



Using Text Understanding to Create Formatted Semantic Web from BIM

Jingming Li^{1,2}(✉)

¹ Midea Building Technology, Midea Group, Foshan, China
jli@hnu.edu.cn

² College of Civil Engineering, Hunan University, Changsha, China

Abstract. The application of BIM in the building life cycle needs to be continuous. The information collected and accumulated in the early stages should flow to the subsequent phases. However, BIM applications currently focus on collision inspection, compliance inspection, and engineering calculation, few models can be successively used in the following stages. Remodeling is required in the operation and maintenance period, resulting in waste. Meanwhile, some of the information accumulated by BIM might be frequently used in the operation and maintenance stage, while some data are relatively rarely used. The semantic web can help manage building information at all stages. But the generation of a semantic web is mostly manually completed. It is necessary to standardize the repeated semantic description in the model and convert BIM into a standard semantic model for information indexing, reducing the resource consumption of model loading and optimizing the efficiency of the operation and maintenance system. When the existing research transforms from BIM to the semantic web, there will be a lack of information and descriptions of the ownership relationship between entities due to the limitation of formats. To realize the standard transformation from BIM to the semantic web, this work proposes a method of using Natural Language Processing (NLP) to understand the text and infer the relationship between entities according to the knowledge map. First, the entities are extracted from BIM, such as air conditioning unit, electric lamp, fan, etc., if the name of the extracted entity is irregular, the names are translated with the help of NLP and Ontology (such as brick or haystack) to obtain the standard definition. By comparing the complete knowledge graph (such as the knowledge graph of the air conditioning system), the relationships can be deduced, and then a standardized semantic model can be generated.

Keywords: BIM · Semantic web · Text understanding

1 Introduction

Global total energy consumption has reached 162 thousand terawatts in 2019. As the world's largest economies, China and the United States account for more than 40% of global energy consumption [1]. Among them, the building sector accounts for 18.35% of China's total energy consumption, and the ratio has reached 40% in the US [2, 3]. In

© The Author(s) 2023

P. F. Yuan et al. (eds.), *Hybrid Intelligence*, Computational Design and Robotic Fabrication,
https://doi.org/10.1007/978-981-19-8637-6_17

the context of tackling climate change, how to effectively reduce building energy consumption has become an important topic of energy conservation and carbon reduction. Using Building Information Model (BIM), Semantic Web, Internet of Things and other information technologies to improve building operation is one of the key research areas.

BIM contains lots of details since early design stages in buildings. IFC is a common format for BIM. In addition to attribute information, IFC also contains a lot of 3D information, and the material and other information carried by IFC can also be used for rendering 3D effects. The model itself is a large dataset, meaning IFC is not an ideal index to manage the data in the Operation and Maintenance (OM) phase.

The Semantic Web is a network of data that includes dates, titles, part numbers, chemical properties, and any other types of data. Brick and Haystack [4] are two universal semantic schemas for defining entities in buildings. Haystack tags resolve data silos among various subsystems (HVAC, lighting, and enterprise scheduling). Brick aims to standardize the semantic description of buildings, including physical, logical, virtual assets and the relationships between them. Using the Semantic Web, Brick can coherently describe many special and custom functions, assets, and subsystems throughout the building life cycle.

Using Brick or Haystack as a description specification for a building reduces the cost of deploying analytics, energy efficiency measures, and smart controls across buildings, demonstrating the integration of numerous subsystems in a modern building: HVAC, lighting, fire protection, security, etc. Simplifies smart analytics and control applications development, as well as reduces reliance on non-standard, unstructured labels specific to building management systems. But the conversion process is still challenging.

1.1 Related Works

Currently, there are many types of research and applications in building compliance, with Ontology, Metadata, and Semantic Web, but most of them are used for building design and model detection. These studies usually extract semantics from BIM models to generate graphs, or directly use BIM models as query objects, and then carry out deductions such as cost budgeting, energy design, and construction hazard identification based on standards [5–7].

Through the integrated application of BIM and Semantic Web technology, the project [8] design conforms to the construction quality specifications. The project can automatically check the size and position of the BIM model components according to the requirements of the specification, thereby reducing the benchmarking workload of the relevant personnel during the construction process.

When extracting semantics from BIM and comparing them with standards, the process usually introduces ontology description technologies such as SPARQL and Web Ontology Language (OWL) or Unified Modeling Language (UML) to improve the standardization and efficiency of retrieving BIM semantics. For instance, McGibney and Kumar [9] summarized BIM model detection based on semantic implementation. The process firstly converts specifications and models to semantics and then extracts the BIM model based on specification requirements, which improves the efficiency of specification and model checking. Bi et al. [10] used ontology for knowledge expression to promote the protection of ancient buildings.

The current applications of the Semantic Web are similar to labeling entities in Brick and Haystack. Labels produce good guidance for data, but cannot effectively represent relationships between entities. Also, the knowledge graph exported from IFC at this stage contains too much redundant data, which reduces the efficiency of data management. Although the complex types of data in the Operation and Maintenance (OM) phase are more suitable for the Semantic Web to exert its value, few studies have applied it to OM because of the difficulties in creating a semantic web. The application of BIM in the building life cycle is continuous. After accumulating information in BIM, only part of the data will be reused in the OM phase.

However, the applications of BIM are mainly concentrated in the application stages such as model collision check, compliance check, and engineering calculation, and rarely extend to the OM stage, causing massive remodeling works and waste of resources. It is necessary to standardize the repeated semantic description in the model and convert the BIM into a standard semantic model for information indexing, which can reduce model loading, reduce resource consumption and optimize the efficiency of data sharing in the building operating system. Currently, the semantic simplification of BIM models is still done manually, which consumes a lot of manpower and time.

1.2 Contributions

This work explores the scheme of building a standard semantic model based on BIM. First, the entities are extracted from the BIM model, and the names are standardized or transformed regarding the ontology dictionary. Then, the belonging and supply relationships between each entity will be reasoned by combining the knowledge graph. Finally, the custom entities in BIM are converted into standard semantic models and the transition of building information from early stages to OM.

2 Methodology

To create the standard conversion from BIM to Semantic Web, this chapter proposes a method of using NLP to understand entities and infer the interrelationships between entities according to the knowledge graph. The implementation framework of this method is shown in Fig. 1.

2.1 Entity Extraction

The first step is to extract entities from the data source, such as the BIM model.

When the existing research converts BIM information to the Semantic Web, due to the lack of information, the conversion in Industry Foundation Classes (IFC) is geographical rather than semantical. It cannot infer the relationship between entities when exporting ifcOWL.

There are options to extract entities from the BIM model, directly extracted through plug-ins, or converted into ifcOWL. The latter will cause information loss in the conversion process. In contrast, directly extracting entities from the model can avoid data loss during the conversion.

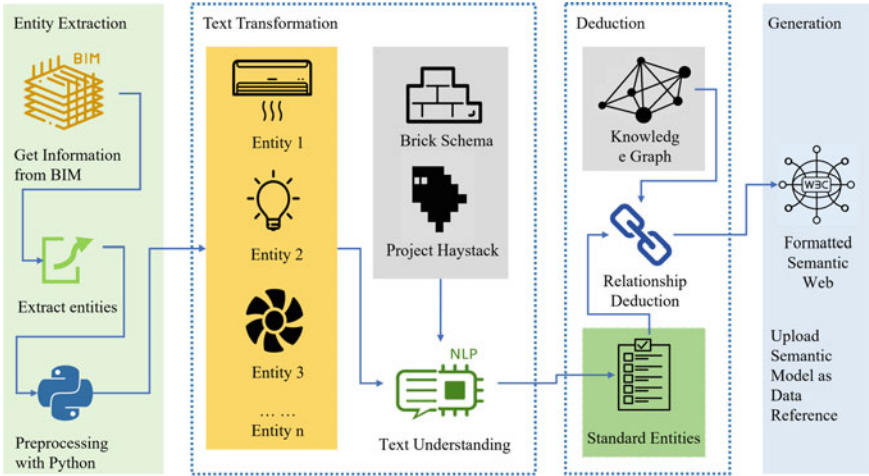


Fig. 1. The method of generating standard SW model from BIM

Direct data extraction from the BIM model can use plug-ins such as Dynamo or Grasshopper, as shown in Fig. 2. First, the script traverses all the categories in the file (such as piping, mechanical, electrical, etc.). Then, it traverses the entities under each category. Finally, the program returns all the entity names and family categories. The built-in conditional grouping module of Dynamo can group entities according to their spatial positions. After deduplication, entities and their positional relationships can be obtained.

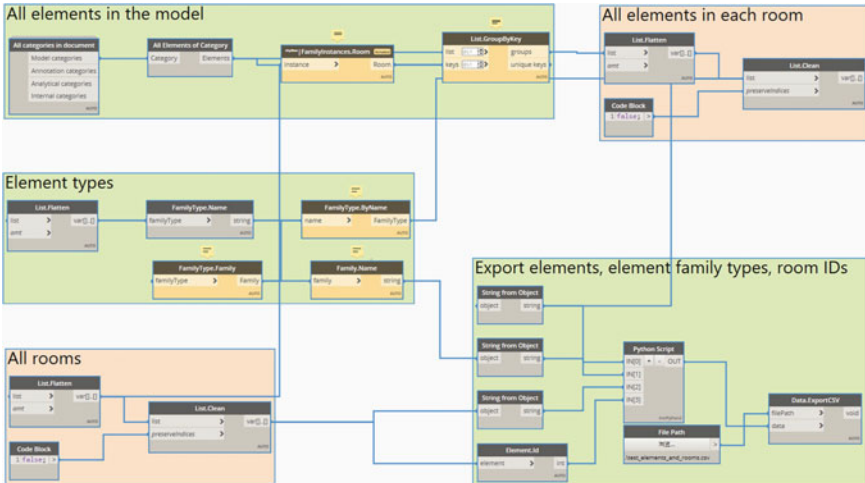


Fig. 2. Extracting entities based on location relationships from BIM

As illustrated in Fig. 3, following similar settings, the extraction of entities can also be carried out according to the supply relationship. The script uses the third-party plug-in (Spring Node) to obtain the associated entities within the view. By setting the category of the parent equipment and the category of the equipment being supplied respectively, and removing the repeating entities, the supply relationships can be obtained.

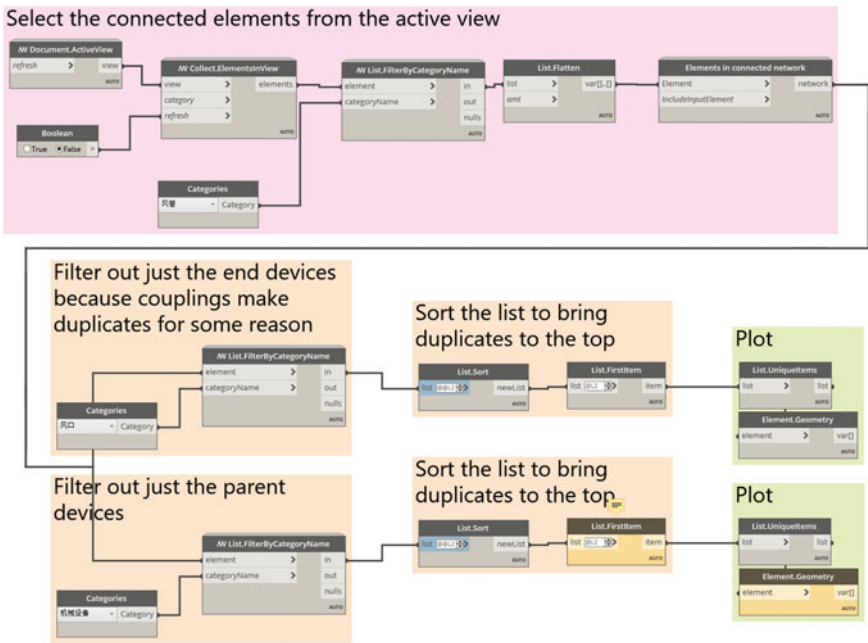


Fig. 3. Extracting entities based on feeding relationships from BIM

2.2 Text Transformation and Relationship Inference

The first step is to determine the entity name. If the names of the extracted entities are irregular, NLP is needed to understand its family name through cosine similarity, as shown in Fig. 4. Knowledge graphs use visualization techniques to describe knowledge resources, which can be used to display, analyze, and reason about the interconnections between entities. The relationship can be derived according to the knowledge graph. The translated standard entity can infer the relationships according to the location in the knowledge graph. The implementation of relational reasoning mainly relies on the SPARQL query on the complete knowledge graph and creates relationships for newly generated entities according to the query results.

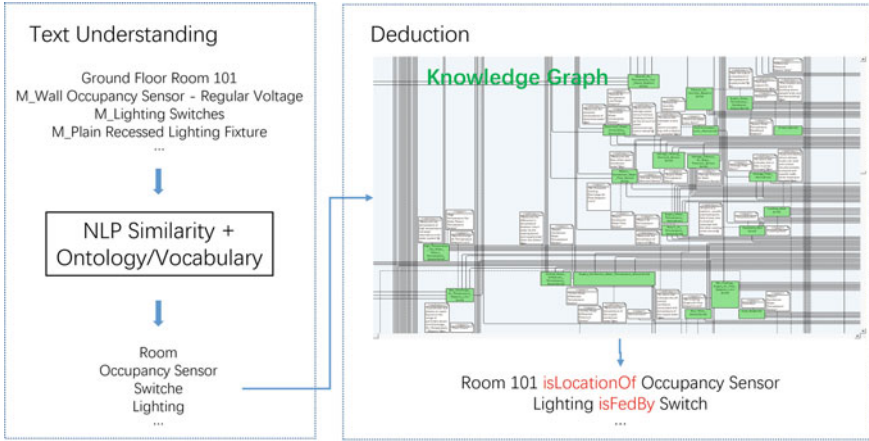


Fig. 4. The reasoning of relationships from custom entity names

A_i and B_i are the vectors transformed from the standard dictionary of name and ontology respectively, and the cosine similarity of the vectors is calculated in turn, as shown in Eq. 1. The closer the cosine similarity is to 1, the closer the two words are to each other, -1 means the words are opposite.

$$\text{similarity} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \tag{1}$$

Through the cosine similarity evaluation of the vector, the entity name can be transformed into the standard vocabulary of the ontology dictionary. The framework uses tools PyTorch and transformers [11]. BERT converts the text into entries with a length of 128. Each entry has a separate 768-digit vector. Pooling will extract the average of all tags and combine them into a unique 768 vector space to produce a “sentence vector”. Based on the pre-trained models, the program converts the non-standard vocabulary and ontology vocabulary into vectors respectively, uses PyTorch to calculate the cosine similarity, and finally selects the closest standard word.

If the naming rules of the model are clear, such as RM_101_IDU_134_T1 refers to No. 134 indoor unit in Room 101, it only needs to be converted into the standard name of the body according to the naming rules. The naming rules may already include the location and supply relationships, and the relationship between entities can be directly disassembled from the naming. Figure 5 shows the process.

By comparing the inferred relationships with a complete knowledge graph (such as the knowledge graph of an air-conditioning water-cooling system), the interrelationships between entities can be deduced, and a standardized semantic model can be generated.

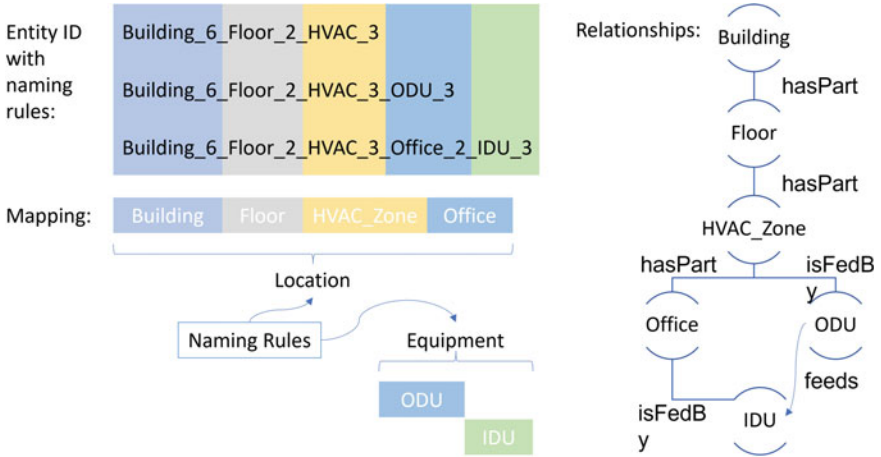


Fig. 5. The reasoning of entity relationships from formatted entity names

3 Experiment Results and Discussions

An experiment was conducted on the HVAC file within Revit. The program extracts examples from BIM files and classifies them into 42 categories. 25% of the categories, which are the data sources and control units, are included in Brick or Haystack. The classification of examples is carried out according to ontology standards, and the results are shown in Table 1. The extracted examples are mainly HVAC and electrical equipment. The entity names may not be standardized, for instance, Trck_BswySysmts_Cooper_RSA_Profile Series_AR111 Closed Back Integral Xfmr is a lighting device.

Figure 6 shows the comparison results between the pre-trained language model and brick and haystack standards. The SentenceTransformer loads the pretrained models [12] and calculates the most similar word, respectively. From the results that even the model specially adjusted for similarity calculation has poor results. The accuracy of the first mock exam is only 60%, and the same model is not stable when dealing with different ontology standards. Paraphrase-mpnet-base-v2, which performs best in classifying and looking for similar texts, has a 40% difference in accuracy between brick standard and haystack standard. However, this problem will be improved with the strengthening of buildings and equipment by the pretrained model, which is also confirmed by the results of different pretrained models.

Compared with the results of brick vocabulary calculation, although the natural language model has been optimized in their respective training sets, its performance is still unstable in the process of practical application, especially in the fields involving specific professional knowledge.

Table 1. The words covered in Brick and Haystack

Model Entities	Entity Type	Brick	Haystack
M_Wall Occupancy Sensor—Regular Voltage	Occupancy sensor	✓	✓
M_Lighting Switches	Switch	✓	✓
M_Plain Recessed Lighting Fixture	Lighting	✓	✓
M_Pendant Light—Linear—2 Lamp	Lighting	✓	✓
M_Sconce Light—Flat Round	Lighting	✓	✓
Trck_BswySystms_Cooper_RSA_Profile Series_AR111 Closed Back Integral Xfmr	Lighting	✓	✓
M_Pile-Steel Pipe	Pipe	–	✓
Water	Water	✓	✓
DC_Tankworks_PLM_5000L	Water tank	✓	✓
Miele MasterCool KF 1911 Vi	Freezer	✓	–

For non-standard named entities, semantic understanding is required. Due to the ontology standards' low coverage of the BIM model and the fact that the pretrained model has not been strengthened for buildings and equipment, the accuracy of standardized text translation is low, and the automation level in the process of generating standardized semantic web from non-standard named entities in BIM file is low, so manual participation is still needed at present.

When using NLP to understand professional knowledge, it still needs to be constrained by a large amount of relevant professional knowledge. For example, when this chapter explores the use of NLP to transform the non-standard language description in BIM into the ontology standard in the field of HVAC, the implementation of this process requires HVAC professional knowledge to provide constraints for machine-reading and understanding and optimize the training set.

However, the pretrained model has not been optimized enough, and there is a large deviation between the results of machine understanding and the actual situation. The accuracy of BIM + NLP in converting non-standard entity names needs to be improved, and the current model also needs to strengthen the language recognition of professional vocabulary in the field of architecture and HVAC.

4 Conclusions

Based on BIM Technology, this work studies the method of standard semantic description of BIM entities, including the extraction of entities and their relationships from BIM and the entity name transformation of NLP semantic understanding.

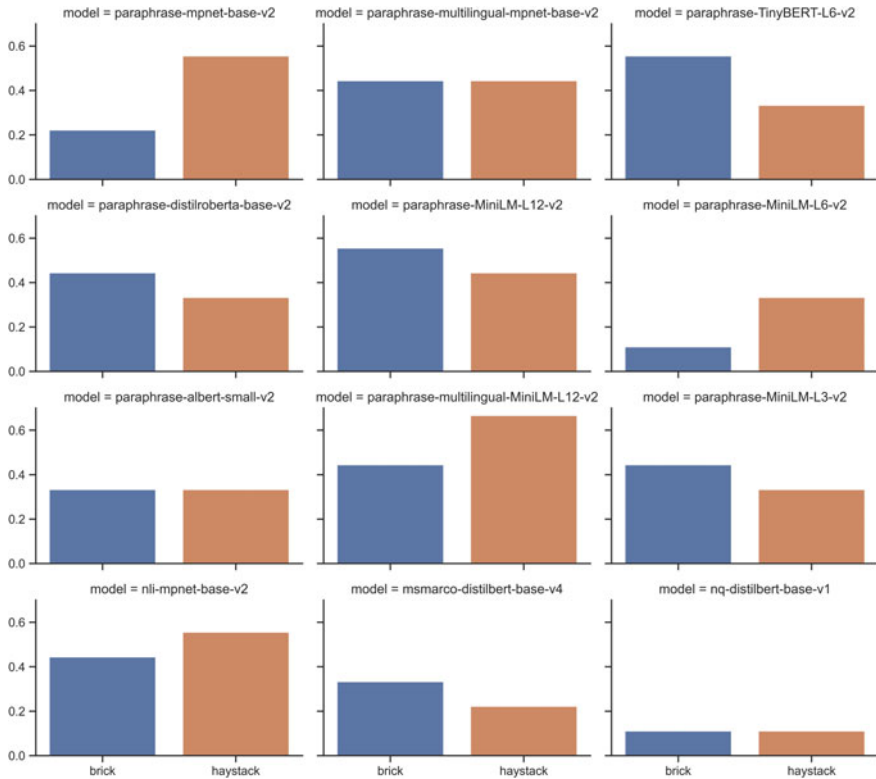


Fig. 6. The results of corrections

Currently, the standardization of BIM entity naming rules is an effective means. Due to the low accuracy in understanding professional HVAC knowledge, the conversion of BIM entities without clear naming rules is still dominated by manual work, and automation is limited. In future research, optimizing the pretrained model through professional knowledge will be the key to optimizing the NLP translation results. With the assistance of ERNIE 3.0 or GPT-3, semantic understanding should be more accurate during BIM to semantic web conversions.

References

1. Ritchie H, Roser M (2021) Energy production and consumption—our world in data. <https://ourworldindata.org/energy-production-consumption#energy-production-and-consumption-by-source>. Accessed 8 Mar 2022
2. Department of Energy (DOE) (2015) Chapter 5: increasing efficiency of building systems and technologies. Quadrenn Technol Rev An Assess Energy Technol Res Oppor, 143–181
3. Yan L (2018) China: energy efficiency report, 2018

4. Quinn C, McArthur JJ (2021) A case study comparing the completeness and expressiveness of two industry-recognized ontologies. *Adv Eng Inf* 47:101233. <https://doi.org/10.1016/j.aei.2020.101233>
5. Staub-French S, Fischer M, Kunz J, Ishii K, Paulson B (2003) A feature ontology to support construction cost estimating. *Artif Intell Eng Des Anal Manuf AIEDAM*. <https://doi.org/10.1017/S0890060403172034>
6. Lork C, Choudhary V, Ul Hassan N, Tushar W, Yuen C, Ng BKK, Wang X, Liu X (2019) An ontology-based framework for building energy management with IoT. *Electron*. <https://doi.org/10.3390/electronics8050485>
7. Zhang H, Gu M, Sun J (2017) The method and the device for BIM model specification detection based on semantic retrieval
8. Farghaly K, Soman RK, Collinge W, Mosleh MH, Manu P, Cheung CM (2022) Construction safety ontology development and alignment with industry foundation classes (IFC). *Electron J Inf Technol Constr*. [https://www.research.manchester.ac.uk/portal/en/publications/construction-safety-ontology-development-and-alignment-with-industry-foundation-classes-ifc\(67e547bd-c261-4c34-9460-afbeb103aedf\).html](https://www.research.manchester.ac.uk/portal/en/publications/construction-safety-ontology-development-and-alignment-with-industry-foundation-classes-ifc(67e547bd-c261-4c34-9460-afbeb103aedf).html). Accessed 8 Mar 2022
9. McGibbney LJ, Kumar B (2015) A framework for regulatory ontology construction within AEC domain. In: *Ontology in the AEC industry: a decade of research and development in architecture, engineering, and construction*
10. Bi Z, Wang H, Lu Y (2014) A construction model of ancient architecture protection domain ontology based on software engineering and CLT. *J Softw* <https://doi.org/10.4304/jsw.9.11.2886-2894>
11. Reimers N, Gurevych I (2019) Sentence-BERT: sentence embeddings using siamese BERT-networks. *EMNLP-IJCNLP 2019–2019 Conf Empir Methods Nat Lang Process 9th Int Jt Conf Nat Lang Process Proc Conf*, 3982–3992
12. Ubiquitous Knowledge Processing Lab (2020) Pretrained models—sentence-transformers documentation. https://www.sbert.net/docs/pretrained_models.html. Accessed 30 Aug 2021

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

