E-mail address: orjeong@gachon.ac.kr (O. Jeong).

[1] https://dictionary.cambridge.org/dictionary/english/neologism.
[2] https://www.urbandictionary.com/.

https://doi.org/10.1016/j.eswa.2019.112965

Table 1
Example neologisms.

Word	The existing meaning	Newly added meaning
(a) Neologisms which are new words		
Selfie	–	Selfies are a style of photography wherein (1) the photographer's own face is included in the photograph, (2) and the camera is held by the photographer when the photo is taken.
Smombie	–	A person walking around unaware of his or her surroundings entirely absorbed in their smartphone.
Chillax	–	A mixture of the terms "chill out" and "relax".
Staycation	–	At times of recession, people prefer to spend time on the beaches of their own country instead of flying to an overseas environment creating a carbon footprint and spending money.
(b) Neologisms which have a new meaning for existing words		
Apple	a kind of fruit	IT brand
Ford	a location where a stream is shallow	motor company
Gangnam style	the lifestyle of the Gangnam, Seoul	K-pop by PSY
Trump	playing card	the 45th president of the United States

Table 2
Definitions.

Term	Definition
Knowledge Graph (KG)	A knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge (Lisa & Wolfram, 2016)
Knowledge Base (KB)	Stored information as a subconcept of knowledge graph
Neologism	1) a new word or expression, or 2) a new meaning for an existing word
Word	An entity/node in the knowledge graph
Relation	A relation/edge in the knowledge graph

apple used to mean a kind of fruit. However, now it also refers to the company behind the iPhone.

In order to make knowledge graphs keep pace with the creation of neologisms, we need to expand the knowledge graphs continuously and effectively. A "natural", or naïve, approach would be to hope that many people over time would manually add neologisms to the knowledge graphs. However, this is not practical, since many neologisms are being created and it is difficult for a person to add and identify related data to any particular neologism.

We believe that a much better way is to automatically crawl the social media for neologisms, to extract new knowledge especially for relations and to automatically add them to the knowledge graphs. After all, neologisms reflect the common sense of a large group of people in a particular country/region and using a particular language. So, crawling social media is a natural way to collect neologisms.

We propose an auto growing knowledge graph, called PolarisX (Polaris Expander). To make PolarisX growing itself, we crawl news sites and social media, extract new relationships, generate knowledge subgraphs and append them to knowledge graphs. PolarisX is a part of Polaris project. Polaris is a project to build a framework for conversational AI (Artificial Intelligence) based on big data. The Polaris project consists of an auto-growing knowledge graph and a big data analysis and prediction system (Yoo, Song & Jeong, 2018). An auto-growing knowledge graph that can deal with neologisms, PolarisX, is a very important component of the Polaris project.

The PolarisX can be helpful to understand a person's commonsense in rapidly expanding knowledge graphs, with established words and neologisms alike, and for core languages and non-core languages alike. We have implemented the PolarisX and validated it using ConceptNet (Havasi, Speer & Alonso, 2009; Speer, Chin & Havasi, 2017), which is an open-source knowledge graph. We describe the architecture of PolarisX and discuss the results of its validation.

In this paper, our major contributions are as follows.

(1) Dealing with Neologisms: the existing knowledge graphs build a graph using the existing data or crowdsourcing. However, these could not respond to neologisms that are newly coined or have new meanings. We propose PolarisX to expand the knowledge graph with real-time social media and news data.
(2) Language Dependency-free: the existing knowledge graphs have lots of data in major languages such as English and French. Most knowledge graphs do not have enough data in other common languages. We expand data sources, figure out new relationship knowledge using BERT (Devlin, Chang, Lee & Toutanova, 2018) and integrate into the expanded knowledge graph. We use a multilingual BERT model so it is no matter to consider the amount of data according to language.

2. Definitions of key terms

We continue to use the following terms in the paper: knowledge graph, knowledge base, neologism, word, and relation. There are terms that are used in the same formal definition, but we summarize those terms briefly to clarify what we intend to convey in the paper.

Table 2 shows the definitions of the key terms used in the paper. Knowledge graphs and knowledge base are used in various definitions from existing research, but we use the definition as shown in Table 2. The biggest difference between the knowledge graph and base is the difference between knowledge and information. The knowledge base contains all the information that is collected such as an ontology. On the other hand, knowledge graphs contain knowledge derived from the knowledge base. So knowledge graphs are the superior concept of the knowledge base. We mentioned it before, knowledge graph can be regarded as a superordinate concept of the knowledge base (Lisa & Wolfram, 2016).

In artificial intelligence applications, ontologies commonly used as knowledge base because ontological representations mainly play a role in semantic modeling of knowledge (Studer, Benjamins & Fensel, 1998). The main difference between a knowledge graph and an ontology could be either a matter of quantity or extended requirements such as a reasoning engine. Therefore, reasoning capabilities are considered as an important characteristic to derive new knowledge and differentiate a knowledge graph from knowledge bases (Guarino, Guarino & Giaretta, 1995; Lisa & Wolfram, 2016; Studer et al., 1998).

In this paper, we use the definition of neologism from the Cambridge dictionary as we mentioned in Section 1. Neologism continues to become meaningful, but existing research only focuses on neologism especially a new meaning for an existing word. In social media, people are free to generate texts so there are many neologisms that are new words or new meanings for existing words. It is very important to figure out the meaning of neologisms to understand the generated texts.

In a knowledge graph, terms such as entity, node, and edge are usually used in the mathematical or network graph. However, since we mainly extract relationships from text and connect them to graphs, we would like to clarify by using the term ‘word’ instead of entity/node, and the term ‘relation’ instead of the edge. Relationships such as composition, aggregation, inheritance, and association are representative relationships that can mainly express relationships among the existing classes. Though there are numerous relationships in the knowledge base, the knowledge graph should be able to extract meaningful knowledge essentially and so it is expressed with four representative relationships (*IsA, HasA, RelatedTo*, and *PartOf*).

3. Related work

With the development of artificial intelligence technology, many studies are being conducted to incorporate AI technologies in various fields such as chatbot (Abdul-Kader & Wood, 2015) and conversational AI (Gao, Galley & Li, 2019). In artificial intelligence technology, it is important to make computers understand a person’s common sense. It is because computers should understand a person’s intention, look for information, and show the result just like a human.

Making computers understand human languages and recognize objects people see has been a longstanding research objective of artificial intelligence (Paulheim, 2016; Singh et al., 2002; Singhal, 2012; Zhang, 2002). In order to realize this objective, data about the real world and human knowledge has to be represented in a form that can be processed by computers. Several forms of knowledge representation have been developed, such as knowledge graphs, semantic nets, and linked open data.

A semantic network is a knowledge base that represents semantic relations between concepts. It is often used as a form of knowledge representation (Sowa, 2006). Linked open data is linked data which is released under an open license data (Yu, 2011). DBpedia (Auer et al., 2007) and Wikidata (Vrandečić & Denny, 2012; Vrandečić & Krötzsch, 2014) are examples of large linked open data sets.

Knowledge graphs have been used in many projects, such as Cyc, Wikidata, Dbpedia, and so on. The descriptions about knowledge graphs are as follows:

- Cyc (Lenat, 1995): is from the Cyc project which dates back to the 1980s. The Cyc knowledge graph has been developed and maintained by CyCorp.
- Wikidata (Vrandečić & Denny, 2012): is a collaboratively edited knowledge graph. It acts as central storage for the structured data of Wikimedia projects including Wikipedia (Bruns, 2008), Wikivoyage, Wikisource, and others. It has data moved from the Freebase graph (Bollacker, Evans, Paritosh, Sturge & Taylor, 2008).
- DBpedia (Auer et al., 2007): is a knowledge graph that is extracted from structured data in Wikipedia.
- YAGO (Suchanek, Kasneci & Weikum, 2007): is also extracted from Wikipedia and builds classification from the category system in Wikipedia and the lexical resource WordNet (Miller, 1995).
- NELL (Carlson, Betteridge, Wang, Hruschka & Mitchell, 2010): is work on a large scale corpus of websites and exploits a coupled process that learns text patterns corresponding to type and relation assertions.
- Google Knowledge Graph:[3] Google launched a service called Knowledge Graph in 2012 to enhance its search engine results with information gathered from a variety of sources.[4] This information is about the topic in addition to a list of links to other sites and presented to users in a box to the right or top of search results. The Knowledge Graph makes use of Wikidata and various other data sources.

Table 3
Number of nodes and edges in representative knowledge graphs.

Name	Words	Relations	Relation types
DBpedia (English)	4806,150	176,043,129	2813
YAGO	4595,906	25,946,870	77
Wikidata	15,602,060	65,993,797	37,781
NELL	2006,896	432,845	425
OpenCyc	118,499	2413,894	18,526
ConceptNet	8000,000	32,755,210	34
Googles Knowledge Graph	570,000,000	18,000,000,000	35,000

Table 3 shows the number of nodes and edges in well-known knowledge graphs. Table 3 is referred from the research and changes the column names to fit the terms we defined as of 2018.

As lots of data are newly created, the knowledge graph needs to be dynamically expanded. Much research has been done (Li, Taheri, Tu & Gimpel, 2016; Paulheim, 2016; Tandon, de Melo, Suchanek & Weikum, 2014; Trivedi, Dai, Wang & Song, 2017; Wang et al., 2013; Yan, Wang, Cheng, Gao & Zhou, 2018) on expanding knowledge graphs. Most knowledge graphs use machine learning methods to add new data and connect new relations. They could expand the knowledge graph when given a new data set. However, they do not automatically expand by gathering data sets, so it is hard to deal with neologisms.

The existing knowledge graphs focus more on the study of linking missing values to complete graphs than on finding and linking neologisms. In order to expand the knowledge graph continuously to deal with neologisms, we have to detect neologisms. Most research to find neologisms is mainly done in the field of linguistics. However, most research detects only neologisms of new meaning. Although detecting neologisms both of new meaning and a new word, they only detect neologisms with new meaning. This is because it is a difficult challenge to figure out the new word from outliers (Abel & Stemle, 2018; Andrés Torres Rivera & Juan-Manuel Torres-Moreno, 2019; Yin & Cheng, 2016; Zhou, 2018).

4. ConceptNet

We use the ConceptNet (Havasi et al., 2009; Liu & Singh, 2004; Speer & Havasi, 2012; Speer et al., 2017) knowledge graph for our study. We expand the ConceptNet knowledge graph by applying our PolarisX automated mechanism for growing a knowledge

[3] https://developers.google.com/knowledge-graph/.
[4] https://en.wikipedia.org/wiki/Knowledge_Graph.

Table 4
Search result for ‘car’ in ConceptNet 5.5.

Relation	Terms
RelatedTo	drive, vehicle, motor, automobile, wheels, auto, automobile, seat, oil, road, ...
PartOf	a tire, a bumper, an engine, a horn, wheels, accelerator, air bag, auto accessory, automobile engine, automobile horn, ...
IsA	a volvo, Honda, an oldsmobile, a BMW, a motor vehicle, vehicle, Volkswagen, ambulance, baggage car, beach wagon, ...
HasA	seats, a seat, windows, an engine, headlights to increase visibility, 4 tires, at least one engine, an engine to power its wheels, a filter, four tires, ...

Table 5
Example relationships found in ConceptNet.

Relationship URI	Description	Examples
/r/RelatedTo	the positive relationship between A and B	car → drive
/r/IsA	A is a subtype or a specific instance of B; every A is a B	car → vehicle
/r/PartOf	A is a part of B	tire → car
/r/HasA	B belongs to A; often the reverse of PartOf	car → tire

Table 6
Number of words in ConceptNet for various languages.

Core languages		Common languages	
Language	Num. of words	Language	Num. of words
English	1803,873	Czech	129,183
French	3023,144	Filipino	17,620
Italian	1078,629	Korean	47,268
Japanese	363,663	Slovak	29,768
Chinese	242,746	Turkish	65,892

graph. ConceptNet builds a knowledge graph of relationships between words used by people to help computers understand common sense.

Table 4 is the result of searching for ‘car’ in ConceptNet 5.5 (Speer et al., 2017). ConceptNet stores the relationships between words, such as “A tire is a part of a car” (tire, *PartOf*, car). ConceptNet includes many relationships: *IsA, PartOf, HasA, FormOf, UsedFor, MadeOf*, and so on. Although there are more relationships and terms associated with ‘car’, we show only four relationships, and for each relationship, 10 terms in order of weight (Havasi et al., 2009; Liu & Singh, 2004; Speer & Havasi, 2012; Speer & Lowry-Duda, 2017; Speer et al., 2017).

Table 5 describes the relationships in ConceptNet that are used most often in our study. We use only 4 relations in the ConceptNet knowledge graph, but there are about 40 relations in it. To use ConceptNet API, we use a particular relationship URI. In the ‘Relationship URI’ column, the prefix is “conceptnet.io”.

Most knowledge graphs contain only English datasets, although some graphs support multiple languages. In particular, ConceptNet has a large amount of data for ten core languages, including English, French, and Japanese.

Table 6 shows the number of words in ConceptNet for five core languages and five non-core languages. We checked the number of words for 68 other common languages. As shown in Table 6, ConceptNet is dominated by English. The number of words for a non-core language is generally much smaller than that for a core language.

The core languages in ConceptNet enjoy another advantage over the non-core languages. ConceptNet provides an API for the core languages, but not for the non-core languages. This means it is easier to use ConceptNet for the core languages than for the non-core languages. To use ConceptNet for a non-core language, we have to download the dataset.

Applying knowledge graphs to various artificial intelligence models helps improve the performance of conversational AI models because computers can easily understand a person’s common sense. As mentioned earlier, however, the existing knowledge graphs have two major limitations. First, common languages other than core languages are difficult to use because the number of data is not enough. Second, the majority of knowledge graphs are difficult to respond to neologisms because they utilize data already collected or collected over a period of time.

5. Proposed method

We would like to propose PolarisX, an auto-extended knowledge graph that complements the limitations of the existing knowledge graphs. It refers to Polaris expander, and Polaris (Yoo et al., 2018) is our big data analysis and prediction framework. PolarisX is a part of Polaris project and used as a knowledge graph to help computers understand a person’s common sense. To complement the limitations of the existing knowledge graphs, we crawl social media especially Twitter and news data in real-time to expand data sources and graph. We extract new relationships using the fine-tuned BERT (Devlin et al., 2018) model and integrate them into the knowledge graph.

5.1. Motivation

The motivating example of PolarisX is described in Fig. 1. Fig. 1 illustrates when the word ‘selfie’ is input for a knowledge graph-based system such as search, question answering, and chatbot. The system on the left side is based on an existing knowledge graph where the new word ‘selfie’ does not exist. On the other hand, a system based on an auto-growing knowledge graph that can deal with neologisms, such as the one on the right side, can also retrieve meaningful knowledge about the word ‘selfie’. In this case, PolarisX could find meaningful knowledge which is used in a person’s common sense by crawling and expanding knowledge graph.

5.2. PolarisX: Auto-growing knowledge graph

Our proposed auto-growing knowledge graph, PolarisX, crawls social media, especially Twitter, and news data in real-time, in order to build a knowledge graph that could deal with neologisms and do not have language dependencies. We fine-tune the Google BERT model to extract new relationships from the crawled data and integrate the new relationships into the knowledge graph.

Fig. 2 shows the architecture of PolarisX. It consists of three major components: the social crawler to expand data resources, the semantic analyzer to determine new relationships, and the knowledge miner to build and extend the knowledge graph. Each major component, in turn, consists of key modules.

Many newly coined words and frequently used words are found in the news. We first use the news data to crawl the words.

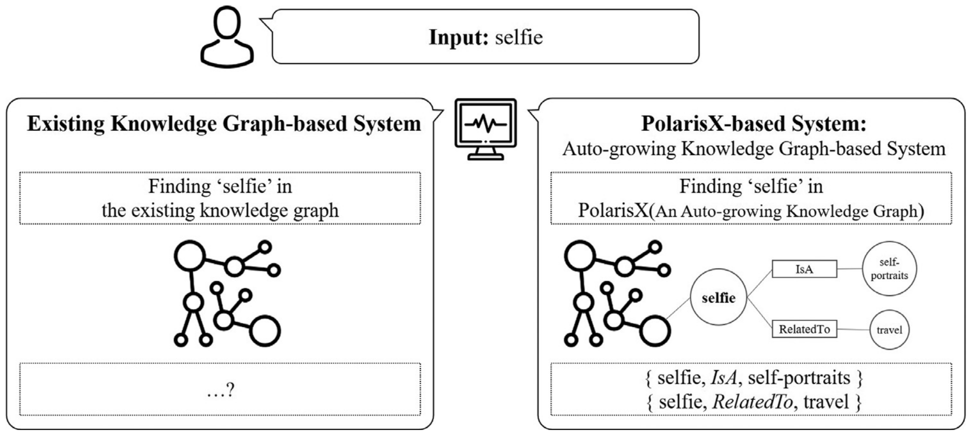

Fig. 1. Motivating example.

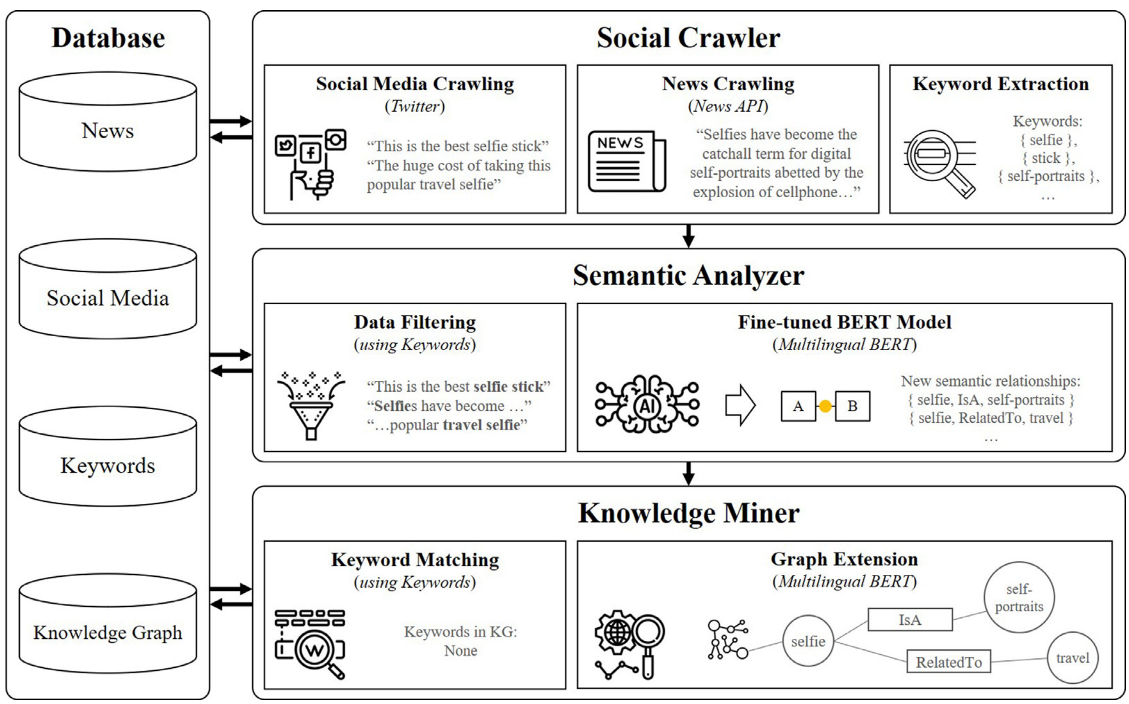

Fig. 2. PolarisX architecture.

Nowadays, many people use social media and create lots of data. So, we expand the dataset by crawling social media. We then build a topic set and figure out new semantic relationships using our fine-tuned BERT model. Then we expand the existing knowledge graph with the newly obtained semantic relationships.

5.2.1. Social crawler

Fig. 3 shows the architecture of the social crawler. The social crawler collects news data and social media data to expand the dataset.

First, we crawl social media, especially Twitter, and news data in real-time. Twitter and News are used to expand data sources on knowledge graphs. Twitter data is collected in real-time using Apache AsterixDB's FeedAdapter (Alkowaileet et al., 2018) function, while news data is collected through NewsAPI.[5]

The 'FeedAdapter' function of AsterixDB helps to collect Twitter easily in real-time. AsterixDB (Alsubaiee, Altowim, Altwaijry & Behm, 2014) is a big data management system to process big data such as social media quickly and deal with the data flexibly.

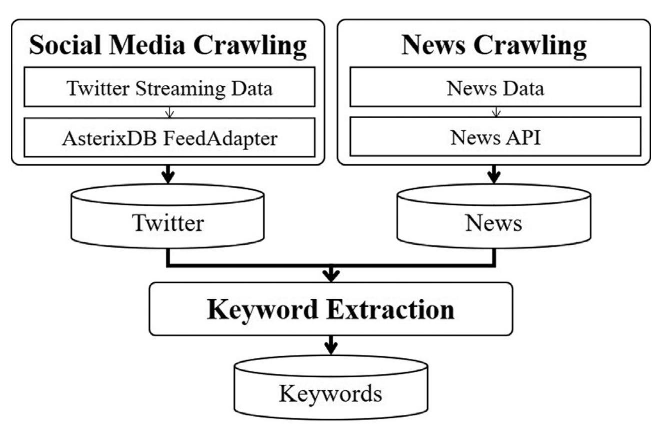

Fig. 3. Social Crawler architecture.

[5] https://newsapi.org/.

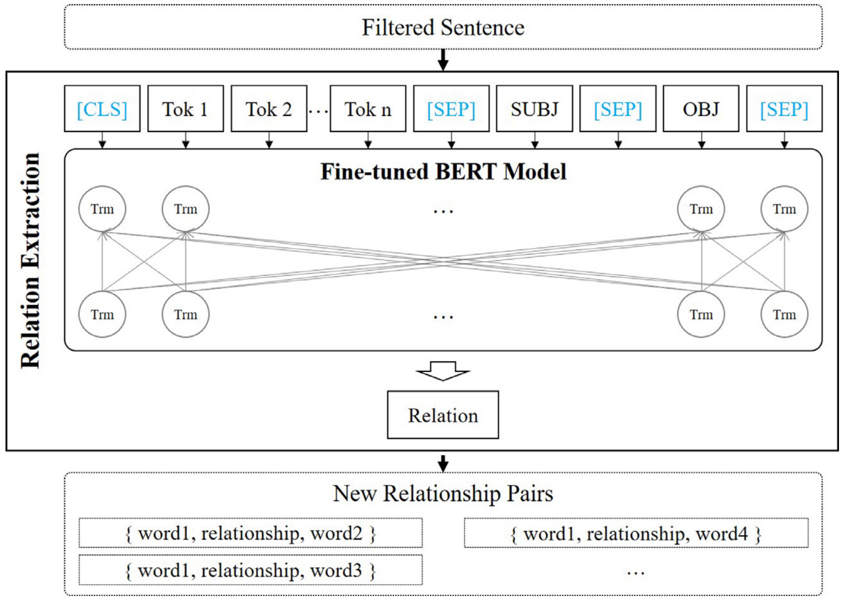

Fig. 4. Flow of semantic analyzer.

It supports the 'FeedAdapter' function for data ingestion, so it integrates data from external sources continuously. In AsterixDB, there is a built-in function for ingesting Twitter stream, called 'push_twitter'. To build a Twitter crawling adapter using the 'FeedAdapter' function, some AsterixDB data models must be defined using the query language SQL++. We also set up an application account with Twitter such as OAuth authentication credentials that include OAuth keys and tokens to access Twitter API.[6]

In order to extract keywords from news and Twitter data, we use LDA(Latent Dirichlet Allocation) (Blei, Ng & Jordan, 2003), the most popular topic modeling algorithm. LDA is a procedural probability distribution model that finds potentially meaningful topics in multiple documents. The topics are extracted by applying the LDA algorithm to each Twitter and news data.

5.2.2. Semantic analyzer

We use the BERT (Devlin et al., 2018) model to extract new relationships for the construction of the automatic expansion knowledge graph. BERT is a pre-trained language model released by Google that has taken up State-of-the-art in 11 tasks in the NLP field. BERT, which learns existing data in advance and is released as a general language model, can be fine-tuned using learning data according to the task you want to perform. We use BERT to tune models to extract relationships between keywords.

In particular, we use the pre-trained multilingual BERT[7] model called 'BERT-base, Multilingual Cased' which is supporting 104 languages. We use the relationship-based data, TACRED (Zhang, Zhong, Chen, Angeli & Manning, 2017) and fine-tune the model, so that the BERT model can be used for the automatically extended knowledge graph. TACRED data is a relation extraction data set made from news or web texts from the corpus of the TAC Knowledge Base Population Challenge. Examples cover 41 types of relationships and are labeled 'no_relation' if they are undefined.

To train the relation extraction model, it is required to figure out the two keywords and the relationship. TACRED dataset includes sentences, subjects, objects, and relationships. These are used to train our BERT-based relation extraction model. Fig. 4 shows the flow of semantic analyzer using the BERT-based relation extraction model. In the training stage, we use [SUBJ] and [OBJ], represented the subject and object respectively in TACRED dataset. We concatenate the tokenized sentence, [SUBJ], and [OBJ]. The concatenated sentence goes into the BERT model as an input. The final output, i.e. label, is a relation. This learned and verified relation extraction model is used to extract relationships in new sentences.

After the model is fine-tuned by training with TACRED dataset, we use the fine-tuned BERT-based relation extraction model to predict the relationship between the keywords. The sentences and keywords from the crawled Twitter and News data are concatenated and used as an input. Then we could extract the relation between two keywords.

We collect social media Twitter and news data in real-time to extract the most-reported keywords. The corresponding sentence with the extracted keywords on Twitter and News data goes as an input of the BERT-based relation extraction model. We extract the relationship between keywords and other words and then make a pair with the extracted entities and relationships such as {*keyword, relation, word*}. We extend the graph with the created pair by mapping it with the existing knowledge graph.

5.2.3. Knowledge miner

Expanding the existing knowledge graph is completed with the knowledge miner if the semantic graph is newly deployed. We extract a new knowledge relationship as a triple {*keyword, relationship, word*} through the semantic analyzer.

The knowledge miner connects the newly discovered relationship to the existing knowledge graph by using string matching. There are two ways to extend the knowledge graph. One is by connecting the matching node and the new semantic graph in the presence of a matched node. Another is by just adding a new node in the absence of a matched node.

Fig. 5 shows a part of the ConceptNet knowledge graph expanded with a matched node. With a newly discovered semantic relationship {sonata, *IsA*, car}, we find a matched relationship *IsA*. If there is a matched edge, we connect the newly extracted keyword 'sonata' to the '*IsA*' relationship.

The extraction of relationships even by using the BERT model can be done only through the dataset, and so it is largely influenced by the size of the dataset, type of label, and so on. To

[6] There is documentation how to ingest data with feed adapters in https://ci.apache.org/projects/asterixdb/feeds.html.

[7] https://github.com/google-research/bert.

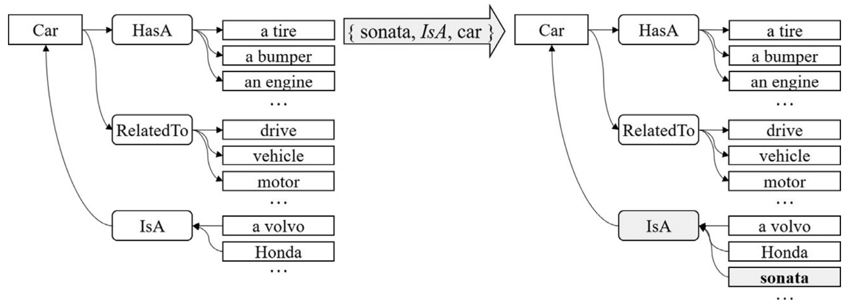

Fig. 5. Expanding the knowledge graph.

Table 7
Search result for ‘sonata’ in ConceptNet 5.5.

Relations	Terms
Synonyms	ソナタ*sonata*(ja), sonata(ca), 奏曲*sonata*(ja), sonata(eu), sonaatti(fi), sonate(de), 奏鳴曲*sonata*(ja),
Related terms	sonata(cs), composition(en), sonate(da), instrument(en), key(en), cantata(en), piano(en), minuet(en), sonatalike(en), tempo(en), sonatas(en), sonatina(en)
Types of sonata	piano sonata(en), sonatina(en), symphony(en)
sonata is a type of	classical music(en)
Parts of sonata	movement(en)
Derived terms	sonata form(en), sonatina(en)
Word forms	sonatas(en)
Context of this term	music(en)

Table 8
Search result for ‘sonata’ in Korean-translated ConceptNet 5.5.

Relations	Terms
Synonyms	소나타*sonata*(ja), 소나타*sonata*(ca), 연주곡*instrumental song*(ja), 소나타*sonata*(eu), 소나티*sonata*(fi), 주명*sonata*(ja), 소나타*sonata*(de)
Related terms	소나타*sonata*(cs), 구성*composition*(en), 소나타*sonata*(da), 악기*instrument*(en), 키*key*(en), 칸타타*cantata*(en), 피아노*piano*(en), 미뉴에트*minuet*(en), 소아 타 이케*soa ta ike*(en), 속도*tempo*(en), 소나타*sonata*(en), 소나 티나*sonatina*(en)
Types of sonata	피아노 소나타*piano sonata*(en), 소나 티나*sonatina*(en), 교향곡*symphony*(en)
sonata is a type of	고전 음악*classical music*(en)
Parts of sonata	운동*movement*(en)
Derived terms	소나타 형식*sonata form*(en), 소나 티나*sonatina*(en)
Word forms	소나타*sonata*(en)
Context of this term	음악*music*(en)

complete the knowledge graph, we could consider distant supervision (Augenstein, 2016) for future work, and we think it should be able to connect data that is not contained in the dataset.

6. Validation

Several experiments have been conducted to validate the proposed auto-growing knowledge graph, PolarisX. It collects one of the social media, Twitter and news data in real-time, and extracts new relationships using the fine-tuned BERT model. We propose a knowledge graph that links the extracted relationships to the graph and expand continuously. To verify our proposed method, we experiment in four ways.

6.1. Dealing with neologisms

To validate the effectiveness of our PolarisX system in automatically expanding a common sense knowledge base, we chose the ConceptNet knowledge graph for the Korean language. ConceptNet contains about 47,000 Korean words and does not support searches for neologisms.

For our validation study, we compared the search results for one word, “Sonata”, using ConceptNet 5.5, Korean-translated ConceptNet 5.5, and ConceptNet 5.5 expanded using PolarisX. The word ‘sonata’ is basically used as a term for music, but there is a car with the same name in South Korea.

We check the percentage of categories that have been searched for ‘sonata’. We used the search data ‘sonata’ for Sonata the car and sonata the music from Naver, a popular search portal in Korea. We compared the search frequencies for the two categories over the course of a year, from January 1, 2017, to December 31, 2017, using Naver Service API.[8]

The search results indicate that 76.58 percent of people searched for the Sonata car, and 31.37 percent for the music sonata. We compared the search results for ‘sonata’ using three versions of the ConceptNet knowledge graph: the unexpanded ConceptNet 5.5, the Korean-translated version of ConceptNet 5.5, and ConceptNet 5.5 expanded by using PolarisX.

Tables 7–9 are, respectively, the result of using ConceptNet 5.5, the Korean-translated ConceptNet, and the expanded ConceptNet using PolarisX for ‘sonata’. ConceptNet 5.5 shows several relationship words for ‘sonata’. However, most of the terms are music-related, so it is impossible to get any information about Sonata the car.

Table 8 is the result of using Korean-translated ConceptNet 5.5. We translated English terms found in ConceptNet for the Korean language into Korean using the Google Translation API. We had thought that when the English terms are translated into Korean, perhaps we would see better results. The Korean-translated Con-

[8] https://developers.naver.com/products/datalab/.

Table 9
Search result for 'sonata' in ConceptNet expanded using PolarisX.

Relations	Terms
Synonyms	ソナタ*sonata*(ja), sonata(ca), 奏曲*sonata*(ja), sonata(eu), sonaatti(fi), 奏鳴曲*sonata*(ja), sonate(de)
Related terms	자동차*car*(ko), 현대*Hyundai*(ko), car(en), sonata(cs), composition(en), sonate(da), instrument(en), key(en), cantata(en), piano(en), minuet(en), sonatalike(en), tempo(en), sonatas(en), sonatina(en)
Types of sonata	쏘나타*sonata*(ko), piano sonata(en), sonatina(en), symphony(en)
sonata is a type of	중형차*Midsize-car*(ko), 자동차*car*(ko), 현대*Hyundai*(ko), car(en), classical music (en)
Parts of sonata	타이어*tire*(ko), 시트*seat*(ko), 엔진*engine*(ko), movement(en)
Derived terms	sonata form(en), sonatina(en)
Word forms	sonatas(en)
Context of this term	music(en)

Table 10
Comparison between ConceptNet 5.5 and PolarisX-expanded ConceptNet 5.5.

Comparison		Comparison 1 (All)		Comparison 2 (Top 5)	
Related category		Music	Car	Music	Car
ConceptNet 5.5	**Ratio (%)**	100	0	100	0
	Num. of terms	29	0	19	0
PolarisX-Expanded	**Ratio (%)**	72.50	**27.50**	59.26	**40.74**
ConceptNet	**Num. of terms**	29	11	16	11

ceptNet did not yield Sonata the car. To make matters worse, the translation was at times incorrect.

Table 9 is the result of using ConceptNet 5.5 expanded with PolarisX. We were able to get a lot of information about Sonata the car.

Table 10 compares the results of using ConceptNet 5.5 and ConceptNet expanded by applying our PolarisX system. We compared the ratio of the resulting terms by the category car and music. Comparison 1 is the result of considering all words, and Comparison 2 is the result of considering only the top five words.

Whereas the unexpanded ConceptNet 5.5 only yields the terms for the music sonata, ConceptNet 5.5 expanded using PolarisX yields the terms for Sonata the car as well as the music sonata. However, we caution that the results do not apply to all ConceptNet relationships, because for our study we have primarily used the relationships *IsA, HasA, RelatedTo*, and *PartOf.*

6.2. Extension of knowledge graph

The PolarisX implements a dynamic graph, which continuously expands. We expand the existing knowledge graph by adding new relationships. We have conducted experiments using existing knowledge graphs to verify the expandability of PolarisX. We tried to show how much the knowledge graph can be expanded during a specific period of time in order to validate its performance.

The PolarisX knowledge graph is an add-on to the ConceptNet 5.5 knowledge graph, which already has a large number of concepts. We chose to compare PolarisX (+ ConceptNet 5.5) against such open source-based dynamic knowledge graphs as DBpedia (Auer et al., 2007), YAGO (Suchanek et al., 2007), NELL (Carlson et al., 2010), OpenCyc (Lenat, 1995), and Probase (Wu, Li, Wang & Zhu, 2012).

Table 11 shows the result of comparison among the existing knowledge graphs. It shows the number of relations and the number of edges. We note that the definition of relationships is slightly different between the knowledge graphs.

Probase has only one type of relationship, namely the *IsA* relationship. It refers to two kinds of *IsA* relationship: the concept-subconcept relationship and the concept-instance relationship. ConceptNet has 40 types of relationships. We have adopted the experimental method proposed by Wu et al. (2012). The graph edge represents data and relationships between the data. A higher total number of edges roughly means a richer knowledge base.

Table 11
Comparison result with the existing knowledge graphs.

Knowledge graphs	# of Relations	# of Edges
DBpedia (English)	2813	176,043,129
YAGO	77	25,946,870
NELL	425	432,845
OpenCyc	18,526	2413,894
Probase	1	20,757,545
PolarisX (+ ConceptNet)	**40**	**32,871,573**

Table 11 shows two things. One is that ConceptNet 5.5 is indeed a large knowledge graph. Another is the difference between the total number of edges in ConceptNet and in PolarisX + ConceptNet. The difference quantitatively shows the expandability of PolarisX.

To measure the expandability of PolarisX, we have selected the period from August 8, 2018, to August 14, 2018. During this period, we collected about 35,000 news articles and about 2.5 million tweets. Using ConceptNet 5.5 as the baseline with 32,755,210 edges, PolarisX expanded the total number of edges to 32,871,573. This means that during a six-day period, PolarisX has automatically expanded the ConceptNet 5.5 knowledge graph with 116,363 new graph edges, which Probase calls knowledge pairs.

6.3. Novelty of the extended knowledge graph

We also experiment to verify the novelty and accuracy of our approach of PolarisX. COMET (Bosselut et al., 2019) and ATOMIC (Sap et al., 2018) are methods to connect graphs based on events, and we apply the experimental methods for the novelty and accuracy of this research. We perform the experiment by using the training and test set provided by ATOMIC dataset.[9] We use metric of BLEU score($n=2$) to verify accuracy. BLEU score (Papineni, Roukos, Ward & Zhu, 2001) is usually used for evaluating the quality of text. There are many variations, but we use BLEU-2 score, which is using bi-grams. We also evaluate the novelty of the expanded knowledge graph using N/T*sro*, the proportion of the newly generated tuples, and N/T*o*, the proportion of the newly added tuples that have a novel object.

ConceptNet is our baseline, but it is very difficult to compare exactly by using the same dataset and metrics. So we evaluate the

[9] https://homes.cs.washington.edu/~msap/atomic/.

Table 12
Comparison on TACRED dataset.

Models	Precision	Recall	F1 score
Logistic Regression (Zhang et al., 2018)	73.5	49.9	59.4
PA-LSTM (Zhang et al., 2017)	65.7	64.5	65.1
C-GCN+PA-LSTM (Zhang et al., 2018)	71.3	65.4	68.2
BERT-based model (our model)	**79.1**	**72.6**	**75.7**

quality and the novelty of our proposed knowledge graph, PolarisX. For PolarisX, we get 9.79 as the BLEU-2 score. Also, N/T*sro* is 100% and N/T*o* is 5.82%. N/T*sro* is 100% because all the training set is not used in the test set. With this result, we could argue that our proposed PolarisX shows fairly quality and novelty as growing automatically.

6.4. Accuracy of semantic analyzer

In PolarisX, a semantic analyzer using a BERT (Devlin et al., 2018) model to extract new relation information plays a key role. The accuracy of the relation extraction model has a significant effect on the performance of the auto growing knowledge graph. This is because the more accurate the relation extraction model is, the more reliable the newly connected relationship knowledge. Therefore, we experiment with the accuracy of the fine-tuned BERT model and verify its performance by comparing it with other existing models.

We use a relation extraction model that has trained BERT for PolarisX. We build and experiment utilizing the TPU environment on Google colab.[10] BERT model has different models depending on case-sensitive, number of layers, number of hidden units, and parameters. We use the ‘BERT-Base, Multilingual Cased’ model, which supports 104 languages and consists of 12 layers, 768 hidden units and 12 heads, to enable to respond to various languages.

To compare the experiment of relation extraction models, we use a traditional technique based model and deep learning-based models. We compare with the logistic regression model of traditional technique. We also compare with deep learning-based models, PA-LSTM and C-GCN+PA-LSTM. PA-LSTM (Zhang et al., 2017) model combines the LSTM(Long Short-Term Memory) model with a form of position-aware attention. C-GCN+PA-LSTM (Zhang, Qi & Manning, 2018) model applies graph convolutional network over pruned dependency tree to PA-LSTM model.

Table 12 shows the result of a comparative experiment with existing models using TACRED dataset. The results using the Logistic Regression are reported in Zhang et al. (2018). Compared with the traditional technique such as logistic regression and deep learning technique such as PA-LSTM and C-GCN+PA-LSTM models, the BERT-based model we proposed shows better performance as 75.7 in F1 score.

Relation extraction is a complex and difficult area in the NLP(Natural Language Processing) field. The model we proposed does not show very high performance but it shows meaningful results compared to existing models using the same data. The accuracy of the relation extraction model is a very important indicator in the knowledge graph because it affects the performance of the proposed PolarisX as a whole.

To verify the proposed PolarisX, new data were collected and we experimented with the newly added amount and accuracy of the knowledge graph. The knowledge graph, which automatically expands by crawling social media and news data in real-time and extracting new relationships, contains more information and shows higher accuracy than the existing knowledge graphs. Then we verify the feasibility of the proposed method.

[10] https://colab.research.google.com/.

7. Conclusion

Knowledge graphs have proven to be effective in helping computers understand human languages and reason. They represent human knowledge in the form of a graph, where the nodes correspond to words and edges correspond to relationships among words. Knowledge graphs, however, have some limitations. The knowledge base is not large except for a small number of core languages. Further, the knowledge base usually does not support searches for neologisms.

In this paper, we described the architecture of the PolarisX that we have designed and implemented. The PolarisX automates the expansion of a common sense knowledge base by crawling news sites and social media to collect neologisms, building new semantic subgraphs, and appending them to the existing knowledge graph.

We use the pre-trained language model, BERT, to extract new relationships. BERT model shows fairly good performance for many NLP tasks, but there is a limit, of course. We fine-tuned the BERT model using the labeled dataset, so the relation extraction model only could extract the relationships in the labels of the dataset.

We believe that PolarisX can help expand the existing common sense knowledge base with neologisms and established words alike. Further, PolarisX can help expand the existing knowledge base for both core languages and non-core languages. For future work, we could build web service as an open source to use PolarisX. Also, we will improve PolarisX to a stand-alone system that enables to show graph expansion for neologisms that users want to know.

PolarisX can be used in various fields, especially on conversational AI. With the auto growing knowledge graph, the computer could understand a person’s common sense and then make a more complicated conversation with the person. If we use PolarisX as the base knowledge data in a chatbot, the chatbot with PolarisX could communicate with a better understanding of the user’s intentions.

Also, it is possible to analyze relationships and sentiments. We could figure out the relationship between two entities using PolarisX, and then connect two entities with the extracted relationship. If we utilize a sentiment dictionary or deep learning model on it, we could find sentiments and relationships between two entities.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests

Credit authorship contribution statement

SoYeop Yoo: Conceptualization, Methodology, Software, Validation, Investigation, Data curation, Writing - original draft, Visualization. **OkRan Jeong:** Writing - review & editing, Supervision, Project administration.

Acknowledgments

This research was supported by Basic Science Research Program through the NRF (National Research Foundation of Korea), and the MSIT (Ministry of Science and ICT), Korea, under the National Program for Excellence in SW supervised by the IITP (Institute for Information & communications Technology Promotion), and Copyright Technology R&D through the MCST (Ministry of Culture, Sports and Tourism), Korea and the KCC (Korea Copyright Commission) (Nos. NRF 2019R1A2C1008412, 2015-0-00932, 2019-CONTEXT-9500).

We would like to sincerely thank the reviewers and Professor Won Kim of the Department of Software at Gachon University for advising us in preparing this paper.

References

Abdul-Kader, S. A., & Wood, J. (2015). Survey on chatbot design techniques in speech conversation systems. *International Journal of Advanced Computer Science and Applications, 6*(7). doi:10.14569/ijacsa.2015.060712.

Abel, A., & Stemle, E. W. (2018). On the detection of neologism candidates as a basis for language observation and lexicographic endeavors: The STyrLogism project. In *Proceedings of the XVIII EURALEX international congress: Lexicography in global contexts* (pp. 535–544). Retrieved from. http://euralex.org/publications/on-the-detection-of-neologism-candidates-as-a-basis-for-language-observation-and-lexicographic-endeavors-the-styrlogism-project/.

Alkowaileet, W., Alsubaiee, S., Carey, M. J., Li, C., Ramampiaro, H., Sinthong, P., et al. (2018). End-to-end machine learning with apache asterixdb. In *Proceedings of the second workshop on data management for end-to-end machine learning - DEEM* (pp. 1–10). ACM Press. doi:10.1145/3209889.3209894.

Alsubaiee, S., Altowim, Y., Altwaijry, H., & Behm, A. (2014). AsterixDB: A scalable, open source BDMS. *Proceedings of the VLDB Endowment, 7*(14), 1905–1916. doi:10.14778/2733085.2733096.

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). *DBpedia: A nucleus for a web of open data* (pp. 722–735). BerlinHeidelberg: Springer. doi:10.1007/978-3-540-76298-0_52.

Augenstein, I. (2016). *Web relation extraction with distant supervision*. University of Sheffield Retrieved from http://etheses.whiterose.ac.uk/13247/ .

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022. (Jan)Retrieved from http://www.jmlr.org/papers/v3/blei03a.html .

Bollacker, K., Evans, C., Paritosh, P., Sturge, T., & Taylor, J. (2008). Freebase. In *Proceedings of the 2008 ACM SIGMOD international conference on management of data - SIGMOD* (pp. 1247–1250). ACM Press. doi:10.1145/1376616.1376746.

Bosselut, A., Rashkin, H., Sap, M., Malaviya, C., Celikyilmaz, A., & Choi, Y. (2019). COMET: Commonsense transformers for knowledge graph construction. *Association for Computational Linguistics (ACL)*. Retrieved from https://par.nsf.gov/biblio/10107989.

Bruns, A. (2008). *Blogs, wikipedia, second life, and beyond: From production to produsage*. IOS Press.

Carlson, A., Betteridge, J., Wang, R. C., Hruschka, E. R., & Mitchell, T. M. (2010). Coupled semi-supervised learning for information extraction. In *Proceedings of the third ACM international conference on web search and data mining - WSDM* (pp. 101–110). ACM Press. doi:10.1145/1718487.1718501.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. Retrieved from http://arxiv.org/abs/1810.04805.

Gao, J., Galley, M., & Li, L. (2019). Neural approaches to conversational AI. *Foundations and Trends® in Information Retrieval, 13*(2–3), 127–298. doi:10.1561/1500000074.

Guarino, N., Guarino, N., & Giaretta, P. (1995). Ontologies and knowledge bases: Towards a terminological clarification. In *Towards Very Large Knowledge Bases: Knowledge Building And Knowledge Sharing* (pp. 25–32). Retrieved from. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.320.8006.

Havasi, C., Speer, R., & Alonso, J. (2009). Conceptnet: A lexical resource for common sense knowledge. In N. Nicolov, G. Angelova, & R. Mitkov (Eds.), *Recent advances in natural language processing V: Selected papers from RANLP 2007* (pp. 269–280). John Benjamins Pub. Co.

Lenat, D. B. (1995). CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM, 38*(11), 33–38. doi:10.1145/219717.219745.

Li, X., Taheri, A., Tu, L., & Gimpel, K. (2016). Commonsense knowledge base completion. In *Proceedings of the 54th annual meeting of the association for computational linguistics (Volume 1: Long Papers)* (pp. 1445–1455). Association for Computational Linguistics. doi:10.18653/v1/P16-1137.

Lisa, E., & Wolfram, W. (2016). Towards a definition of knowledge graphs. In *Joint Proceedings of the posters and demos track of 12th international conference on semantic systems - SEMANTiCS and 1st international workshop on semantic change & evolving semantics (SuCCESS)* Retrieved from https://www.researchgate.net/publication/323316736_Towards_a_Definition_of_Knowledge_Graphs.

Liu, H., & Singh, P. (2004). ConceptNet — A practical commonsense reasoning tool-kit. *BT Technology Journal, 22*(4), 211–226. doi:10.1023/B:BTTJ.0000047600.45421.6d.

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM, 38*(11), 39–41. doi:10.1145/219717.219748.

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2001). BLEU. In *Proceedings of the 40th annual meeting on association for computational linguistics - ACL* (p. 311). Association for Computational Linguistics. doi:10.3115/1073083.1073135.

Paulheim, H. (2016). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web, 8*(3), 489–508. doi:10.3233/SW-160218.

Pujara, J., Miao, H., Getoor, L., & Cohen, W. (2013). *Knowledge graph identification* (pp. 542–557). BerlinHeidelberg: Springer. doi:10.1007/978-3-642-41335-3_34.

Rivera, A. T., & Torres-Moreno, J. M. (2019). Detecting new word meanings: A comparison of word embedding models in Spanish. In *Proceedings of the conférence en recherche d'informations et applications - CORIA,16th French information retrieval conference*. doi:10.24348/coria.2019.CORIA_2019_paper_15.

Sap, M., Bras, R. Le, Allaway, E., Bhagavatula, C., Lourie, N., Rashkin, H., et al. (2018). ATOMIC: An atlas of machine commonsense for if-then reasoning. In *Proceedings of the thirty-second AAAI conference on artificial intelligence (AAAI)* Retrieved from https://www.semanticscholar.org/paper/ATOMIC%3A-An-Atlas-of-Machine-Commonsense-for-If-Then-Sap-Bras/8209a8703d8c48aaca1523cfa307dd1c069e58f3.

Singh, P., Lin, T., Mueller, E. T., Lim, G., Perkins, T., & Li Zhu, W. (2002). *Open mind common sense: Knowledge acquisition from the general public* (pp. 1223–1237). Berlin, Heidelberg: Springer. doi:10.1007/3-540-36124-3_77.

Singhal, A. (2012). Introducing the knowledge graph: Things, not strings. Official Google Blog.

Sowa, J. F. (2006). Semantic networks. *Encyclopedia of cognitive science*. Chichester: John Wiley & Sons, Ltd. doi:10.1002/0470018860.s00065.

Speer, R., Chin, J., & Havasi, C. (2017). ConceptNet 5.5: An open multilingual graph of general knowledge - semantic scholar. In *Proceedings of the AAAI* (pp. 4444–4451). Retrieved from. https://www.semanticscholar.org/paper/ConceptNet-5.5%3A-An-Open-Multilingual-Graph-of-Speer-Chin/26aa6fe2028b5eefbaa40ab54ef725bbbe7d9810.

Speer, R., & Havasi, C. (2012). Representing general relational knowledge in conceptnet 5. In *Proceedings of the eighth international conference on language resources and evaluation (LREC)* (pp. 3679–3686). Retrieved from. https://aclanthology.info/papers/L12-1639/l12-1639.

Speer, R., & Lowry-Duda, J. (2017). ConceptNet at semeval-2017 task 2: Extending word embeddings with multilingual relational knowledge. Retrieved from http://arxiv.org/abs/1704.03560

Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering, 25*(1–2), 161–197. doi:10.1016/S0169-023X(97)00056-6.

Suchanek, F. M., Kasneci, G., & Weikum, G. (2007). Yago: A core of semantic knowledge. In *Proceedings of the 16th international conference on world wide web - WWW* (pp. 697–706). ACM Press. doi:10.1145/1242572.1242667.

Tandon, N., de Melo, G., Suchanek, F., & Weikum, G. (2014). WebChild. In *Proceedings of the 7th ACM international conference on web search and data mining - WSDM* (pp. 523–532). ACM Press. doi:10.1145/2556195.2556245.

Trivedi, R., Dai, H., Wang, Y., & Song, L. (2017). Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In *Proceedings of the 34th international conference on machine learning* (pp. 3462–3471). Retrieved from. http://proceedings.mlr.press/v70/trivedi17a.html.

Vrandečić, D., & Denny (2012). Wikidata. In *Proceedings of the 21st international conference companion on world wide web - WWW Companion* (p. 1063). ACM Press. doi:10.1145/2187980.2188242.

Vrandečić, D., & Krötzsch, M. (2014). Wikidata: A free collaborative knowledgebase. *Communications of the ACM, 57*(10), 78–85. doi:10.1145/2629489.

Wang, Z., Li, J., Wang, Z., Li, S., Li, M., Zhang, D., et al. (2013). XLore: A large-scale English–Chinese bilingual knowledge graph. In *Proceedings of the international semantic web conference* (pp. 121–124).

Wu, W., Li, H., Wang, H., & Zhu, K. Q. (2012). Probase. In *Proceedings of the international conference on management of data - SIGMOD* (p. 481). ACM Press. doi:10.1145/2213836.2213891.

Yan, J., Wang, C., Cheng, W., Gao, M., & Zhou, A. (2018). A retrospective of knowledge graphs. *Frontiers of Computer Science, 12*(1), 55–74. doi:10.1007/s11704-016-5228-9.

Yin, L., & Cheng, F. (2016). Neologisms detection in a overlapping topical complex network. In *Proceedings of the 12th international conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD)* (pp. 830–834). IEEE. doi:10.1109/FSKD.2016.7603284.

Yoo, S., Song, J., & Jeong, O. (2018). Social media contents based sentiment analysis and prediction system. *Expert Systems with Applications, 105*, 102–111. doi:10.1016/j.eswa.2018.03.055.

Yu, L. (2011). Linked open data. In *A developer's guide to the semantic web* (pp. 409–466). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-642-15970-1_11.

Zhang, L. (2002). *Knowledge graph theory and structural parsing*. Twente University Press (TUP) Retrieved from https://research.utwente.nl/en/publications/knowledge-graph-theory-and-structural-parsing.

Zhang, Y., Qi, P., & Manning, C. D. (2018). Graph convolution over pruned dependency trees improves relation extraction. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 2205–2215). Association for Computational Linguistics. doi:10.18653/v1/D18-1244.

Zhang, Y., Zhong, V., Chen, D., Angeli, G., & Manning, C. D. (2017). Position-aware attention and supervised data improve slot filling. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 35–45). Association for Computational Linguistics. doi:10.18653/v1/D17-1004.

Zhou, J. (2018). Internet newborn word recognition method under conditional random field model. In *Proceedings of the international conference on virtual reality and intelligent systems (ICVRIS)* (pp. 532–535). IEEE. doi:10.1109/ICVRIS.2018.00136.