

Article

# A Survey of Techniques for Constructing Chinese Knowledge Graphs and Their Applications

Tianxing Wu<sup>1,2</sup>, Guilin Qi<sup>1,\*</sup>, Cheng Li<sup>1</sup> and Meng Wang<sup>1</sup>

<sup>1</sup> School of Computer Science and Engineering, Southeast University, Nanjing 211189, China; wutianxing@seu.edu.cn (T.W.); licheng1205@seu.edu.cn (C.L.); wangmengsd@outlook.com (M.W.)

<sup>2</sup> School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore

\* Correspondence: gqi@seu.edu.cn; Tel.: +86-25-52090910

Received: 8 August 2018; Accepted: 10 September 2018; Published: 11 September 2018



**Abstract:** With the continuous development of intelligent technologies, knowledge graph, the backbone of artificial intelligence, has attracted much attention from both academic and industrial communities due to its powerful capability of knowledge representation and reasoning. In recent years, knowledge graph has been widely applied in different kinds of applications, such as semantic search, question answering, knowledge management and so on. Techniques for building Chinese knowledge graphs are also developing rapidly and different Chinese knowledge graphs have been constructed to support various applications. Under the background of the “One Belt One Road (OBOR)” initiative, cooperating with the countries along OBOR on studying knowledge graph techniques and applications will greatly promote the development of artificial intelligence. At the same time, the accumulated experience of China in developing knowledge graphs is also a good reference to develop non-English knowledge graphs. In this paper, we aim to introduce the techniques of constructing Chinese knowledge graphs and their applications, as well as analyse the impact of knowledge graph on OBOR. We first describe the background of OBOR, and then introduce the concept and development history of knowledge graph and typical Chinese knowledge graphs. Afterwards, we present the details of techniques for constructing Chinese knowledge graphs, and demonstrate several applications of Chinese knowledge graphs. Finally, we list some examples to explain the potential impacts of knowledge graph on OBOR.

**Keywords:** knowledge graph; intelligent applications; OBOR

## 1. OBOR Background

In 2013, China proposed the cooperation initiative of building the new “Silk Road Economic Belt” and “21st Century Maritime Silk Road”, referred as “One Belt One Road (OBOR)”. With existing and effective regional cooperation platforms and multilateral mechanisms between China and related countries, China actively develops the partnership with Eurasian countries along OBOR and promotes common development in the field of economics, politics, culture, science and technology. By making use of the historical symbol “Silk Road”, it aims to build a community with mutual political trust, economic integration, cultural inclusiveness and common development of science and technologies. Therefore, the OBOR initiative is seen as a bid to enhance regional connectivity and embrace a brighter future. From 2013 to 2018, the OBOR initiative has attracted lots of countries to join and made significant progress, such as “China–Pakistan Economic Corridor”, “Regional Comprehensive Economic Partnership”, “Trans-Eurasia Logistics” etc.

In the 21st century, the strength of science and technology gradually becomes an important standard to measure the comprehensive strength of a country. The development of science and

technologies also promotes the development of economy. Therefore, the cooperation of science and technologies with the countries along OBOR is also an important part of OBOR policies. As one of the core components of modern science and technologies, artificial intelligence has attracted much attention all over the world. In China, developing artificial intelligence has even been raised to a national strategy. Knowledge graph, the backbone of artificial intelligence, has already shown its powerful capability of knowledge representation and reasoning in different kinds of intelligent applications, such as semantic search, question answering, knowledge management, decision making, information integration etc. Under the background of the OBOR initiative, cooperating with the countries along OBOR on studying knowledge graph techniques and applications will greatly promote the development of artificial intelligence. At the same time, the accumulated experience of China on developing the techniques of constructing knowledge graphs is also a good reference to develop non-English knowledge graphs, so we mainly discuss the techniques of constructing Chinese knowledge graphs and their applications, as well as analyse the impact of knowledge graph on OBOR in this paper.

The rest of this paper is organized as follows. Section 2 introduces the basic concept and development history of knowledge graph and typical examples of Chinese knowledge graphs. Section 3 presents the techniques of constructing real-world Chinese knowledge graphs in detail. Section 4 describes different intelligent applications supported by Chinese knowledge graphs. Section 5 lists some examples to explain the potential impacts of knowledge graph on OBOR, and we conclude in the last section.

## 2. Knowledge Graph

### 2.1. Introduction to Knowledge Graph

The concept of knowledge graph was originally proposed by Google Knowledge Graph project on 16 May 2012. This project aimed at using structured knowledge to enhance Google search engine, and improve search quality and user experience. As shown in Figure 1, when searching “Yao Ming” in Google, in addition to the web pages related to “Yao Ming” shown on the left side, some structured information of “Yao Ming” will also be presented on the right side (e.g., date of birth, place of birth, height and his wife). Such comprehensive returned results rely on the background knowledge graph. Compared with the traditional keyword matching-based web search, knowledge graph-based semantic search can better understand users’ query intents, thus providing more concise and intelligent search results. For example, when searching “Height of Yao Ming” in Google, it not only returns the web pages related to “Yao Ming” and “Height” respectively, but also directly returns “2.29 m” to users. Subsequently, some well-known companies claimed that knowledge graph is one of the core components of their architectures for building the next generation of intelligent search engines. To enhance Bing search, Microsoft built a large-scale knowledge graph called “Satori”. Baidu company has constructed three different kinds of knowledge graphs including entity graph, concern graph and intention graph and apply them to its search engine. Sougou, another Chinese company which is famous for its search engine, also built a knowledge graph called “Knowledge Cube” to improve the search quality.

Nowadays, knowledge graph has become a very popular terminology in the field of artificial intelligence, database and semantic Web, and it refers to any collection of knowledge represented in the form of graph, such as semantic web knowledge bases (e.g., DBpedia [1] and Yago [2]), RDF datasets and formal ontologies. The nodes in the graph structure can be concepts, entities and literals, and edges are different kinds of relationships. As shown in Figure 2, entities represent real-world individuals (e.g., “P.R. China” and “Beijing”). A concept represents a set of individuals with the same characteristics, for example, “P.R. China”, “Greece”, “the United States” etc., compose a set corresponding to the concept “Country”. Literals refer to the strings which indicate specific values of some relations, such as string “中国”, the Chinese name of entity “P.R. China”. Edges between these nodes represent different relationships between entities, concepts and literals, such as “P.R. China” is a “Country” and the capital of “P.R. China” is “Beijing”. All these relationships and their related entities, concepts or literals

are stored in the form of triples in knowledge graphs which is the basic storage unit of knowledge graphs. Triples organize knowledge in the form of <subject, predicate, object>, e.g., <P.R. China, is-a, Country> and <P.R. China, Chinese name, “中国”>.

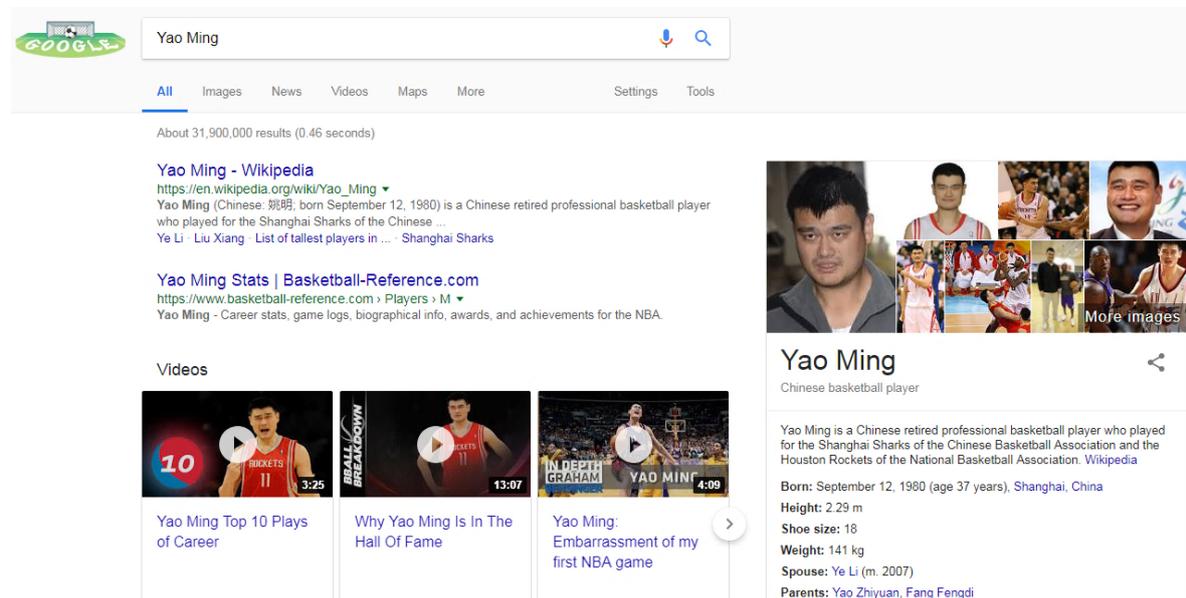


Figure 1. Search results of “Yao Ming” in Google search engine.

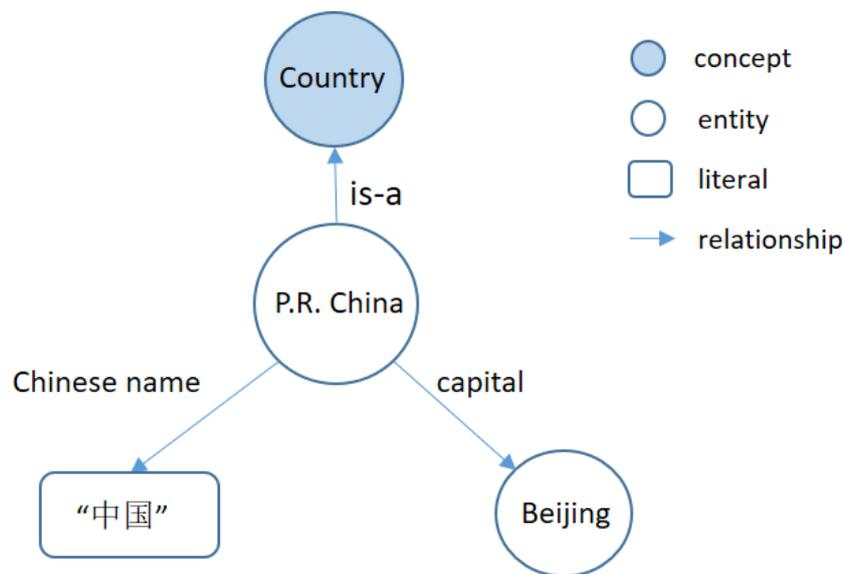


Figure 2. An example for a part of a knowledge graph.

The core target of artificial intelligence is to provide machines with the ability of learning and reasoning, thus lots of researchers focus on studying the techniques of machine learning, logical deductions, etc. One key issue here is how to represent and store the acquired knowledge for machines to reuse. Knowledge graph is a powerful solution to this problem. It makes machines more intelligent by providing machine-readable knowledge, which enhance the capability of learning and reasoning. Knowledge graphs in different languages and different domains are now promoting the development of artificial intelligence.

## 2.2. The Development of Knowledge Graph

Knowledge graph is essentially originated from semantic network [3]. Semantic network was proposed in late 1950s and early 1960s, and it is a graph-based data structure to store knowledge, where the graph can be directed or undirected. It is quite convenient to utilize semantic network to represent and store natural language sentences, which can be further applied in machine translation, question answering and natural language understanding. In 1970s, lots of works began to study the relationship between semantic network and first-order predicate. For example, Simmons et al. [4] presented an algorithm to convert semantic network to predicate logic, while Schubert [5] proposed a method leveraging semantic network to represent the conjunctions and quantifiers in first-order predicate logic. In 1980s, the mainstream of artificial intelligence research had become knowledge engineering and expert system, especially rule-based expert system. During this period, the theory of semantic network became more mature and lots of research worked on reasoning based on semantic network [6]. More importantly, the research on semantic network started to turn to knowledge representation and reasoning with strict logical semantics. From late 1980s to 1990s, the research on semantic network focused on modeling the relationship between concepts. Based on this, terminology logic and description logic were proposed. The representative work in this period is the CLASSIC language [7] proposed by Brachman et al. and FaCT reasoning machine [8] implemented by Horrocks. In the 21st century, semantic network has a new application scenario, i.e., semantic Web. The semantic Web was proposed by Tim Berners-Lee, the founder of the Web, and his collaborators. They aimed to implement an extension of the Web through W3C standards so that data can be shared and reused in different applications. In this phase, knowledge is often organized as schemata or ontologies. After that, with the advent of Linked Data [9], different kinds of knowledge bases were published and interlinked on the Web, which form a large-scale global knowledge base, called Linked Open Data Cloud (see Figure 3 (<https://lod-cloud.net/>)). Such open accessed knowledge bases laid the foundation for the success of Google Knowledge Graph project.

Compared with the early semantic network, knowledge graph has its own characteristics. First of all, knowledge graph consists of concepts, entities, literals and their relationships. While early semantic network was used to represent the natural language sentences, not structured knowledge. Secondly, compared with manually built semantic networks, knowledge graph construction mainly depends on automatic knowledge discovery techniques, which are often applied in online encyclopedias, text, and databases. Finally, knowledge graph construction emphasizes knowledge fusion from different sources and knowledge cleansing, which are not the focus of early semantic network. With the constant improvement of knowledge discovery techniques and enrichment of data sources (text, images, sensors, video, etc.), the scale of different knowledge graphs is growing explosively. Taking the Linked Open Data Cloud as the example, there were about 2 billion RDF triples in 2007, the number increased to 31 billion in 2011 and by the end of August 2017, it released more than 40 billion RDF triples.

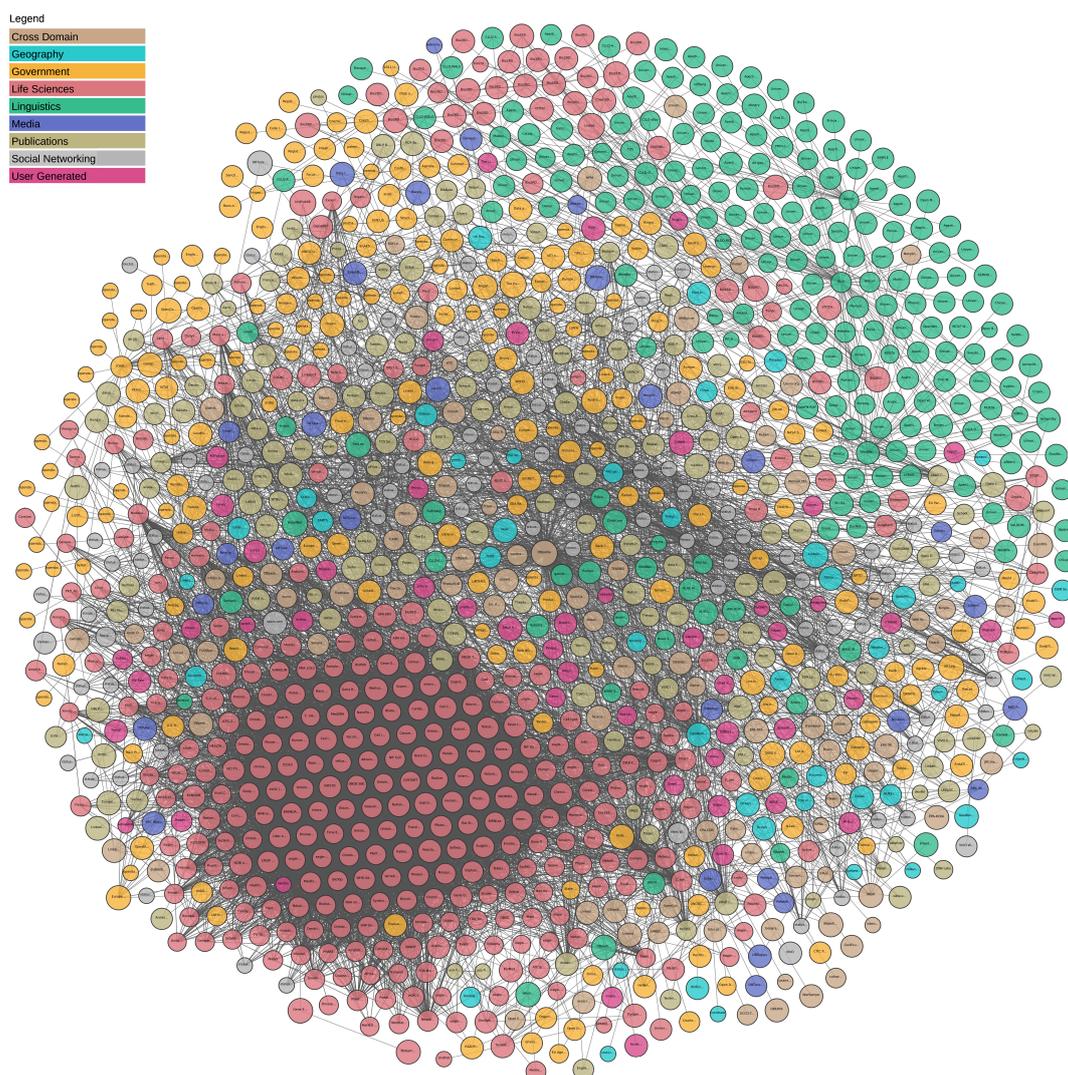


Figure 3. Linked Open Data Cloud diagram.

### 2.3. Typical Chinese Knowledge Graphs

Although Linked Open Data Cloud contains some multilingual knowledge graphs, most of the knowledge (including concepts, entities, triples, etc.) is still denoted in English, and the number of Chinese knowledge is quite few. For example, DBpedia contains more than 5 million English entities, but only contains less than 1 million Chinese entities. Yago does not have a Chinese version though it contains a certain number of Chinese labels. Lack of real world practical Chinese knowledge graphs hinders the development of semantic applications, artificial intelligence and semantic Web itself in Chinese. This is why Chinese researchers should make great efforts to construct Chinese knowledge graphs, and the good news is that they have achieved very encouraging progress, such as the large-scale Chinese encyclopedic knowledge graphs Zhishi.me [10], CN-DBpedia [11], XLORE [12,13], PKU-PIE (<http://openkg.cn/dataset/pku-pie>), Belief-Engine (<http://www.belief-engine.org>) and the Chinese schema-level knowledge graphs cnSchema (<http://cnschema.org>) and Linked Open Schema [14–16].

**Zhishi.me:** Zhishi.me is the first Chinese encyclopedic knowledge graph which is built from three largest Chinese online encyclopedias: Baidu Baike (<https://baike.baidu.com>), Hudong Baike (<http://www.baik.com>) and Chinese Wikipedia (<https://zh.wikipedia.org>). It adopts the methods similar to those of DBpedia to extract structured knowledge from these three online encyclopedias and link equivalent entities across them by fixed rules, so as to build a large-scale Chinese general knowledge graph. It now contains more than 10 million entities and 125 million triples.

**CN-DBpedia:** CN-DBpedia is another landmark of Chinese knowledge graphs after Zhishi.me. Similar to Zhishi.me, CN-DBpedia is also a large-scale Chinese encyclopedic knowledge graph, which is extracted from Baidu Baike, Hudong Baike and Chinese Wikipedia, the same data sources with Zhishi.me. Different from Zhishi.me, CN-DBpedia, not only extracts structured knowledge from the three largest Chinese online encyclopedias, but also integrates, complements and corrects the extracted knowledge so as to greatly improve the quality of the knowledge graph. In addition, CN-DBpedia is also a constantly updated knowledge graph. There are 9.4 million entities and 80 million triples in CN-DBpedia.

**PKU-PIE:** PKU-PIE is also a Chinese encyclopedic knowledge graph which is constructed by Peking University. It extracts knowledge from multi-sources such as Wikipedia, DBpedia, Baidu Baike, etc. and focuses on integrating and associating them with a self-defined predicate system and categories. The developers of PKU-PIE do not publish any paper or technical report, so that we cannot know the technical details of constructing this knowledge graph. PKU-PIE now contains about 9 million of entities and 40 million triples.

**XLORE:** Most Chinese knowledge graphs such as Zhishi.me, CN-DBpedia and PKU-PIE do not consider cross-lingual knowledge sharing between a Chinese knowledge graph and the one in other languages. The large-scale English-Chinese encyclopedic bilingual knowledge graph constructed by Tsinghua University called XLORE is an important work to solve this problem. It is constructed by extracting semi-structured data from Chinese and English online encyclopedias respectively and mining the equivalent relations between Chinese and English entities. There are more than ten million bilingual entities in XLORE.

**Belief-Engine:** Belief Engine is also an English-Chinese bilingual knowledge graph built from Baidu Baike, Hudong Baike and Wikipedia. Different from XLORE, Belief-Engine only extracts declarative knowledge from these three online encyclopedias and further produces concept-level common knowledge by conceptualizing the declarative knowledge. Each piece of common knowledge will get a belief value in this process. Belief Engine contains around 50 million triples and also does not have any paper to introduce its construction details.

The above-mentioned work focuses on building entity-level knowledge graphs. Different from them, cnSchema and Linked Open Schema are two Chinese schema-level knowledge graphs. cnSchema manually defines thousands of classes, data types, properties and so on. It reuses, interconnects and expands existing schemata of Schema.org, Wikidata, etc. and considers the characteristics of Chinese languages to provide data descriptions and interface definitions. Linked Open Schema is a large-scale English-Chinese bilingual schema-level knowledge graph built from English and Chinese social web sites. It contains more than 700 thousand bilingual concepts and about 2.4 million relations (including *equal*, *subclassof* and *relate* relations) between the extracted concepts in social Web sites.

Currently, there already exists much knowledge of the OBOR initiative in Chinese knowledge graphs such as Zhishi.me, CN-DBpedia, XLORE, etc., which can support different applications regarding OBOR, such as semantic search and question answering. Figure 4 shows a part of the knowledge of OBOR in Zhishi.me, for example, it tells us that the “*initiative country*” of OBOR is “*China*”.

## infobox

节日	12月16日 (“一带一路”国际日)
外文名	One Belt One Road/The Belt and Road Initiative
壮文名	Sai Ndeu Roen Ndeu
含义	丝绸之路经济带、海上丝绸之路
中文名	一带一路
线路	北线；中线；南线；中心线
核心区	新疆、福建
<b>倡议国</b>	<b>中国</b> China
枢纽港口	上海、天津、宁波、广州、福州等
枢纽机场	上海、广州

Figure 4. A part of the knowledge of OBOR in Zhishi.me.

### 3. Techniques of Constructing Chinese Knowledge Graphs

In this section, we focus on introducing the techniques already used in constructing real-world Chinese knowledge graphs. We summarize a general framework of these techniques which is shown in Figure 5. This framework contains four stages: knowledge extraction, knowledge integration, knowledge quality improvement and knowledge update.

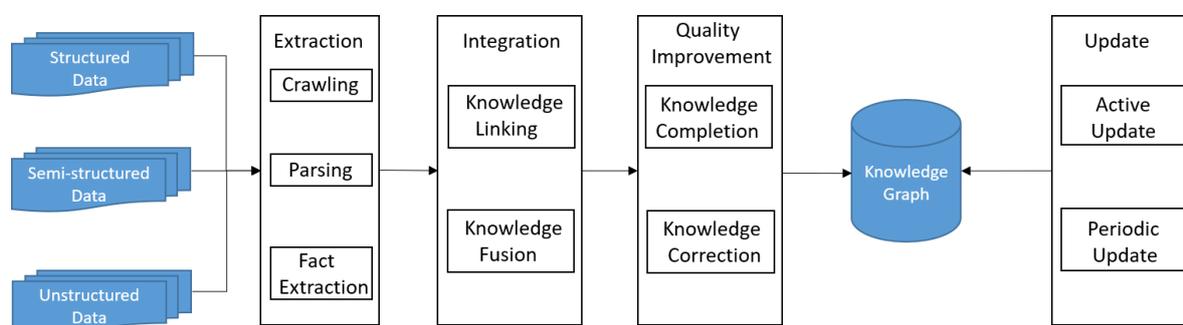


Figure 5. The general framework of Chinese knowledge graph construction.

**Knowledge Extraction:** The first stage of Chinese knowledge graph construction is knowledge extraction. There are mainly three types of data sources, i.e., structured data, semi-structured data and unstructured data, for knowledge extraction. Structured data refers to the data with strict data model structure such as the data stored in tables and relational databases. Semi-structured data refers to data with a certain structure which is not very rigid, such as xml data. Unstructured data often refers to the information that does not have a pre-defined data model, such as text.

**Knowledge Integration:** Most Chinese knowledge graphs extract knowledge from various sources, so it is necessary to integrate such heterogeneous knowledge. Some Chinese knowledge graphs only choose to link knowledge, which is to link equivalent entities, concepts, properties etc. between different sources, and others may further fuse equivalent knowledge after knowledge linking.

**Knowledge Quality Improvement:** Lots of data sources (e.g., Wikipedia, Baidu Baike, etc.) for constructing Chinese knowledge graphs are freely edited by crowds, so mistakes in knowledge graphs are unavoidable. Thus, it is important to improve the quality of data in knowledge graphs through error detection and correction techniques. In addition, knowledge graphs are incomplete and knowledge completion is also necessary for the improvement of data quality.

**Knowledge Update:** Since emerging entities continuously appear in news and other Web streams and the knowledge of existing entities may change as time goes on, update mechanisms are important to keep knowledge graphs more complete and accurate at a specific time point.

In the following subsections, we will describe each stage in detail.

### 3.1. Knowledge Extraction

Most of existing Chinese knowledge graphs, including Zhishi.me, CN-DBpedia, XLORE, etc., extract knowledge from structured or semi-structured data. Taking Zhishi.me as an example, it takes three largest Chinese online encyclopedias, i.e., Baidu Baike, Hudong Baike and Chinese Wikipedia, as its data sources and extract a large number of knowledge from them. As shown in Figure 6, Zhishi.me extracts entities and their related knowledge from structured data (e.g., *Infobox*) and semi-structured data (e.g., *Labels*, *Abstracts*, *Categories* and *Redirects*) in the articles of these three online encyclopedias, and we also list some generated RDF triples which are used to denote knowledge. In addition to Zhishi.me, CN-DBpedia and XLORE both adopt a similar way to extract knowledge from online encyclopedias. Linked Open Schema extracts hierarchical categories and social tags (semi-structured data) as the sources (examples are given in Figure 7) for schema-level knowledge mining, i.e., discovering semantic relations among categories and tags.

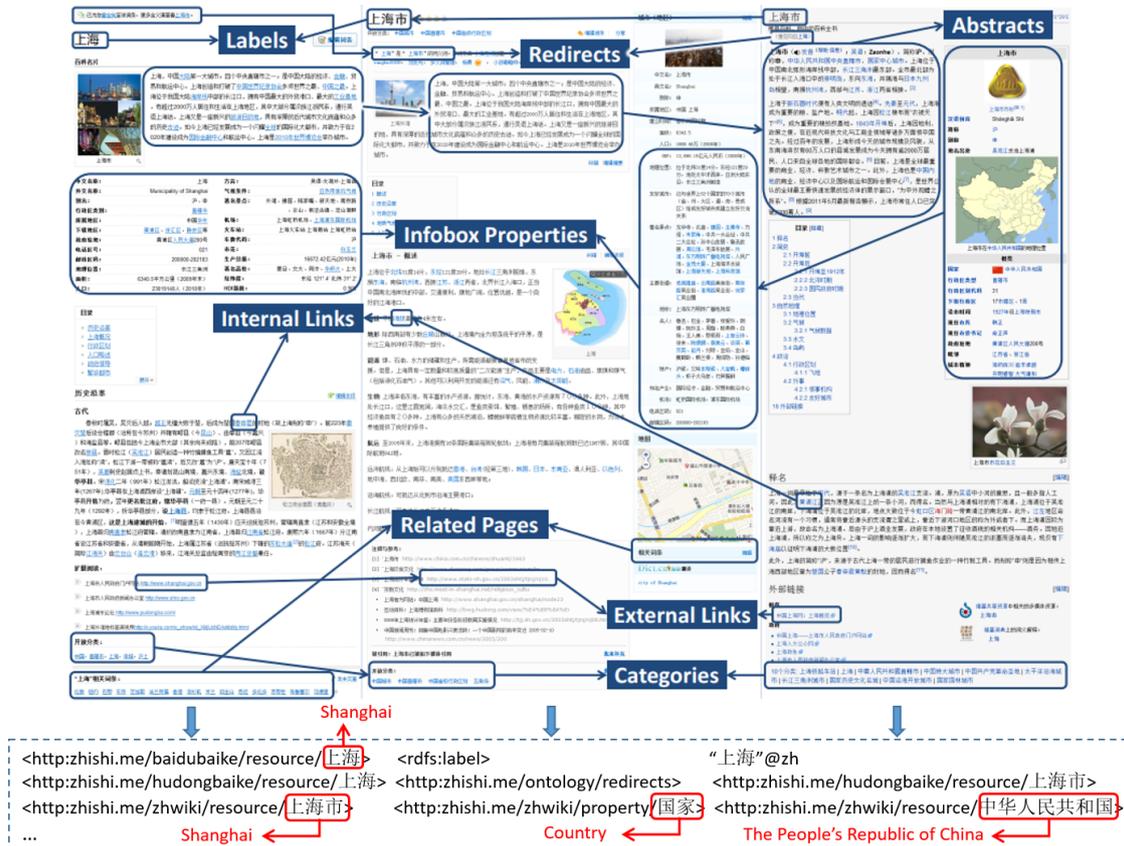


Figure 6. Sample pages from Baidu Baike, Hudong Baike and Chinese Wikipedia.

**大家电** 平板电视 空调 冰箱 洗衣机 家庭影院 DVD/电视盒子  
迷你音响 冷柜/冰吧 家电配件 功放 回音壁/Soundbar  
Hi-Fi专区 电视盒子 酒柜

**厨卫大电** 燃气灶 油烟机 热水器 消毒柜 洗碗机

**厨房小电** 料理机 榨汁机 电饭煲 电压力锅 豆浆机 咖啡机 微波  
炉 电烤箱 电磁炉  
壶/热水瓶 电饼铛  
厨房电器 养生壶/

**生活电器** 取暖电器 净化器  
电风扇 电插座 电话机  
电风扇 冷风扇 其它  
电风扇

**个人防护** 剃须刀 剃/脱毛器  
直发器 按摩椅 拔  
毛器 血糖仪 体温计 耳

**五金家装** 电动工具 手动工具  
电灯 洁身器 水槽  
电铃 电气开关 插座

**tags in NetEase Microblog**

思想(164598)	80后(150209)	音乐(135346)	美食(117911)	旅游(110475)	听音乐(100219)
电影(98796)	摄影(92603)	时尚(89056)	健康(88294)	90后(86369)	爱上网(71477)
乐观(71254)	看书(70016)	看电影(63973)	心理(63888)	善良(61821)	低调(59498)
户外运动(57353)	奋斗ing(57291)	星座(56285)	原创(56044)	财经(49783)	直率(48847)
爱睡觉(48280)	漫画(45165)	亲子(45125)	自由职业(41550)	完美主义(41059)	
自由主义者(40212)	努力ing(39631)	宅女(39611)	学生(39494)	阳光(38646)	幽默(38468)
执着(36734)	宠物(36582)	动漫(36435)	游戏(35557)	70后(34187)	三分钟热度(32632)
学习ing(32472)	宅男(31205)	爱狗(30182)	打篮球(26438)	内讧(25339)	攒钱ing(24575)
温柔(23546)	我宅(23527)	成长ing(23156)	爱做饭(22358)	逛论坛(21995)	减肥ing(21892)

Figure 7. Examples of extracted categories and tags in Linked Open Schema.

### 3.2. Knowledge Integration

Most Chinese knowledge graphs extract knowledge from various sources. For example, Zhishi.me, CN-DBpedia and XLORE extract knowledge from Baidu Baike, Hudong Baike and Chinese Wikipedia. Categories and tags in Linked Open Schema are extracted from different kinds of social Web sites in different languages. Thus, it is necessary to integrate such heterogeneous knowledge. Existing Chinese knowledge graphs integrate knowledge in two ways: knowledge linking and knowledge fusion.

#### 3.2.1. Knowledge Linking

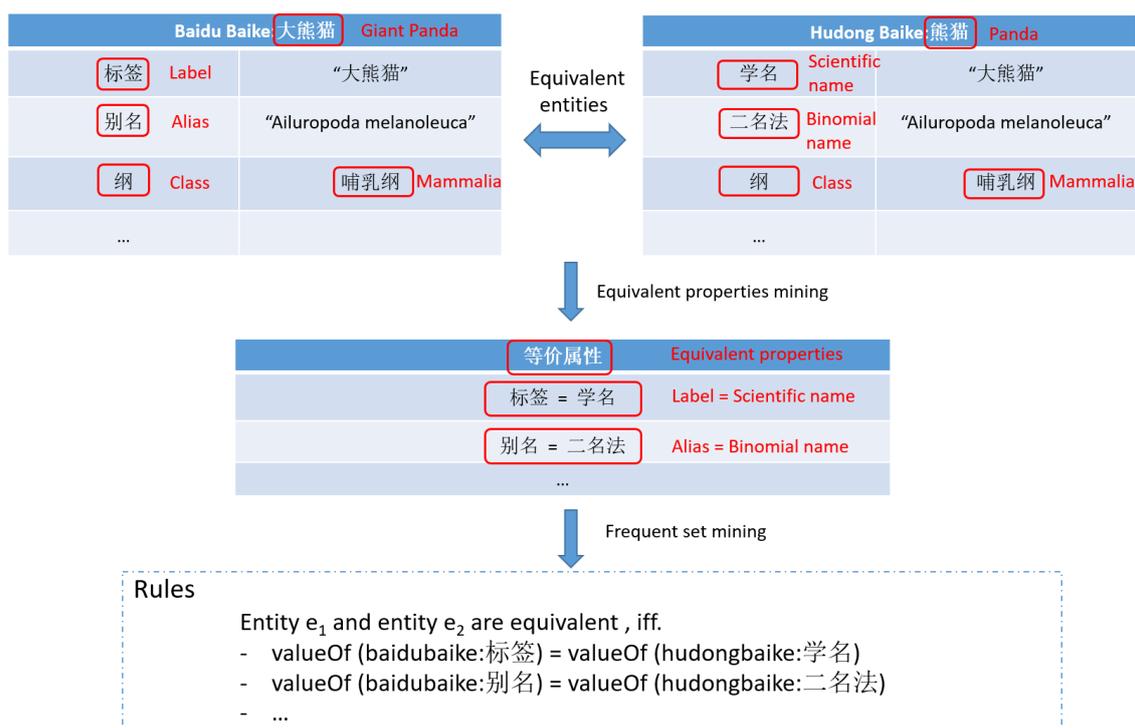
Knowledge linking aims to discover semantic relations (e.g., equivalence relations between entities and subClassOf relations between concepts) between entities or concepts from different sources. Such relations are taken as links.

##### Monolingual Knowledge Linking

Niu et al. [17] proposed a semi-supervised learning method based on the EM algorithm to find equivalent entities across different online encyclopedias in Zhishi.me. This is a new technique which is not used in constructing English encyclopedic knowledge graphs, because they only rely on English Wikipedia but Zhishi.me extracts knowledge from three Chinese online encyclopedias. This method uses a small number of existing equivalent entities as seeds, then iteratively mines and refines matching rule sets from these seeds and find new equivalent entities of high confidence using these rules.

In this method, a small number of existing equivalent entities linked by the relation **owl:sameAs**, as well as their properties, are used as seeds to discover more equivalent entities. For each pair of existing equivalent entities, the method firstly tries to mine equivalent properties from them. Then frequent sets, each of which consists of several pairs of equivalent properties, are mined from these equivalent properties through association rule mining. The mined frequent equivalent property sets are used for building matching rules and applied on the whole data set to find new equivalent entities. For example, suppose we have a pair of equivalent entities (see Figure 8) from Baidu Baike

and Hudong Baike. Both of the entities refer to the panda. Given all of the properties of the two entities, equivalent properties can be discovered by equivalent property values (i.e., the same or synonymous values). For example, in Figure 8, property “标签 (Label)” in Baidu Baike is equivalent to property “学名 (Scientific name)” in Hudong Baike, and “别名 (Alias)” in Baidu Baike is equivalent to “二名法 (Binomial Name)” in Hudong Baike. If the equivalent property set (consisting of the above equivalent properties pairs) is a frequent set, which means that it can also be mined from many other equivalent entities, then it will be used to construct matching rules. One example of the matching rules is also given in Figure 8, which indicates that entity  $e_1$  and  $e_2$  are equivalent if and only if the values of corresponding equivalent properties in the frequent set are all equivalent.



**Figure 8.** An example of equivalent properties and matching rules mining.

Newly found equivalent entities with high confidences will be added to the set of seeds and the method starts the next iteration. The iteration procedure of discovering rules as well as equivalent entities follows the Expectation-Maximization (EM) algorithm. The E-step discovers new equivalent entities using matching rules, and the M-step mines new matching rules from the equivalent entities discovered in E-step by maximizing the precision of these rules. The algorithm iterates until no new equivalent entities can be found any more. Semi-supervised learning based on EM algorithm automatically discovers dataset-specific matching rules while preserving high accuracy and good coverage. These derived rules have the capacity to find the most discriminative data characteristics for different data sources, so it is not domain-specific or dataset-specific.

Wu et al. [14] proposed a semi-supervised learning method based on self-training to schema-level knowledge linking in Linked Open Schema. This method uses a small number of initial labeled data and the self-training algorithm to mine equivalence, subClassOf and relate relations among the categories and tags from different social Web sites. In each iteration of the self-training process, the proposed method learns an SVM classifier. Then the classifier is applied to the unlabeled data. Here, rules are utilized to revise the misclassified results. Afterwards, newly found relations will be added to the labeled data. The whole process will terminate if the predicted labels do not change any more. To measure the relatedness of concept pairs (i.e., a pair of categories, or tags, or a category and a tag), different features are designed for training the SVM classifier. This method jointly exploits

the advantages of rules and machine learning techniques which can acquire better learning results. The self-training algorithm is simple but quite efficient and suitable for large-scale knowledge linking. The drawback of this method is that it needs manually designed features for training and different features needs to be re-designed in different application scenarios.

### Cross-Lingual Knowledge Linking

Cross-lingual knowledge linking is also an important task of knowledge linking and it can largely enrich the cross-lingual knowledge and facilitate knowledge sharing across different languages. This is a new task which no longer only depends on the cross-lingual links between Wikipedia pages to link knowledge of different languages.

Wang et al. [18] proposed an approach for building cross-lingual links across the online encyclopedias of different languages. The approach uses only language-independent features of articles (an article corresponds to an entity) and employs a linkage factor graph model to predict new cross-lingual links between entities. This approach is used in XLORE. In this work, after analysing the existing cross-lingual links, several factors including link homophily, category homophily and author interests are found to be helpful in cross-lingual entity linking. For example, if the articles of two entities share more common links (outlinks and inlinks) and categories, and authors of these articles share more common interests, these two entities are more likely to be equivalent. Hence, this method designs four features to measure the similarities of common outlinks, inlinks, categories and author interests between the articles of two given entities in different languages. These features are used to construct the node feature function of a linkage factor graph model. In addition, the edge feature function and constraint feature function are also defined. Edge feature function considers the relations between nodes, which is based on the assumption that if an entity links to other two equivalent entities, then these three entities tend to be equivalent to each other. Constraint feature function defines 1-to-1 linking constraint on all relations between nodes. Based on the linkage factor graph model, the joint distribution is defined over the three feature functions. Given a set of labeled data, the parameters can be learned, and new cross-lingual links between entities can be predicted with this model.

Wang et al. [19] also argued that when there are insufficient existing cross-lingual links between entities as the training data, discovering new cross-lingual links becomes more challenging. Therefore, they proposed an approach that boosts cross-lingual entity linking by concept annotation. This approach is also used in XLORE. In this approach, concept annotation aims to identify important concepts in the article of the given entity and link them to the corresponding articles in the same online encyclopedia. It first extracts key concepts in each article by matching all the n-grams in the input article with the elements in a controlled vocabulary, and then uses a greedy disambiguation algorithm to match the corresponding articles of the key concepts. Matched articles will be linked to the article, thus enriching the inner links in an online encyclopedia. After concept annotation, several link-based features, which measure the similarities of common links (outlinks and inlinks) and categories between two articles, are designed for predicting new cross-lingual links. A regression-based learning model is proposed to learn the weights of different similarity features, which are used to rank the candidate cross-lingual links. To find as many new cross-lingual links as possible, the method adopts an iteration framework which allows the concept annotation and cross-lingual link prediction to mutually reinforce.

### 3.2.2. Knowledge Fusion

Different from knowledge linking, knowledge fusion aims to integrate knowledge from different data sources rather than link them to each other. For the purpose of fusing equivalent knowledge in different data sources, CN-DBpedia utilizes a normalization module to normalize entities, attributes and attribute values which have the same meaning but different string labels. For example, in Baidu Baike, “生日(birthday)” is used to describe the birth time of a person, while in Hudong Baike, “出生日期 (birth date)” describes the same attribute. After using the normalization module,

“生日 (birthday)” is used to describe this attribute in a unified way. In many scenarios, an integrated knowledge graph is more convenient and efficient for use than several knowledge graphs linked to each other. Hence, knowledge fusion strategies are applied in some Chinese knowledge graphs (e.g., CN-DBpedia). However, there may exist conflicts when fusing knowledge from different data sources and it may introduce errors into knowledge graphs and need many human efforts. Current strategies applied to knowledge fusion in constructing Chinese knowledge graphs are still quite simple, and some language-independent approaches [20,21] proposed in English knowledge fusion can be considered to be used in this part.

### 3.3. Knowledge Quality Improvement

#### 3.3.1. Knowledge Completion

Chinese knowledge graphs built from various sources are incomplete in terms of type information, infobox information and so on. Therefore, completion is necessary for the quality improvement of these knowledge graphs. Knowledge graph completion techniques include type inference and infobox completion which will be introduced in detail in this subsection.

#### Type Inference

Type information refers to the axioms stating that an entity is of a certain type. These axioms can be denoted as triples, each of which is composed of a TypeOf relation linking from a class to an entity, e.g., “*President of the United States*” TypeOf “*Donald Trump*” and “*Country in Europe*” TypeOf “*Italy*”. Type information can also benefit many applications in different research fields, such as entity search [22,23] and question answering [24].

**Attribute-Driven Type Inference:** Existing methods of type inference in Wikipedia to build English knowledge graphs rely on linguistic rules, which cannot be applied to Chinese knowledge graphs. Thus, Wu et al. [25] proposed a new language-independent approach to entity type inference in Wikipedia, and it is used in Zhishi.me. This approach is based on an attribute-driven type inference assumption: “*in Wikipedia, if an instance contains representative attributes of one of its classes, there may exist a TypeOf relation from the class to the instance with high probability*”. For example, given attributes “*actors, release date, director*” of a certain entity, people may infer that the type of this entity is “*Movie*”. However, given attributes “*name, foreign name*”, people cannot infer the type of the entity because too many classes have the attribute “*name*” or “*foreign name*”. The proposed approach first extracts the attributes of entities from infoboxes in articles of Wikipedia. Considering that the attributes of many entities are still incomplete or missing, for each entity, it utilizes a vector similarity metric and a class similarity metric to acquire its most similar entities which already have attributes, and the attributes of the most similar entities are used to complement attributes of the given entity. Then, it adopts a general algorithm incorporating several language-independent rules to iteratively get the attributes of classes. Finally, it uses a random graph walk model to compute the probability of a TypeOf relation existing between a class and the given entity. Attribute-driven type inference adopts a language-independent feature, i.e., attributes, to build the semantic associations between entities and classes. It can infer the types of entities of different languages from online encyclopedias such as the multilingual Wikipedia.

Another work [26] has a very similar idea with [25]. It also applies the above attribute-driven type inference assumption and designs a random graph walk model to compute the probability of a TypeOf relation existing between a class and the given entity. The differences are that (1) this work does not mine more attributes of entities from other similar entities, and (2) it also proposes several language-dependent heuristics to infer entity types from abstracts and infoboxes of the articles in Chinese online encyclopedias.

**Cross-Lingual Type Inference:** Xu et al. [27] proposed a cross-lingual type inference approach which types Chinese entities with some widely accepted taxonomies in English, such as the taxonomy of DBpedia. This is a new task which does not exist in constructing English knowledge graphs, because

many non-English knowledge graphs fail to type their entities due to the absence of a reasonable local hierarchical taxonomy. Building a widely accepted taxonomy requires huge human efforts, so the proposed approach can greatly reduce human efforts and boost type inference by typing Chinese entities with English types in well-defined taxonomies. This approach first exploits existing cross-lingual links between Chinese and English entities to construct the training data, and then learns a hierarchical classification model to predict the English types (from the taxonomy of DBpedia) of a given Chinese entity. This approach is used in CN-DBpedia.

In the process of constructing training data, if a Chinese entity and an English entity are already known as equivalent or there exists a cross-lingual link between these two entities in Wikipedia, the Chinese entity will be labeled with the types of the English entity. For example, the Chinese entity “威廉莎士比亚” is labeled with the types of entity “*William Shakespeare*”, i.e., “*English male stage actors*”, “*English writers*”, “*English poets*” etc. In this way, a massive amount of labeled data is generated as training data. Then, a supervised hierarchical classification model is built, which accepts an untyped Chinese entity (with its features) as input and outputs all valid English types in the taxonomy of DBpedia. We use an example in Figure 9 to illustrate the model. In the training phase, it first trains a binary classifier for each type in the hierarchy except the root node “*Thing*”. The binary classifier of each type uses all entities belonging to this type as positive examples and those entities belonging to this type’s sibling-type or super-type as negative examples. For example, the classifier of type “*Actor*” uses all entities belonging to this type as positive examples. Entities belonging to sibling types “*Writer*” and “*Poet*” as well as those belonging to super-type “*Person*” are taken as negative examples. After training the classifier of each type in the taxonomy, the hierarchical classification model is made up of these binary classifiers in a hierarchy structure.

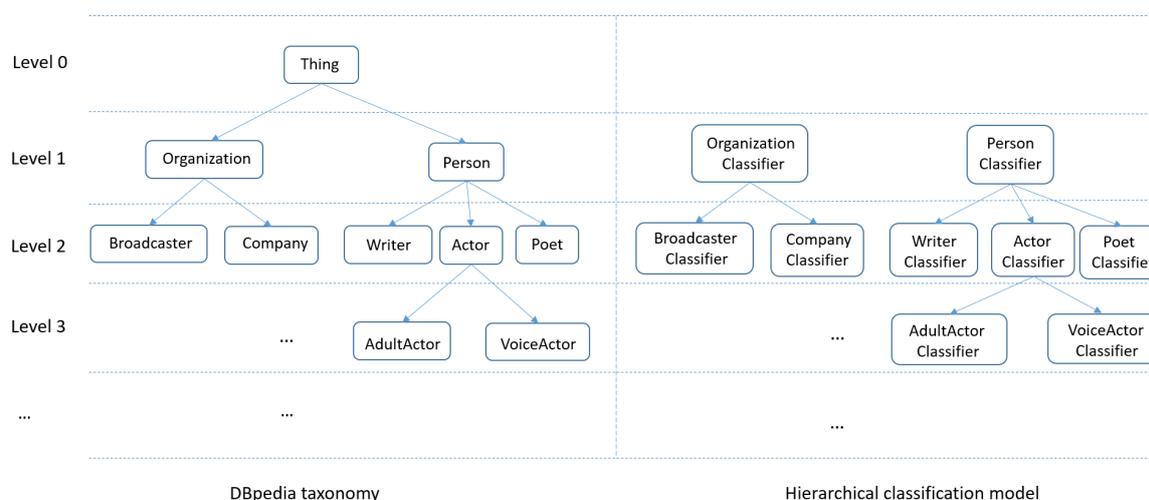


Figure 9. Part of DBpedia taxonomy and hierarchical classification model.

When using this model to predict types for a given entity, it traverses the DBpedia taxonomy in a top-down manner to search for types of this entity. It first visits the types in the first level of the taxonomy which are the sub-types of root node “*Thing*” and add them to the candidate types set. Then for each candidate type, the approach runs the classifier of the type to test whether the given entities belong to this type or not. If the result is true, it adds all its sub-types to the candidate type set. The process ends when no more candidate types can be processed. All types, which are classified as true for the given entity during this process, are attached to the entity as its types.

### Infobox Completion

Infoboxes explicitly contain structured facts of entities in terms of triples (each one is denoted as  $\langle \text{subject}, \text{predicate}, \text{object} \rangle$ ), such as  $\langle \text{Yao Ming}, \text{Height}, 2.29 \text{ m} \rangle$ . As shown in Figure 6, Zhishi.me

extracts properties from the infoboxes in Baidu Baike, Hudong Baike and Chinese Wikipedia respectively. Similarly, CN-DBpedia and XLORE both extract infobox information from these three online encyclopedias. However, structured facts about entities in infoboxes are incomplete, and many facts implicitly exist in the articles corresponding to the entities in online encyclopedias, so infobox completion from unstructured data is also an important part of knowledge completion.

Xu et al. [11] presented a hybrid Long Short-Term Memory Recurrent Neural Network framework to extract knowledge from free text and this approach is used in CN-DBpedia for infobox completion. CN-DBpedia treats the infobox completion task as extracting the object for a given pair of subject (i.e., entity) and predicate (i.e., property) from the article corresponding to the given entity. This approach models the extraction problem as a sequence to sequence learning problem. The input is a sentence containing tokens, and the output is the label of each token. The label is either 1 or 0, predicting whether a token is a part of the object for the predicate. For example, the sentence in Figure 10 is taken from the article corresponding to the entity “Yao Ming” in Baidu Baike. The extractor of “BirthPlace” will label “Shanghai” and “Xuhui District” as true for “Yao Ming”. To relieve the burden of human efforts, this approach also adopts a distant supervision method to automatically construct training data. It constructs sentences that express the facts in the infobox as its training data. For example, if  $\langle \text{Yao Ming, Height, 2.29 m} \rangle$  occurs in the infobox of Yao Ming, this method will construct the sentence “The height of Yao Ming is 2.29 m.” and correctly label the object in the sentence. To cover as much as possible useful information, the approach trains a hybrid representation of words, phrases and types for each token in the input token sequences. However, training the Long Short-Term Memory Recurrent Neural Network model has very high time complex since lots of parameters need to estimate.

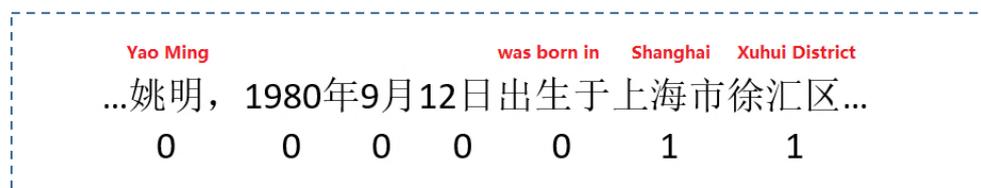


Figure 10. An example for extracting objects from text.

### 3.3.2. Knowledge Correction

Baidu Baike, Hudong Baike and Chinese Wikipedia are all collaborative online encyclopedias where the content is edited freely by users. There inevitably exist mistakes and inconsistencies in these online encyclopedias. Therefore, it is very important to improve the quality of knowledge extracted from these sources through knowledge correction. The process of knowledge correction consists of two stages, i.e., error detection and error correction.

Xu et al. [11] proposed two methods for error detection and apply them in CN-DBpedia. The first is rule-based error detection. This method uses a lot of predefined rules to detect errors which do not conform to the rules. For example, it uses the domain and range of properties to find errors. The range of attribute “birthday” should be date and any other types of values are incorrect. The second method is based on user feedbacks. CN-DBpedia provides a user exploration interface to browse the knowledge. The interface allows users to provide feedbacks on the correctness of the facts in CN-DBpedia (see Figure 11).

Q InfoBox				
Chinese Name	中文名	一带一路	One Belt One Road	<input type="button" value="👍"/> <input type="button" value="🗑️"/>
Initiative Country	倡议国	中国	China	<input type="button" value="👍"/> <input type="button" value="🗑️"/>
English Name	外文名称	One Belt One Road/The Belt and Road Initiative		<input type="button" value="👍"/> <input type="button" value="🗑️"/>
Hub Airport	枢纽机场	上海	Shanghai	<input type="button" value="👍"/> <input type="button" value="🗑️"/>
Hub Airport	枢纽机场	广州	Guangzhou	<input type="button" value="👍"/> <input type="button" value="🗑️"/>

**Figure 11.** The interface of CN-DBpedia for providing feedbacks.

After error detection, Xu et al. also presented an error correction method based on crowd-sourcing. Error facts are assigned to different contributors of crowd-sourcing for correction. Then it adopts a simple-yet-effective method, i.e., majority vote, to aggregate the multiple and noisy contributor inputs to create consistent outputs.

### 3.4. Knowledge Update

Emerging entities appear rapidly and the knowledge relevant to entities may also change over time. There will be some outdated knowledge existing in knowledge graphs overtime without update. It is important to prevent knowledge graphs from being obsolete through adopting update mechanisms.

**Periodical update:** Periodical update is adopted by most existing Chinese knowledge graphs such as Zhishi.me, XLORE, etc. It refers to replace the knowledge graph with a new version after a period of time. The problem of periodical update is that if the update cycle is too long, so there will be a large amount of outdated knowledge existing in the knowledge graph. If the update cycle is too short, there will be a huge consumption of computation resources.

**Active update:** To overcome the drawbacks of periodical update, active update techniques are used to update knowledge graphs at a higher frequency and lower cost. Most knowledge graphs do not adopt active update strategies. DBpedia is updated actively via the update stream provided by Wikimedia foundation. However, most of Chinese online encyclopedias do not provide such update stream. Hence, Liang et al. [28] proposed a new active update approach, and it is used in CN-DBpedia. This approach is based on the assumption that hot entities and those entities semantically related to them are more likely to change or to be new entities than others in knowledge graphs. In this approach, it firstly extracts hot entities from the titles of hot news, popular search keywords of search engines, hot topics of social networks and so on. Then it checks whether these entities need to be updated or whether they are new entities which do not exist in the current knowledge graph. If the result is true, the information of these entities in online encyclopedias will be synchronized (updated or inserted) to the knowledge graph. To obtain more hot entities, the approach uses hyperlinks in the newest articles corresponding to existing hot entities to find more relevant entities. Here gives an example of active update, the news title “Donald Trump was elected as President of the United States” contains an entity “Donald Trump”, and the attribute value of “profession” needs to be updated as “President of the United States”.

To improve the efficiency of update with limited resources, this approach trains a supervised learning predictor based on the history change frequencies of all entities in the knowledge graph and uses it to predict the future change frequencies of hot entities. When resources is limited, hot entities with higher predicted change frequencies will be synchronized to the knowledge graph first. Since most entities in the knowledge graph do not need to be updated, it will cost fewer computation resources when only updating a very small part of entities in the knowledge graph. A low consumption of computation resources in each update cycle allows this approach to update the knowledge graph with

a higher frequency. However, the amount of updated knowledge in each cycle still accounts for a small proportion of all outdated knowledge in the knowledge graph, compared to that in periodical update.

#### 4. Applications of Chinese Knowledge Graphs

At present, Chinese knowledge graphs have shown their important roles in many different intelligent applications, many of which are relevant to OBOR and will be introduced in the following subsections. Note that knowledge graphs in different languages can be applied to semantic search, question answering, intelligence analysis, decision making, and other applications in different languages, respectively. These applications in different languages have a similar way to use knowledge graph, and the only difference exists in some applications which need linguistic techniques to support themselves. For example, question answering in different languages needs to utilize different methods to parse the syntactic structure of a given textual question.

##### 4.1. Semantic Search

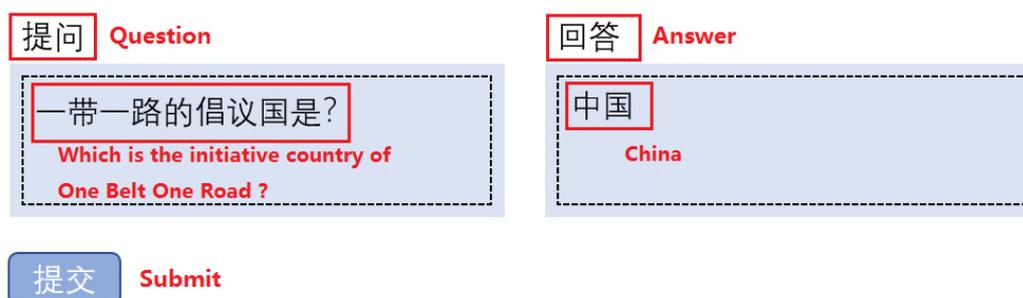
Chinese knowledge graphs are capable of providing high-quality structured background knowledge and common-sense knowledge to aid search engines to better understand the user query intents, thus providing more concise and intelligent search results. This technology overcomes the limitations of the traditional keyword-based search model and converts the Web-based search to the semantic search. This makes the search engines more intelligent and effectively improves the quality of the search results. Now most of the Chinese search engines such as Baidu (<https://www.baidu.com>), Sogou (<https://www.sogou.com>) and (<https://www.so.com>) have built their own knowledge graph to enhance search experience of users. As shown in Figure 12, when you search “一带一路 (One Belt One Road)” in Baidu search engine, in addition to the web pages related to OBOR shown on the left side, members of OBOR will also be recommended to users on the right side. This is the application of knowledge graph techniques. The search engine firstly identifies the entities in users’ search keywords, then obtains the knowledge relevant to these entities in knowledge graphs and finally returns the structured knowledge to users.



Figure 12. Search results of “One Belt One Road” in Baidu search engine.

#### 4.2. Question Answering

Intelligent question answering (QA) is an important application of artificial intelligence. It can provide users with the answers for their questions with accurate and concise natural language. It can greatly improve efficiency and reduce cost of manual intervention. This kind of dialogic information acquisition relies on the accuracy and reliability of the QA system. Knowledge graph can help improve the intelligence of robots, IOT and other devices by providing high-quality background knowledge. As shown in Figure 13, when submitting a question “一带一路的倡议国是? (Which is the initiative country of OBOR?)” to the QA system supported by a Chinese knowledge graph, this system will firstly use natural language processing technologies to convert the problem to mining what is the attribute value of attribute “倡议国 (initiative country)” for the entity “一带一路 (OBOR)”, then obtain the attribute value “中国 (China)” in the given knowledge graph and finally return the answer to users. Currently, many researchers are working on the techniques of Knowledge Base Question Answering (KBQA) [29,30] and a lot of QA platforms have been developed in China, and they have already introduced in knowledge graphs to ensure a better user experience, such as robot “Xiaodu” from Baidu Company, robot “Xiaomi” from Alibaba Company, intelligent voice assistant “Siri” from Apple Company and Microsoft’s robot “Cortana”.



**Figure 13.** An example of Chinese QA system supported by knowledge graph.

#### 4.3. Intelligence Analysis and Decision Making

Knowledge graph can represent and link a variety of types of data, such as text, multimedia, sensors, etc., which builds associations among different kinds of big data. Such associated big data can be easily used to perform information analysis and assist decision making. Knowledge graph is now playing a very important role in intelligence analysis. As shown in Figure 14, in the scenario of the anti-fraud intelligence analysis, we can build an intelligence knowledge graph by fusing the information from different data sources and adding business rules established by domain experts. Potential fraud risks can be identified through data inconsistency detection in the knowledge graph. For example, borrower Mr. Zhang and borrower Mr. Wu claimed that they are colleagues when filling in their personal information, but the company names they filled are different. Besides, the phone number of Mr. Zhang also belongs to another borrower Mr. Zhen. These inconsistencies indicate the potential fraud so that the bank will decline the requests of loan respectively from Mr. Zhang, Mr. Wu and Mr. Zhen. In addition, knowledge graph has already been applied in many other kinds of intelligence analysis to assist decision making such as police intelligence analysis, stock intelligence analysis, etc.

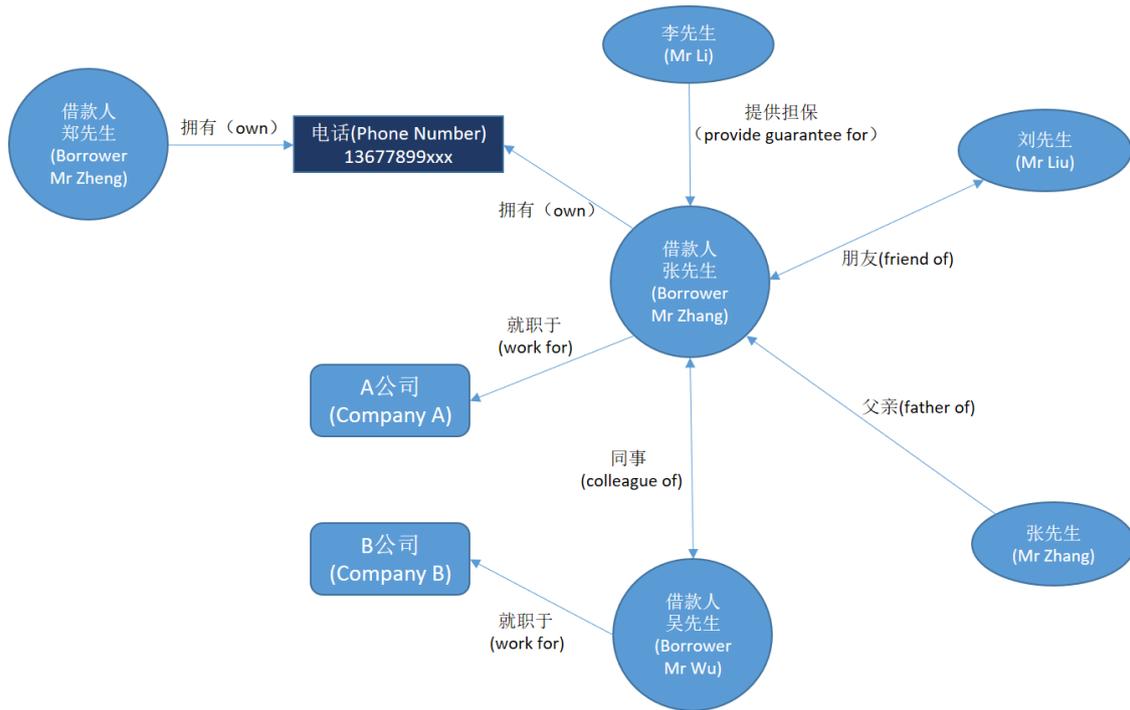


Figure 14. An example of intelligence analysis: anti-fraud.

4.4. Information Integration

With the development of World Wide Web, people can acquire a great deal of information from plenty of different data sources, which leads to the problem of information heterogeneity and it becomes more difficult to integrate such information according to user requirements by traditional methods. Knowledge graph can represent, link and integrate different kinds of data from various data sources, which makes it become an effective tool of information integration. In recent years, China has become the center of global manufacturing with the development of China’s economy. As the scale of manufacturing information in various sources is in an explosive growth, we can integrate such information from different data sources into one knowledge graph for the convenience of utilizing and visualizing. Figure 15 gives a part of the knowledge graph of equipment manufacturing.

Here, we also give another example regarding OBOR policies in Figure 16. After the OBOR initiative was proposed, Chinese government has published a lot of policies in different occasions, such as international conferences, press conferences, official announcement and so on. We can utilize natural language processing techniques to extract entities and relationships from these policies and use them to compose an OBOR knowledge graph. In this way, the information related to OBOR policies, which is distributed over different sources, is integrated together. Figure 16 visualizes a part of our constructed knowledge graph for OBOR policies.

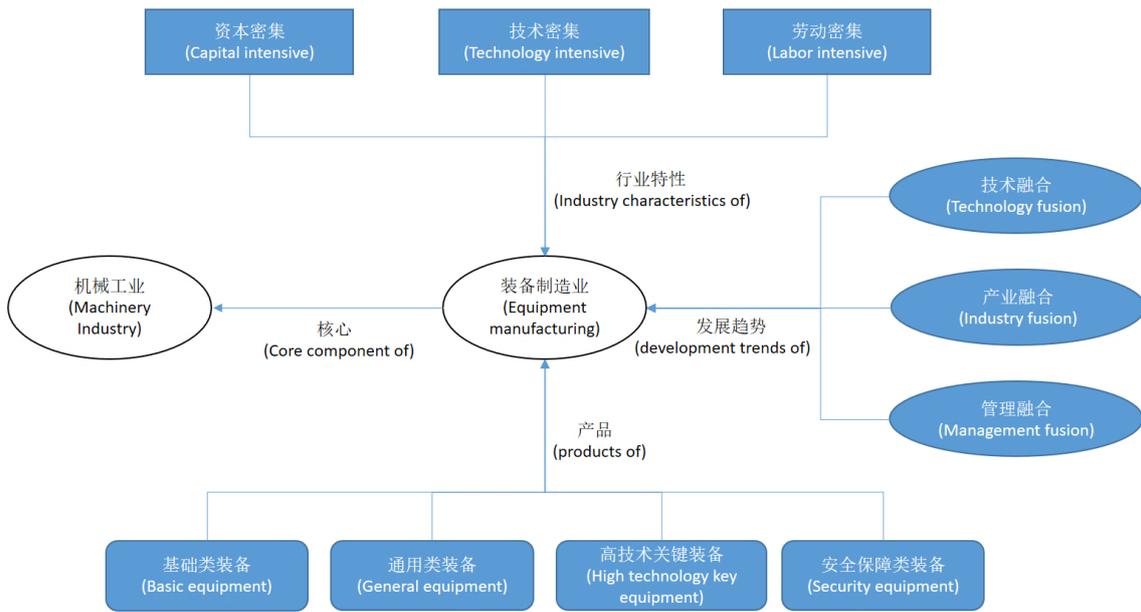


Figure 15. Part of the knowledge graph for equipment manufacturing.

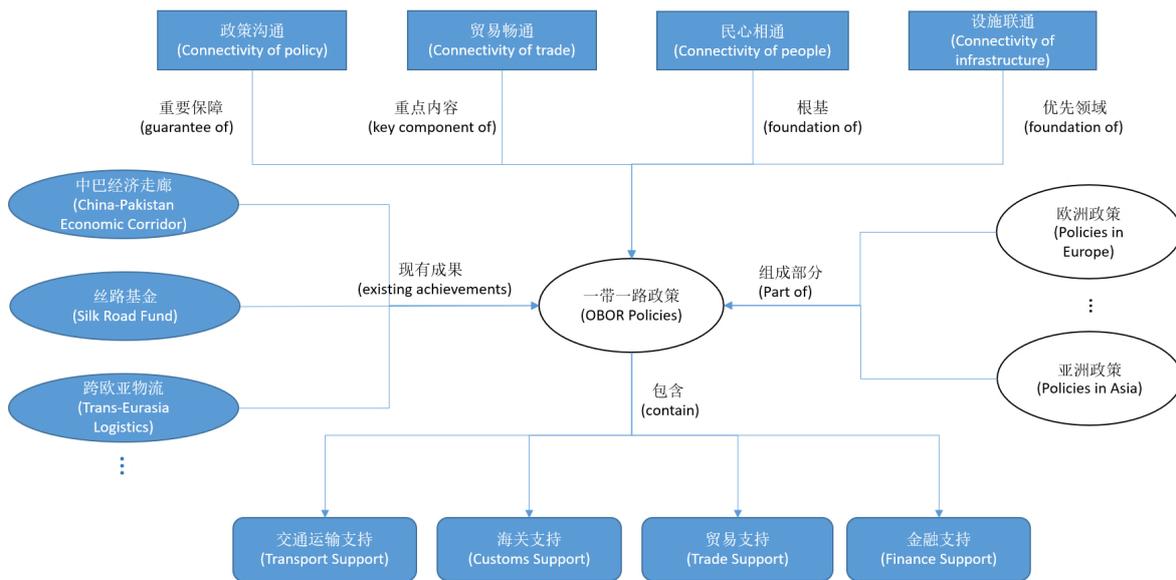


Figure 16. Part of the knowledge graph for OBOR policies.

4.5. Applications in other Vertical Domains

In the financial field, knowledge graph is promoting the development of intelligence on multiple scenarios. We can formulate marketing strategy by integrating customers’ data from different sources to build customer knowledge graphs. Through extracting data from financial reports, the prospectus, company announcements etc., we can construct company related knowledge graphs to assist financial researchers to make investment decisions. By modeling various factors that have effect on stock, we can build stock knowledge graphs. At present, a lot of financial companies in China have introduced knowledge graph technologies to get the first place of financial AI. Taking Company ZTE as an example, after the news of the USA restricting exports to ZTE was released, if we have the knowledge graph of ZTE and its related companies including suppliers, partners and competitors (see Figure 17), we can quickly screen out affected listed companies and mine investment opportunities or control the investment risk.

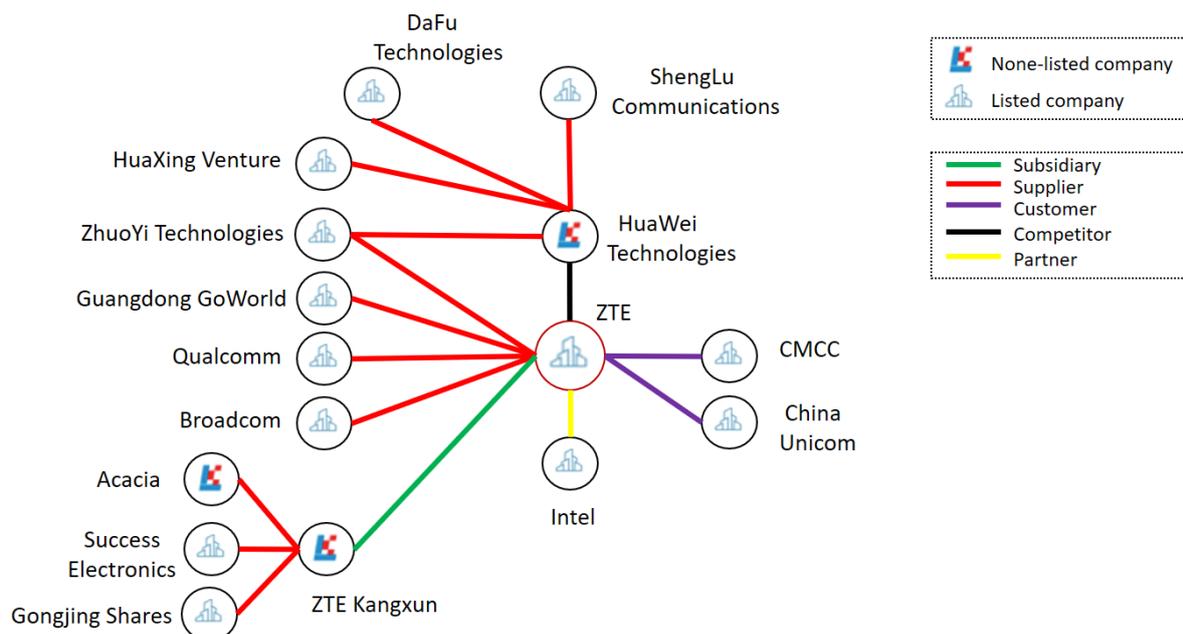


Figure 17. Part of the knowledge graph related to ZTE.

In the field of e-commerce, many companies (e.g., Alibaba) are also deploying knowledge graph technologies. An E-commerce platform not only contains huge number of goods, but also contains products, manufacturers, suppliers and other kinds of objects related to these goods. How to effectively manage and model these objects and their relationships has become an important challenge. Knowledge graph can effectively model and manage the knowledge for its flexible and powerful ontology modeling ability. Alibaba's knowledge graph takes the goods, standard products, brands and standard classifications as the core and utilizes entity recognition, entity linking and semantic analysis technologies to integrate heterogeneous knowledge. It is a huge commodity knowledge graph containing billions of triples. It is widely used in Alibaba's core business such as commodity search, shopping guidance, platform management, intelligent QA, branding etc. It greatly boosts the development of the e-commerce platform of Alibaba.

In other fields such as social network, biomedical science and the Internet of things, knowledge graph is also playing an increasingly important role. This is because of wide applications of knowledge graph so that it has become so popular in recent years.

#### 4.6. OpenKG: A Chinese Open Knowledge Graph Community

OpenKG (<http://www.openkg.cn/>) (see Figure 18), a Chinese open knowledge graph community, aims to promote the open access and interconnection of Chinese knowledge graphs. So far, there have been 56 institutions from industry and academia contributing their own knowledge graphs. It has attracted 80 high-quality knowledge graphs in total including Zhishi.me, CN-DBpedia, XLORE, etc., and covers 15 different kinds of knowledge graphs such as encyclopedia, finance, health care, agriculture, business, social network, etc. Besides, there also 45 different open tools used for knowledge graph construction, integration, completion and reasoning. OpenKG integrates the most abundant knowledge graph resources in China.

The OpenKG community also aims at interlinking different Chinese knowledge graphs to promote the open and interconnection of Chinese knowledge graphs. For example, it has already established links between the major large-scale Chinese encyclopedic knowledge graphs including Zhishi.me, CN-DBpedia, XLORE, Belief-Engine, PKU-PIE, etc. This greatly promotes the open and interconnection of Chinese general encyclopedic knowledge graphs. These linked encyclopedic knowledge graphs (<http://link.openkg.cn>) are open to the public and users can download the whole linked dataset

from the website of OpenKG or visit data through the interface provided by OpenKG. In the future, the OpenKG community will continue to link knowledge graphs in vertical fields.



Figure 18. OpenKG: A Chinese open knowledge graph community.

## 5. Impacts of Knowledge Graph on OBOR

Since a lot of people from the countries along OBOR are not familiar with the OBOR initiative and its related policies, applications supported by knowledge graphs can help them know more about OBOR quickly and easily. After extracting and integrating knowledge of OBOR from various sources, a knowledge graph on OBOR can be built which is shown in Figure 16. In the following subsections, we will list two examples to demonstrate the potential impacts of knowledge graph for OBOR.

### 5.1. Semantic Search on OBOR

When people want to know more about OBOR, such as what OBOR is, its related policies, its benefits and so on, they may choose to search “OBOR” in search engines. However, traditional keyword matching-based search engines will return a lot of web pages related to OBOR to users and it may take great efforts for users to find their desired information. In contrast, knowledge graphs integrate knowledge from different sources together and knowledge graph-based search can return the most related results to users directly. As shown in Figure 19, when users search “*what is One Belt One Road policy*” in Google, it will return the definition of OBOR in Google Knowledge Graph to users which can help users quickly know what OBOR is. In addition, knowledge graphs are very helpful to people who want to know more about OBOR but do not know what to search for. All relevant knowledge of OBOR which is integrated in the knowledge graph can be returned to users for reference. For example, when users search “*One Belt One Road policy*” in Bing search engine (see Figure 20), it will return relevant knowledge of OBOR (i.e., contents, oversight, infrastructure network, etc.) in a structured table. People can easily get a detailed understanding of OBOR policies by searching “*One Belt One Road policy*” only.

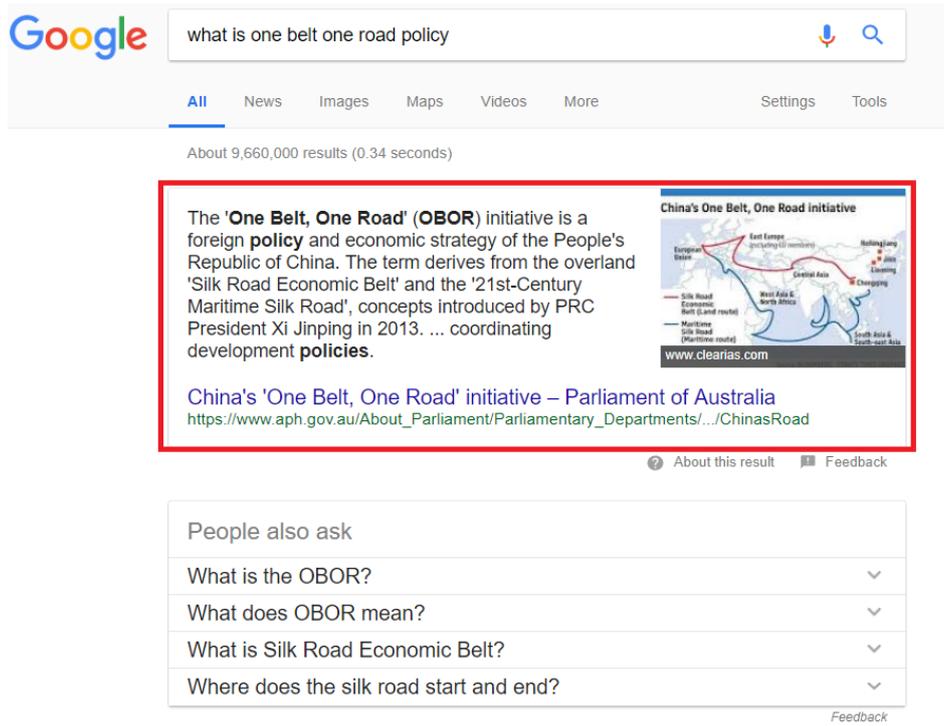


Figure 19. Search results of “what is OBOR policy” in Google.

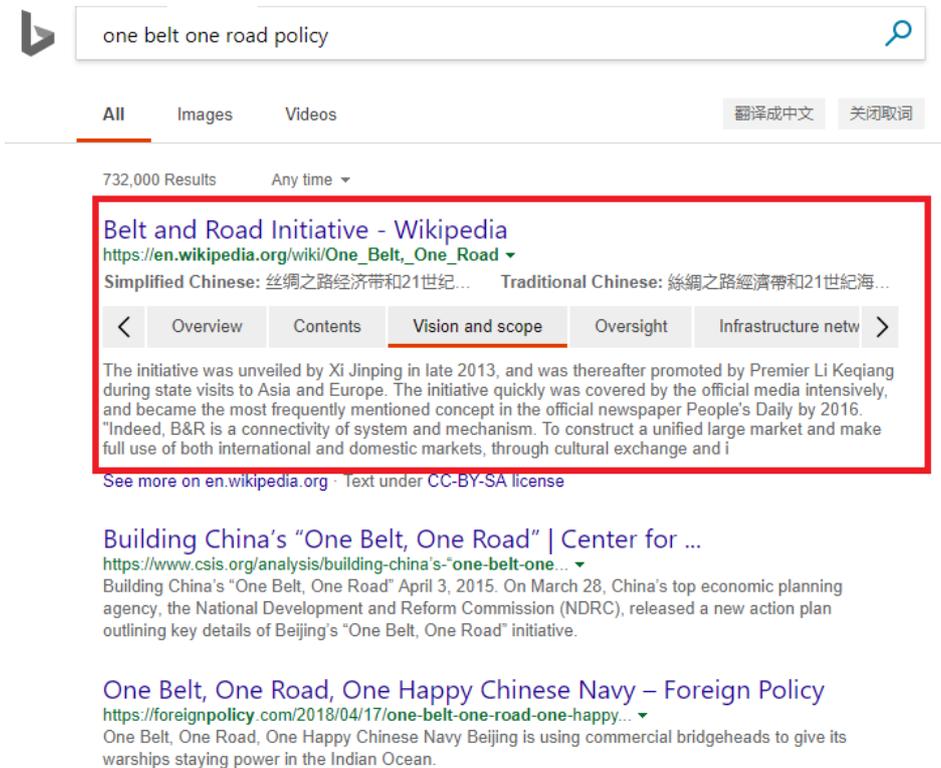
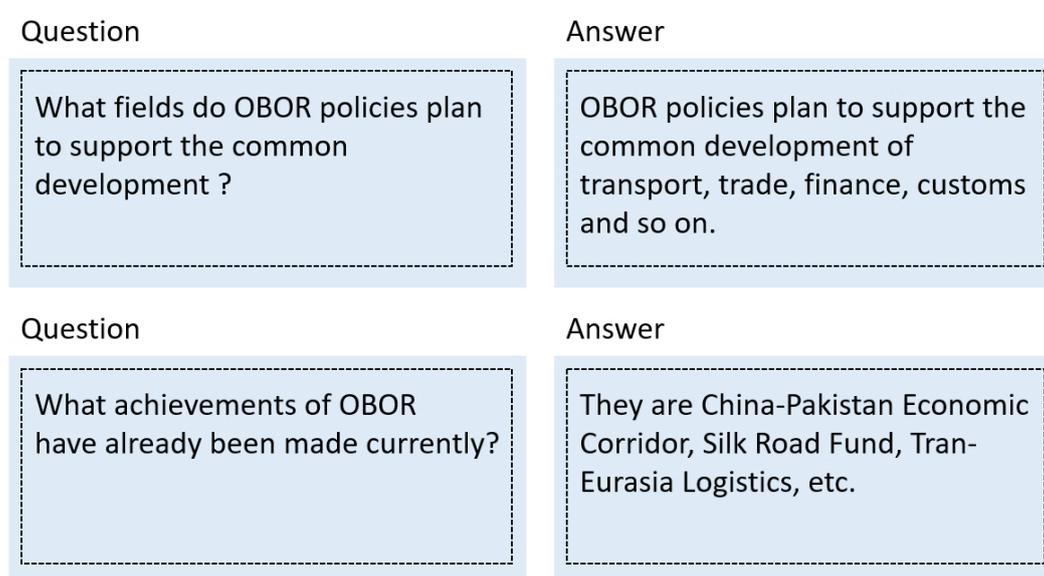


Figure 20. Search results of “OBOR policy” in Bing.

## 5.2. QA on OBOR

As an alternative to search engines, QA systems can provide users with more concise answers for their questions. It may be more convenient for users to directly get the answer via QA systems and do not need to find answers by themselves from lots of Web pages. Knowledge graph can provide QA systems with high-quality structured knowledge and guarantee the accuracy and reliability of the answers from QA systems. People can ask questions about OBOR in a knowledge graph-based QA systems and get the answers directly (see Figure 21). With the help of this kind of QA systems, they can easily become familiar with OBOR.

With the above examples, we can see that knowledge graph and its related applications can help people know more about OBOR quickly and conveniently. People can easily know the definition of OBOR, OBOR policies, OBOR benefits, etc., through knowledge graph-based search engines and QA systems. It may further promote the popularity of OBOR and the cooperation among countries along OBOR.



**Figure 21.** Questions and answers of OBOR in a knowledge graph-based QA system.

## 6. Conclusions

After proposing the OBOR cooperation initiative, China has actively built cooperative partnership with the countries along OBOR to jointly develop economy, politics, culture, science and technologies. As an important part of OBOR policies, cooperation in science and technologies is attracting increasing attention from many countries. Knowledge graph techniques, as an important branch of artificial intelligence, are getting increasing attention because of its powerful capability of knowledge representation and reasoning, and it is also widely applied in semantic search, QA, knowledge management and many other scenarios.

Nowadays, a large number of knowledge graphs were published and interlinked to each other on the World Wide Web, which forms a large-scale global knowledge base, known as the Linked Open Data Cloud. Most of the knowledge in current knowledge graphs is represented in English, which hinders the development of non-English knowledge graphs and their applications. Thus, both academia and industry need technologies and experience to construct non-English knowledge graphs. In recent years, with the emergence of the first large-scale Chinese encyclopedic knowledge graph-Zhishi.me, Chinese knowledge graphs has begun to develop rapidly. General knowledge graphs such as CN-DBpedia, XLORE and domain-specific knowledge graphs have been published one after another. A lot of techniques for constructing Chinese knowledge graphs have been proposed, including techniques

of knowledge extraction, knowledge integration, knowledge quality improvement and knowledge update. One problem here is that few works use the techniques of natural language processing based on Chinese linguistic characteristics to mine knowledge from text in real-world Chinese knowledge graph construction, but text is one of the most important mining sources for knowledge graph construction and should not be neglected. Thus, how to apply the techniques of Chinese open information extraction [31–33] from text to real-world Chinese knowledge graph construction worth further study. The rapid development of Chinese knowledge graphs and their techniques promote the development of related Chinese applications, including semantic search, QA, intelligence analysis and so on. Many of these applications are related to OBOR and have great impacts on OBOR. They can greatly promote the popularity of OBOR and the cooperation among countries along OBOR. Meanwhile, if the countries along OBOR aim to promote the development of the knowledge graph in their native languages, especially non-English languages, the techniques and applications of Chinese knowledge graphs are also very good references.

**Author Contributions:** T.W., G.Q. and C.L. summarized the history of knowledge graph development and the techniques of constructing Chinese knowledge graphs. M.W. investigated the applications of Chinese knowledge graphs. T.W., G.Q. and C.L. wrote the paper. All authors read and approved the final manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China (NSFC) under Grant No. U1736204.

**Acknowledgments:** We thank Prof. Anna Visvizi and Prof. Miltiadis D. Lytras who gave a lot of suggestions on how knowledge graphs impact the implementation of OBOR.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lehmann, J.; Isele, R.; Jakob, M.; Jentzsch, A.; Kontokostas, D.; Mendes, P.N.; Hellmann, S.; Morsey, M.; Van Kleef, P.; Auer, S.; et al. DBpedia—a large-scale, multilingual knowledge base extracted from Wikipedia. *Semant. Web* **2015**, *6*, 167–195.
2. Mahdisoltani, F.; Biega, J.; Suchanek, F.M. Yago3: A Knowledge Base from Multilingual Wikipedias. In Proceedings of the Biennial Conference on Innovative Data Systems Research, Asilomar, CA, USA, 6–9 January 2015.
3. Sowa, J.F. *Principles of Semantic Networks: Explorations in The Representation of Knowledge*; Morgan Kaufmann: Burlington, MA, USA, 1991.
4. Simmons, R.F.; Bruce, B.C. Some Relations Between Predicate Calculus and Semantic Net Representations of Discourse. In Proceedings of the International Joint Conference on Artificial Intelligence, London, UK, 1–3 September 1971; Volume 2, pp. 524–529.
5. Schubert, L.K. Extending the expressive power of semantic networks. *Artif. Intell.* **1976**, *7*, 163–198. [[CrossRef](#)]
6. Fahlman, S.E.; Touretzky, D.S.; Van Roggen, W. Cancellation in a Parallel Semantic Network. In Proceedings of the International Joint Conference on Artificial Intelligence, Vancouver, BC, Canada, 24–28 August 1981; pp. 257–263.
7. Brachman, R.J.; McGuinness, D.L.; Patel-Schneider, P.F.; Borgida, A. “Reducing” classic to practice: Knowledge representation theory meets reality. *Artif. Intell.* **1999**, *114*, 203–237. [[CrossRef](#)]
8. Horrocks, I. The FaCT System. In Proceedings of the International Conference on Automated Reasoning with Analytic Tableaux and Related Methods, Oisterwijk, The Netherlands, 5–8 May 1998; pp. 307–312.
9. Bizer, C.; Heath, T.; Berners-Lee, T. Linked Data-The Story So Far. *Int. J. Semant. Web Inf. Syst.* **2009**, *5*, 1–22. [[CrossRef](#)]
10. Niu, X.; Sun, X.; Wang, H.; Rong, S.; Qi, G.; Yu, Y. Zhishi.me-Weaving Chinese Linking Open Data. In Proceedings of the International Semantic Web Conference, Bonn, Germany, 23–27 October 2011; pp. 205–220.
11. Xu, B.; Xu, Y.; Liang, J.; Xie, C.; Liang, B.; Cui, W.; Xiao, Y. CN-DBpedia: A Never-Ending Chinese Knowledge Extraction System. In Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Arras, France, 27–30 June 2017; pp. 428–438.

12. Wang, Z.; Li, J.; Wang, Z.; Li, S.; Li, M.; Zhang, D.; Shi, Y.; Liu, Y.; Zhang, P.; Tang, J. XLORE: A Large-scale English-Chinese Bilingual Knowledge Graph. In Proceedings of the International Semantic Web Conference (Posters & Demos), Sydney, Australia, 21–25 October 2013; Volume 1035, pp. 121–124.
13. Li, M.; Shi, Y.; Wang, Z.; Liu, Y. Building a Large-Scale Cross-Lingual Knowledge Base from Heterogeneous Online Wikis. In *Natural Language Processing and Chinese Computing*; Springer: Cham, Switzerland, 2015; pp. 413–420.
14. Wu, T.; Wang, H.; Qi, G.; Zhu, J.; Ruan, T. On building and publishing Linked Open Schema from social Web sites. *J. Web Semant.* **2018**, *51*, 39–50. [[CrossRef](#)]
15. Wang, H.; Wu, T.; Qi, G.; Ruan, T. On Publishing Chinese Linked Open Schema. In Proceedings of the International Semantic Web Conference, Riva del Garda, Italy, 19–23 October 2014; pp. 293–308.
16. Wu, T.; Qi, G.; Wang, H. Zhishi.schema Explorer: A Platform for Exploring Chinese Linked Open Schema. In Proceedings of the Chinese Semantic Web and Web Science Conference, Wuhan, China, 8–12 August 2014; pp. 174–181.
17. Niu, X.; Rong, S.; Wang, H.; Yu, Y. An Effective Rule Miner for Instance Matching in a Web of Data. In Proceedings of the ACM International Conference on Information and Knowledge Management, Maui, HI, USA, 29 October–2 November 2012; pp. 1085–1094.
18. Wang, Z.; Li, J.; Wang, Z.; Tang, J. Cross-lingual Knowledge Linking cross Wiki Knowledge Bases. In Proceedings of the International Conference on World Wide Web, Lyon, France, 16–20 April 2012; pp. 459–468.
19. Wang, Z.; Li, J.; Tang, J. Boosting Cross-Lingual Knowledge Linking via Concept Annotation. In Proceedings of the International Joint Conference on Artificial Intelligence, Beijing, China, 3–9 August 2013; pp. 2733–2739.
20. Dong, X.; Gabrilovich, E.; Heitz, G.; Horn, W.; Lao, N.; Murphy, K.; Strohmann, T.; Sun, S.; Zhang, W. Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 601–610.
21. Wang, H.; Fang, Z.; Zhang, L.; Pan, J.Z.; Ruan, T. Effective Online Knowledge Graph Fusion. In Proceedings of the International Semantic Web Conference, Bethlehem, PA, USA, 11–15 October 2015; pp. 286–302.
22. Ding, J.; Ding, W.; Hu, W.; Qu, Y. An EBMC-Based Approach to Selecting Types for Entity Filtering. In Proceedings of the AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015; pp. 88–94.
23. Tonon, A.; Catasta, M.; Demartini, G.; Cudré-Mauroux, P.; Aberer, K. Trank: Ranking Entity Types using the Web of Data. In Proceedings of the International Semantic Web Conference, Sydney, Australia, 21–25 October 2013; pp. 640–656.
24. Kalyanpur, A.; Murdock, J.W.; Fan, J.; Welty, C. Leveraging community-built knowledge for type coercion in question answering. In Proceedings of the International Semantic Web Conference, Bonn, Germany, 23–27 October 2011; pp. 144–156.
25. Wu, T.; Qi, G.; Luo, B.; Zhang, L.; Wang, H. Language-Independent Type Inference of the Instances from Multilingual Wikipedia. *Int. J. Semant. Web Inf. Syst.* **2019**, Unpublished work.
26. Wu, T.; Ling, S.; Qi, G.; Wang, H. Mining Type Information from Chinese Online Encyclopedias. In Proceedings of the Joint International Semantic Technology Conference, Chiang Mai, Thailand, 9–11 November 2014; pp. 213–229.
27. Xu, B.; Zhang, Y.; Liang, J.; Xiao, Y.; Hwang, S.W.; Wang, W. Cross-Lingual Type Inference. In Proceedings of the International Conference on Database Systems for Advanced Applications, Dallas, TX, USA, 16–19 April 2016; pp. 447–462.
28. Liang, J.; Zhang, S.; Xiao, Y. How to Keep a Knowledge Base Synchronized with Its Encyclopedia Source. In Proceedings of the International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 3749–3755.
29. Cui, W.; Xiao, Y.; Wang, H.; Song, Y.; Hwang, S.W.; Wang, W. KBQA: Learning question answering over QA corpora and knowledge bases. *Proc. VLDB Endow.* **2017**, *10*, 565–576. [[CrossRef](#)]
30. Lai, Y.; Jia, Y.; Lin, Y.; Feng, Y.; Zhao, D. A Chinese Question Answering System for Single-Relation Factoid Questions. In Proceedings of the National CCF Conference on Natural Language Processing and Chinese Computing, Dalian, China, 8–12 November 2017; pp. 124–135.

31. Tseng, Y.H.; Lee, L.H.; Lin, S.Y.; Liao, B.S.; Liu, M.J.; Chen, H.H.; Etzioni, O.; Fader, A. Chinese Open Relation Extraction for Knowledge Acquisition. In Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (Volume 2: Short Papers), Gothenburg, Sweden, 26–30 April 2014; pp. 12–16.
32. Qiu, L.; Zhang, Y. ZORE: A Syntax-based System for Chinese Open Relation Extraction. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1870–1880.
33. Wang, H.; Qi, Z.; Hao, H.; Xu, B. A Hybrid Method for Chinese Entity Relation Extraction. In *Proceedings of the National CCF Conference on Natural Language Processing and Chinese Computing*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 357–367.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).