INTERNATIONAL DESIGN CONFERENCE – DESIGN 2022
https://doi.org/10.1017/pds.2022.84

Patent-KG: Patent Knowledge Graph Extraction for Engineering Design

H. Zuo ✉, Y. Yin and P. Childs

Imperial College London, United Kingdom
✉ hz2019@ic.ac.uk

Abstract

This paper builds a patent-based knowledge graph, patent-KG, to represent the knowledge facts in patents for engineering design. The arising patent-KG approach proposes a new unsupervised mechanism to extract knowledge facts in a patent, by searching the attention graph in language models. The extracted entities are compared with other benchmarks in the criteria of recall rate. The result reaches the highest 0.8 recall rate in the standard list of mechanical engineering related technical terms, which means the highest coverage of engineering words.

Keywords: knowledge representations, artificial intelligence (AI), data-driven design

1. Introduction

In 2019, more than 3 million patents and 1.5 million scientific papers (World Intellectual Property Organization, 2019) were published worldwide. This can be contrasted with a person's reading capacity of 264 papers per year (Van Noorden, 2014) on average. The ever-growing quantity of data provides both considerable challenges and opportunities for designers. On one hand, it is hardly possible for an individual to fully search and comprehend a specific domain, as the published data grow every day. On the other hand, (Swanson, 1986) hypothesized that a scientific discovery can be established by systematically studying existing knowledge. The vast amount of data are of high diversity, and can be reused as incentives and stimuli for new knowledge. Reusing existing knowledge to speed up the idea generation has already been used in the domain of design(Shi et al., 2017, Sarica et al., 2020, McCaffrey and Spector, 2018, Fu et al., 2013).

A new concept - knowledge graph - is introduced to represent the knowledge in data in a new format. A knowledge graph is defined (Wang et al., 2017) as "a multi-relational graph composed of entities and relations which are regarded as nodes and different types of edges, respectively" to represent the knowledge. A knowledge graph expresses knowledge in the format of a triple, which includes a head, a relationship and a tail. This single piece of information can be regarded as a quantum of knowledge and knowledge graph is the accumulation of a great number of knowledge facts. For example, consider Figure 1 illustrating the knowledge fact: Albert Einstein was born in German Empire. The knowledge triple is representing it as: (Albert Einstein, BornIn, German Empire) and it is transformed into the images with two nodes as the head and tail, one edge as the relationship. Many acknowledged facts can be composed together in a knowledge graph. To build the knowledge graph, it is fundamental to identify and extract the knowledge in the text. However, existing rule-based and supervised machine learning based methods are both subject to human based rules and annotations, which is less likely to generalize to a larger size data.

To address the challenge highlighted, this paper aims to propose an unsupervised method to extract knowledge facts in patents and facilitate knowledge reuse for engineering design, with potential applications to other domains.

(Albert Einstein, **BornIn**, German Empire)
(Albert Einstein, **HusbandOf**, Mileva Marić)
(Lieserl Einstein, **DaughterOf**, Albert Einstein)
(Albert Einstein, **SonOf**, Hermann Einstein)
(Albert Einstein, **GraduateFrom**, University Of Zurich)
(Albert Einstein, **WinnerOf**, Nobel Prize in Physics)
(Albert Einstein, **ExpertIn**, Physics)
(Nobel Prize in Physics, **AwardIn**, Physics)
(The theory of relativity, **TheoryOf**, Physics)
(Albert Einstein, **SupervisedBy**, Alfred Kleiner)
(Alfred Kleiner, **ProfessorOf**, University of Zurich)
(The theory of relativity, **ProposedBy**, Albert Einstein)
(Hans Albert Einstein, **SonOf**, Albert Einstein)

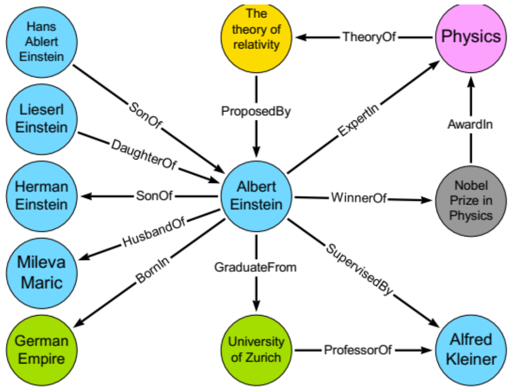


Figure 1. An example of a knowledge graph

2. Related work

2.1. Network for engineering design

When considering knowledge reuse, common knowledge sources include encyclopedias, patent documents, scientific literature, and reports. The growing need for engineering design tools and insights has driven researchers to construct large databases based on these sources. Shi (2018) proposed a pipeline to crawl design knowledge from design posts and Elsevier published scientific literature, and constructed a structured ontology network to provide a semantic level understanding of the design knowledge. This semantic level representation of understanding was applied for design information retrieval and insight for idea generation process. Sarica et al. (2020) trained a semantic engineering knowledge graph from patent data, to overcome the limited coverage of traditional keyword retrieval. A case study demonstrated that the proposed method improves the efficiency to assist the early stage of design work. Chen et al. (2013) proposed an information system in Wikipedia to extract core information inside the selected articles. These pieces of extracted information were analysed in pairs and frequency to gain insight into relationships, and thus to support the conceptual design stage. Siddharth et al. (2021) used syntactic rules to extract knowledge triple (head, relationship, tail) from patent data to build an Engineering knowledge graph. Listing these rules is laborious and these rules cannot be listed comprehensively, which will cause incomplete extraction. Our goal is to extract the head, tail and relationship without these rules and achieve better results. Patents were chosen as the data source as this form of documents records the content of inventions and contains a large quantity of scientific and technological information (Aristodemou and Tietze, 2018).

2.2. Information Extraction

Information extraction is a task to identify and recognize words or phrases in text. This task is a fundamental activity in knowledge graph construction and various methods have been proposed. Normal methods include rule-based information extraction, supervised information extraction and unsupervised information extraction (Li et al., 2020).

2.2.1. Rule-based information extraction

Rule-based information extraction (Li et al., 2020) extracts entities based on predefined rules. After analyzing the characteristics of entities, artificial rules of the intended information, such as syntactic rules or POS tagging, need to be constructed to match and identify the entity in the text. For example, to extract disease name such as type 1 diabetes, type 2 diabetes and type 3 diabetes, the rule is 'type + number + diabetes'. The limitation is evident: the rule can never be listed and formulated comprehensively by human resources.

https://doi.org/10.1017/pds.2022.84 Published online by Cambridge University Press

2.2.2. Supervised and unsupervised information extraction

Supervised machine learning with labelled data is expensive in terms of human labour, thus making it difficult to apply for large-scale information extraction. So, weak supervised learning or unsupervised learning is proposed. For example, distant supervised learning (Mintz et al., 2009) will apply similar labelled data text to the target text, thus avoiding the tagging task. TextRunner (Etzioni et al., 2008), Reverb (Fader et al., 2011), and Ollie (Schmitz et al., 2012) are examples of three unsupervised learning mechanisms. These three methods transform knowledge facts in plain text into triples without predefined classification, which is also called open information extraction. However, the quality of unsupervised approaches normally cannot compare with supervised learning because noise will be introduced as there is no label data as a filter. In Natural language processing, language models(LM) such as BERT and GPT-2 (Radford et al., 2019) have demonstrated significant capability in some related tasks, such as: sentence classification (Wang et al., 2018). In the structure of LMs, multi-head attentions are used that mimics cognitive attention, which will enhance the importance of the aiming part from the input data and fade others. Wang et al. (2020) further applies the attention mechanism in language models for information extraction inside a sentence.

The aim in this study is to propose an unsupervised method to build a patent-based knowledge graph. Inspired by the attention mechanism in language models, we aim to apply the attention mechanism to extract the relationships between technical terms, thus

3. Construction of patent-KG

3.1. Data source

Patent data, for this study was gathered from 2016 to 2021, with Cooperative Patent Classification (CPC) codes start in ‘F’(European Patent Office, 2021) – referring to “Mechanical engineering; lighting; heating; engines or pumps”. In total, 457,815 patents were retrieved after filtering.

3.2. Data pre-processing

The purpose of this step is to process and prepare each sentence of the abstract into a list of tokens which are applied as heads and tails.

First the abstract will be split into sentences with spaCy (Honnibal Matthew, 2017). Note that in spaCy, the default sentence segmentation will only split a sentence on punctuation such as ‘.’, ‘!’ or ‘?’ applying a general language rule. In patent text, a period tends to be used in case of an abbreviation or to signify the end of one sentence or claim. Normally the sentence in a patent can be very long, the semicolon ‘;’ is used frequently allowing incorporation of multiple discrete parts for a sentence. Therefore the ‘;’ token is added as a sentence boundary and overwritten in spaCy. In this way, the splitting can avoid long sentences and can be more accurate.

Second, the sentence will be split into tokens, and then noun phrases will be recognized and combined. Patents contain a large number of technical terms, some of which are brand new technical terms formed in the patent. There are three ways of technical terms formation in patents normally (Andersson et al., 2016):

(1) orthographical unit, e.g. bookcase, airplane, curveball. These words can be recognized by spaCy normally via its POS tagging.

(2) multi-word unit (MWU), e.g. airplane wings, knowledge graph, natural language processing. These words can be recognized by spaCy normally via its POS tagging.

(3) combined with hyphenation (e.g. H-theorem, mother-in-law). The default tokenizer in spaCy will split on hyphens. To avoid this, the existing infix definition is overwritten and a regular expression that treats a hyphen between letters as an infix is added.

Third, the sentence will be split into tokens, and then phrasal verbs will be recognized and combined. A phrasal verb is the combination of a verb and a particle, such as an adverb or a preposition, e.g. relate to, positioned through, engageable with. The verb contains the action information while the particle contains the additional information. Both can be part of the relationship, so it is also recognized and combined after tokenization.

https://doi.org/10.1017/pds.2022.84 Published online by Cambridge University Press

3.3. Patent-KG construction

3.3.1. Dependency patterns

After the data pre-processing stage, the technical terms as nouns are recognized as heads and tails, the next step is to find and extract the relationships between them. Different from SAO structure (Cascini et al., 2004) which only extract verbs to represent the relationships between technical terms, the patent texts are analyzed and the four interested dependency patterns (Marie-Catherine de Marneffe, 2016) are listed and corresponding examples are as follows:

(1) dobj: direct object

The direct object of a verb phrase is the noun phrase which is the (accusative) object of the verb.

"The sensor sends a signal" dobj(sends, signal)

"The electricity light the bulb" dobj (light, bulb)

(2) nsubj: nominal subject

A nominal subject is a noun phrase which is the syntactic subject of a clause. There are two scenarios, the governor of this relation is a verb or a copular verb. When the relation is a copular word, the root of the clause is the complement of the copular verb, which can be an adjective. However, the scenario of an adjective is not considered in patent patterns because it cannot form a knowledge fact.

"The baby is cute" nsubj (cute, baby) * not considered as knowledge facts*

"She left him a note" nsubj(left, she)

(3) cop: coupla

A coupla is the relation of a function word used to link a subject to a nonverbal predicate, including the expression of identity predication.

"HTR is the fourth generation nuclear power station" cop (station, is)

"Bill is a good person" cop (person, is)

(4) pobj: object of a preposition

The object of a preposition is the head of a noun phrase following the preposition.

"The colls in the axial magnetic bearings" pobj (in, axial magnetic bearing)

"Place the card inside the slot" pobj (inside, slot)

3.3.2. Match

After defining the relationships between , the next step is to find the relationships (mostly verbs) between them. The technical terms are grouped in two as a head and tail pair (h, t), then the match stage will find the best relationship between the (h, t) pair to generate a knowledge triple (h, r, t) by searching the attention matrix. The attention matrix is formed within transformer-based language model BERT(Bidirectional Encoder Representations from Transformers).

BERT (Devlin et al., 2018) is a large and pre-trained Transformer network, with 12 layers where each layer consists of 12 attention heads. Fig 2 illustrates how the attention score is computed by selecting only one attention head in the first layer.

Firstly, the input tokens (x1,x2,x3,x4) are transformed into a sequence of vectors [a1,a2,a3,a4]. Then each vector is transformed into a query and a key vector by the linear transformation matrix Wq and Wk. Starting with a query vector, e.g. q1, the query vector will have a dot product with the key vector of all the other, e.g. k1, k2, k3, k4 (including the other key vectors and itself). Then softmax is applied over all the scores, a1,1, a1,2, a1,3, a1,4 to normalize them to be positive and sum to one.

The attention mechanism will generate normalized weights $\hat{a}_{1,1}$, $\hat{a}_{1,2}, \hat{a}_{1,3}, \hat{a}_{1,4}$ which decide how "important" for each other the token is when calculating the next representation on the current token. There are 12 heads in each layer, so more than one head enables BERT to learn more about the structure of the text. BERT also stacks multiple layers of attention, each of which computes based on the output of the previous layer. Through this repeated structure, the attention heads from deeper layers are able to form richer representation after previous computation. In total BERT's architecture comprises 12 layers with 12 heads, resulting in a total of 12×12 = 144 different attention heads. Clark et al. (2019) analysed and visualized all these attention heads, revealing that the attention heads include patterns such as

https://doi.org/10.1017/pds.2022.84 Published online by Cambridge University Press

finding direct objects of verbs, determiners of nouns, objects of prepositions, and objects of possessive pronouns.

With respect to finding interesting patterns in patents mentioned in 3.3.1, an experiment was conducted and head 8-10 are shown and chosen to have the best results. Figure 3 shows two attention map examples on different interested attention patterns in patents, all these examples are computed on head 8-10. The line from one to another indicate the "importance" between each other, the deeper the colour is, the more important it is. The aiming words in a sentence are coloured red to highlight it from the others. For example, in the pattern of "dobj", the object word "signal" is coloured red to see if the verb is allocated with higher weight. The line between "signal" and right verb "send" has the deepest colour which means the attention head is recognizing the "dobj" pattern.

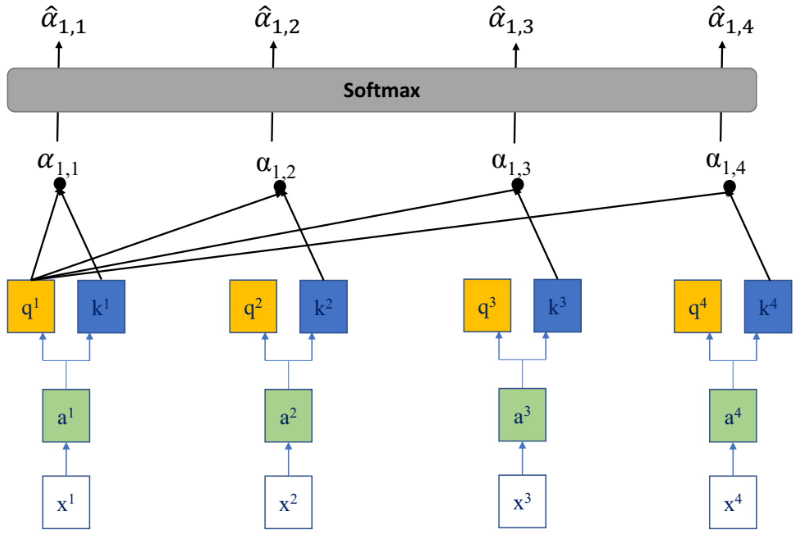


Figure 2. Computing process inside one head of attention

The attention mechanism in BERT calculates the attention from token to token. However, the words in our sentence are split into tokens and then re-combined if noun phrases and phrasal verbs are recognized. Therefore, the attention is converted from a token-to-token map to word(phrase)-to-word(phrase) map. For attention from a phrase, the mean of the attention weights is calculated over the tokens. For attention to a phrase, the sum of attention weights is calculated over the tokens. These transformations preserve the property that the attention from one to other sums to be one.

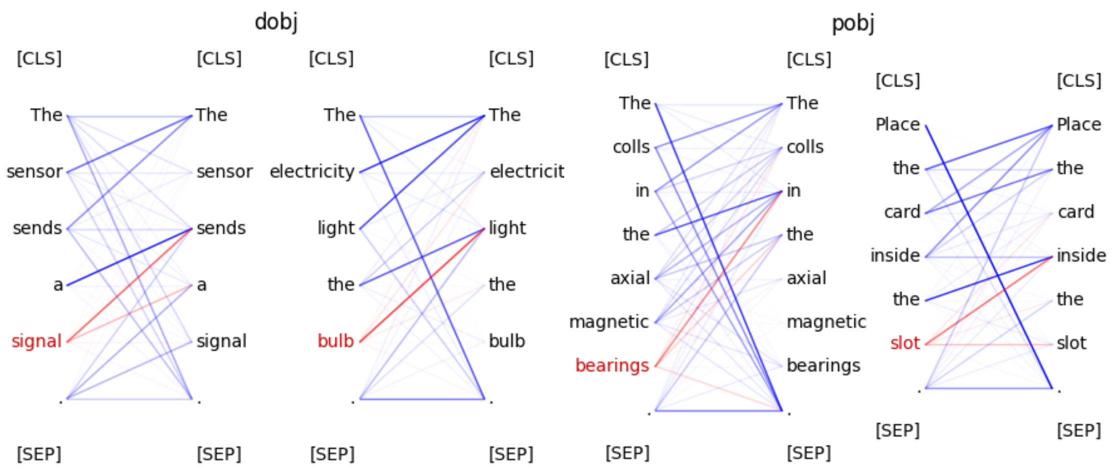


Figure 3. Attention examples on language patterns

After the calculation, the attention graph(matrix) will be obtained such as shown in Fig 4 (on the right). To interpret this, the rows mean the key (from), the columns mean the query (to). In the attention graph, the beam search is used to find the best matched relationship (mostly verbs)

https://doi.org/10.1017/pds.2022.84 Published online by Cambridge University Press

candidate fact. For every head and tail pair (h, t) in a sentence, the beam search will search backwards: t→r→h, computing through the k words with k-highest attention scores between the head and tail. Taking the sentence "the magnetic force provided levitates the shaft" as an example, the head-tail pair is (the magnetic force, the shaft), the beam size equals 2, the search computing processes are as the following:

(1) First, instead of searching forward, the searching algorithm is searching backward. The reason for searching backward is that the later words have stored the knowledge from previous words, while the previous words may not have read about later words. We need to add the tail "the shaft" to the beam, mark the head "the magnetic force" as the ending position and initialize the total attention degree as 0.

(2) Find the token with the largest attention score with the tail "the shaft", add that to the candidate (the shaft, levitates) and update the attention score 0.7761. Mark the token "levitates" as added to prevent search again. Then find the attention score between relationship "levitates" and "the magnetic force" which is 0.2496. The total attention score is 0.7761+0.2496= 1.0257

(3) Find the token with the second largest attention score with the tail "the shaft", add that as a candidate, (the shaft, provided) and update the attention score 0.0154. Mark the token "provided" as added to prevent search again. Then find the attention score between relationship "provided" and "the magnetic force" which is 0.5684. The total attention score is 0.0154+0.5684=0.5838

(4) The search will be stopped because the number of candidates reached the limit of beam size 2, and also reached the marked ending position "the shaft". The two candidate facts now both have attention scores from tail to relationship and relationship to head. The candidate with the highest attention scores, (the magnetic force, levitates, the shaft) will be kept and returned.

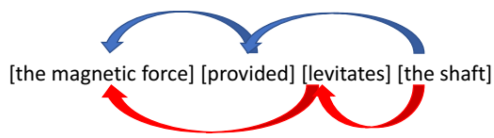
4. Evaluation

As shown in Table 1, patent-KG extracts the knowledge facts from 457,815 patents in Section F from 2016-2021. In total there are 4,157,377 entities and 10,991,896 edges. In this section, the outcome of two further evaluation experiments conducted with Patent-KG to demonstrate its feasibility and usefulness are reported. The evaluations are designed and conducted from in parts in terms of knowledge graph, the coverage of entities and the coverage of relationships.

The quality of entities and edges are evaluated as follows.

(h,t) - (the magnetic force, the shaft)

Step	Candidate facts:	Total attention score
1	(the shaft,	0
2	(levitates, the shaft	0.7761
	(The magnetic force, levitate, the shaft)	1.0257
3	(The magnetic force, levitate, the shaft) (provided, the shaft	1.0257 0.0154
	(The magnetic force, levitate, the shaft) (the magnetic force, provided, the shaft)	1.0257 0.5838
4	(The magnetic force, levitate, the shaft)	1.0257



	the magnetic force	provided	levitates	the shaft
the magnetic force	0.4243	0.3603	0.0341	0.0051
provided	0.5684	0.3450	0.0271	0.0073
levitates	0.2496	0.1256	0.3752	0.0196
the shaft	0.0638	0.0154	0.7761	0.1317

Figure 4. The searching process inside the attention graph

Table 1. The size of patent-KG

Number	Value
Number of Patents（F section）	457,815
Number of Entities	4,157,377
Number of edges	10,991,896
Number of phrasal verbs	317,789

4.1. Entities

In order to demonstrate the feasibility and usefulness of Patent-KG, a standard list of mechanical engineering related technical terms are built as golden concepts to compare with other benchmarks.
In the standard list, 3 categories (engines or pumps, engineering in general, lighting and heating) belonging to mechanical engineering are chosen and 13 subcategories, and totally 180 corresponding terms are chosen. Note that the meaning in the terms has some overlap between each other, the terms in one subcategory can also be classified into another. This list just choses the one of the appropriate subcategories. Table 2 lists the categories, the subcategories and part of the terms.

Table 2. Standard list of mechanical engineering related technical terms

Categories	lists
Engines or Pumps	
COMBUSTION ENGINES; HOT-GAS OR COMBUSTION-PRODUCT ENGINE PLANTS	heat engine, combustion chamber, working fluid, pistons, turbine blades, rotor, nozzle., internal combustion engine, four-stroke engine, compression-ignition engine, turbomachinery…
MACHINES OR ENGINES FOR LIQUIDS; WIND, SPRING, OR WEIGHT MOTORS; PRODUCING MECHANICAL POWER OR A REACTIVE PROPULSIVE THRUST	pelton wheels, wear-protection couplings, water current turbine, water wheels, Francis turbines, propeller turbines, Kaplan turbines, flywheel, fluid accumulator…
Engineering in general	
FLUID-PRESSURE ACTUATORS; HYDRAULICS OR PNEUMATICS IN GENERAL	pressure intensifier, isobaric pressure exchange, pneumatically operated actuator, hydraulic attachment, hydraulic circuit, positioner, electropneumatic transducer, nozzle-flapper system, pyrotechnic micro-actuator…
ENGINEERING ELEMENTS AND UNITS; GENERAL MEASURES FOR PRODUCING AND MAINTAINING EFFECTIVE FUNCTIONING OF MACHINES OR INSTALLATIONS; THERMAL INSULATION IN GENERAL	nails, staples, fastener, rod, anchor, toggle, dowels, bolts, hooks, gear, belts, chains, couplings, cranks, magnetic bearing…
STORING OR DISTRIBUTING GASES OR LIQUIDS	gas filling compartment, gasometers, gas container, gas reservoir , pressure vessels, gas cylinder, gas tank, replaceable cartridge…
Lighting, Heating	
LIGHTING	candle, flash light, illumination, headlight, LED, lamp, Incandescent mantles, pressure vessel…
STEAM GENERATION	evaporator, boiler, Inhalator, vaporizer, atomizer, Rankine cycle, working fluid, surface condenser, cooling tower…
COMBUSTION APPARATUS; COMBUSTION PROCESSES	Otto cycle, stove, chamber, burner, superheater, reheater…
REFRIGERATION OR COOLING; COMBINED HEATING AND REFRIGERATION SYSTEMS; HEAT PUMP SYSTEMS; MANUFACTURE OR STORAGE OF ICE; LIQUEFACTION SOLIDIFICATION OF GASES	freezer, refrigerator, compressor, rectifiers, cryogen, vapor-compression, Stirling cycle, defrost…
DRYING	dryer, convection, supercritical drying, dehydration, thermodynamics, moisture, filtration, centrifugation, temperature…

https://doi.org/10.1017/pds.2022.84 Published online by Cambridge University Press

With the standard list of mechanical engineering related technical terms, we then evaluate how many terms are contained in patent-KG and other benchmarks. There are 4 publicly accessible and related engineering datasets can be applied as benchmarks, which are WordNet, ConceptNet, B-link and TechNet. The retrieval rate CR is applied as the metric of concept retrieval which is an indication of coverage of the specific field. The retrieval rate is calculated as equation (1):

$$C_R = \frac{n_C}{N_C} \tag{1}$$

where n means the number of covered items in database, while N means the total items.
Table 3 shows the total recall rate and the three individual recall rates of the patent-KG and other benchmarks. The WordNet (Miller, 1998) and ConceptNet (Speer et al., 2017) performs at a lower recall rate because the construction of both mainly involves single-word terms and focus from a general level. The results between Patent-KG and TechNet by Sarica are close because both focus on patents and have a similar method to extract the terms. However, it should be noted that patent-KG only chooses the patents from 2016 to 2021, with a focus in the 'F' section, while TechNet has a broader time range and broader cover of disciplines. It can be observed that the total recall rate of the patent-KG reaches 0.82 and outperforms the other 4 benchmarks, which means a more engineering specific coverage than other methods. Specifically, patent-KG covers more technical terms in the category of 'Engines and pumps', while more specific terms are chosen as examples in this category.

Table 3. The comparison between Patent-KG and other baselines

	WordNet	ConceptNet	Feng (2017)	Sarica (2019)	Patent-KG
Total recall rate	0.46	0.56	0.63	0.79	0.82
Engines or Pumps	0.30	0.37	0.54	0.67	0.77
Engineering in general	0.40	0.42	0.66	0.81	0.82
Lighting, heating	0.68	0.88	0.70	0.88	0.86

4.2. Retrieved relationships

A standard list as shown in Table 4 of mechanical engineering related technical relationships is built as a benchmark to test whether it is covered in patent-KG. There are 5 publicly related engineering datasets published before patent-KG. However, some of these do have semantic relationships (B-link, TechNet), or some (Wordnet) use a linguistic English dictionary which mostly provides the taxonomic semantic relationships, such as hypernymy and hyponymy. It is unfair to have a quantifying comparison with a database with different purpose on relationship, so only the result of patent-KG is computed.

Table 4. Standard list of mechanical engineering related relationships

Name	Value
relationships	accelerate, add, assemble, block, compute, manufacture, select, prevent, have, made by(with), move, hold, connect to, connect through, include, followed with, reach, pull, lift...
retrieval rate	0.67

https://doi.org/10.1017/pds.2022.84 Published online by Cambridge University Press

5. Conclusions

A patent can be characterized by its long sentence and complex syntactic structure. With limited labelled data in patents, information extraction tasks can only be done using a traditional rule-based method or an unsupervised processing method. In this paper, we propose an unsupervised method to use the attention mechanism in the language model-BERT, to extract knowledge facts in patents. The quality of the extracted entities and relationships are demonstrated by comparing with other benchmarks. The entities recall rate shows that patent-KG is more engineering specific even with a smaller time scope and disciplines coverage. The relation extraction result suggests that the attention mechanism in language model works for knowledge extraction in patents, without complex rules.

The limitation is that the outcome does not perform well on passive sentences and very long sentences. To achieve better quality and broader coverage of knowledge graph for engineering design, the language models can be fine-tuned over patent text, and the attention mechanism can even be combined with syntactic rules. The knowledge graph is a fundamental step toward knowledge intelligence. In the developing direction of knowledge re-use, enhanced quality directs knowledge reuse from human and computer together as the collaborator, toward automatic knowledge generation with the computer as the creator.

References

Andersson, L., Lupu, M., Palotti, J., Hanbury, A. & Rauber, A. 2016. When is the Time Ripe for Natural Language Processing for Patent Passage Retrieval? Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. Indianapolis, Indiana, USA: Association for Computing Machinery.

Aristodemou, L. & Tietze, F. 2018. The state-of-the-art on Intellectual Property Analytics (IPA): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data. World Patent Information, 55, 37-51.

Cascini, G., Fantechi, A. & Spinicci, E. Natural language processing of patents and technical documentation. International Workshop on Document Analysis Systems, 2004. Springer, 508-520.

Chen Hao, D. Z., Li Yuan, Wang Zhengping, Xuan Yang, Yang Jidong. 2019. Automatic high-precision adjustment control system for common electric valve.

Chen, P.-L., Li, S.-C. & Hung, M.-T. Co-occurrence analysis in innovation management: Data processing of an online brainstorming platform. 2013 Proceedings of PICMET'13: Technology Management in the IT-Driven Services (PICMET), 2013. IEEE, 688-694.

Clark, K., Khandelwal, U., Levy, O. & Manning, C. D. J. a. E.-P. 2019. What Does BERT Look At? An Analysis of BERT's Attention. Available: https://ui.adsabs.harvard.edu/abs/2019arXiv190604341C [Accessed June 01, 2019].

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. J. a. P. A. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding.

Etzioni, O., Banko, M., Soderland, S. & Weld, D. S. J. C. O. T. A. 2008. Open information extraction from the web. 51, 68-74.

European Patent Office. 2021. CPC classification [Online]. 2021. Available: https://www.cooperativepatentclassification.org/cpcConcordances [Accessed].

Fader, A., Soderland, S. & Etzioni, O. Identifying relations for open information extraction. Proceedings of the 2011 conference on empirical methods in natural language processing, 2011. 1535-1545.

Fu, K., Cagan, J., Kotovsky, K. & Wood, K. 2013. Discovering Structure in Design Databases Through Functional and Surface Based Mapping. Journal of Mechanical Design, 135.

Honnibal Matthew, M. I. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.

Hu, M. 2018. Computer aided patent processing: Natural language processing, machine learning, and information retrieval. Ph.D., Drexel University.

Huang, Z., Xu, W. & Yu, K. J. a. P. A. 2015. Bidirectional LSTM-CRF models for sequence tagging.

Li, J., Sun, A., Han, J., Li, C. J. I. T. O. K. & Engineering, D. 2020. A survey on deep learning for named entity recognition.

Luo, L., Yang, Z., Yang, P., Zhang, Y., Wang, L., Lin, H. & Wang, J. J. B. 2018. An attention-based BiLSTM-CRF approach to document-level chemical named entity recognition. 34, 1381-1388.

Marie-Catherine De Marneffe, C. D. M. 2016. Stanford typed dependencies manual [Online]. Available: https://downloads.cs.stanford.edu/nlp/software/dependencies_manual.pdf [Accessed August 23,2021].

https://doi.org/10.1017/pds.2022.84 Published online by Cambridge University Press

Mccaffrey, T. & Spector, L. J. a. E. 2018. An approach to human–machine collaboration in innovation. 32, 1-15.
Miller, G. A. 1998. WordNet: An electronic lexical database, MIT press.
Mintz, M., Bills, S., Snow, R. & Jurafsky, D. Distant supervision for relation extraction without labeled data. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2, 2009. Association for Computational Linguistics, 1003-1011.
Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. & Sutskever, I. J. O. B. 2019. Language models are unsupervised multitask learners. 1, 9.
Sarica, S., Luo, J. & Wood, K. L. 2020. TechNet: Technology semantic network based on patent data. Expert Systems with Applications, 142, 112995.
Schmitz, M., Soderland, S., Bart, R. & Etzioni, O. Open language learning for information extraction. Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning, 2012. 523-534.
Shi, F. 2018. A Data-Driven Text Mining and Semantic Network Analysis for Design Information Retrieval.
Siddharth, L., Blessing, L. T. M., Wood, K. L. & Luo, J. J. a. E.-P. 2021. Engineering Knowledge Graph from Patent Database. Available: https://ui.adsabs.harvard.edu/abs/2021arXiv210606739S [Accessed June 01, 2021].
Speer, R., Chin, J. & Havasi, C. Conceptnet 5.5: An open multilingual graph of general knowledge. Thirty-first AAAI conference on artificial intelligence, 2017.
Swanson, D. R. J. T. L. Q. 1986. Undiscovered public knowledge. 56, 103-118.
Van Noorden, R. 2014. Scientists may be reaching a peak in reading habits. Nature.
Verberne, S., D'hondt, E., Oostdijk, N. & Koster, C. H. 2010. Quantifying the challenges in parsing patent claims.
Wang, A., Singh, A., Michael, J., Hill, F., Levy, O. & Bowman, S. R. J. a. P. A. 2018. GLUE: A multi-task benchmark and analysis platform for natural language understanding.
Wang, C., Liu, X. & Song, D. 2020. Language Models are Open Knowledge Graphs. Available: https://ui.adsabs.harvard.edu/abs/2020arXiv201011967W [Accessed October 01, 2020].
Wang, Q., Mao, Z., Wang, B., Guo, L. J. I. T. O. K. & Engineering, D. 2017. Knowledge graph embedding: A survey of approaches and applications. 29, 2724-2743.
World Intellectual Property Organization. 2019. World Intellectual Property Indicators 2019 [Online]. Available: https://www.wipo.int/publications/en/details.jsp?id=4464 [Accessed August 23, 2021].