[BoldFont=STHeiti, ItalicFont=STKaiti]STSong [BoldFont=STHeiti] STFangsong

# Building Large-Scale Cross-lingual Knowledge Base from Multi-Encyclopedia

Mingyang Li†, Yao Shi†, and Zhigang Wang†

†Tsinghua National Laboratory for Information Science and Technology,
Department of Computer Science and Technology,
Tsinghua University, Beijing 100084, China
{lmy13@}tsinghua.edu.cn
{fantasysy@}sina.com

**Abstract.** Cross-lingual Knowledge Bases are important for global knowledge sharing. However, there are few Chinese-English knowledge bases due to the following reasons: 1) the scarcity of Chinese knowledge; 2) the number of current cross-lingual language links is limited; 3) the incorrect relation in semantic taxonomy. In this paper, a large-scale cross-lingual knowledge base(CLKB) is build to solve the above problems. The CLKB is based on English and Chinese Wikipedia as well as Baidu Baike and Hudong Baike. An extension method is used to increase language links while a pruning approach is used to refine taxonomy. Totally, there are XXX concepts, XXX instances, XXX properties referred in the CLKB. Moreover, the paper provides visualization webpages of knowledge and a SPARQL endpoint accessing to the established CLKB.

**Keywords:** Knowledge Base, Semantic Web, Ontology, Cross-lingual

## 1 Introduction

With Linked Open Data(LOD) developing recent years, an increasing number of knowledge bases are generated for information sharing. Large-scale knowledge bases in LOD project, such as DBpedia[10], YAGO[7] and Freebase[3] are often used for information extraction[4], entity linking[14], recommendation[13, 5, 6] and many other applications. These knowledge bases are not only cross-domain but also multilingual, which is benefit for global knowledge sharing.

DBpedia, the nucleus of LOD, extracts structured information from Wikipedia and contains approximately 38.3 million things. Now DBpedia provides 124 versions of non-English language including Chinese. However, there are only XXX instances and 11 classes in Chinese. Compared with the 4.58 million things in English, the quantity of Chinese information is obviously insufficient for further research and application.

A cross-lingual knowledge base can effectively promote global knowledge sharing, understanding and expanding. However, to build an practical knowledge base, some problems must be addressed: (1) The imbalance of different language source size leads to less entities. Comparing the number of articles in English

Wikipedia with those in Chinese Wikipedia, which are 5 million and 800 thousand separately, it's obvious that to structure a knowledge base in Chinese with Wikipedia is more challenging. (2) The low proportion of existing cross-lingual links in Wikipedia affects the quality of a bilingual knowledge base. Statistically, only XX% articles have cross-lingual links between English and Chinese in all articles of the two languages. Especially scarce evident links for infobox attribute. (3) The large but not rigorous category system in Wikipedia causes incorrect semantic relations in taxonomy. For example,example

Currently, there are several massive knowledge encyclopedias in Chinese, including Hudong Baike and Baidu Baike. To solve the imbalance problem, we utilize such monolingual sources to enrich our Chinese information. In this paper, we propose a pipeline to integrate four resources, English Wikipedia, Chinese Wikipedia, Hudong Baike and Baidu Baike, into one cross-lingual knowledge base, which contains XXX concepts, XXX instances and XXX properties. During the whole procedure, we also expand language links and make judgement on the semantic relations. Based on the obtained result, we develop an online website, which supports keyword search and SPARQL endpoint to our knowledge base. Specifically, our work makes the following contributions:

- We propose a method to build a Chinese-English cross-lingual knowledge base combining multi-encyclopedia. Among them, Chinese encyclopedias are utilized to help balance and enrich information in two languages.
- We extend the cross-lingual link set by employing a cross-lingual knowledge linking discovery approach for concept and instance, and analyzing templates in Wikipedia for property.
- We prune the original taxonomy, which is extracted from encyclopedia category system, to retrieve more precise *subClassOf* and *instanceOf* relations in ontology.
- Both website and SPARQL endpoint are provided for public query operations over our knowledge base.

The rest of the paper is organized as follows. Section 2 introduces the related concepts and defines the problem of building a cross-lingual knowledge base. Section 3 presents the extraction approaches in concept, instance and property level. Section 4 describes the procedure of building a knowledge base using extracted result. Section 5 demonstrates the data situation of established knowledge base. Section 6 reviews the related literatures. Section 7 concludes this paper and points out some future directions of this work.

## 2 Problem Definition

In this section, we firstly introduce the encyclopedias used to build and enrich our knowledge base. Then we give definitions about ontology and knowledge base. At last we describe our task in this work.

## 2.1 Encyclopedias

**Wikipedia** Nowadays, Wikipedia is the largest data store of human knowledge. It was launched in 2001 and has hold over 35 million articles in 288 languages by 2015. Out of these, English articles contribute most. The imbalance of different language articles makes ontologies based on Wikipedia-only behave badly in cross-lingual. Thus, more Chinese encyclopedias are necessary to enrich Chinese source.

Among the large-scale monolingual Chinese encyclopedias currently, Baidu Baike and Hudong Baike are the most content-rich. Hudong Baike was founded in 2005 and contains more than 12 million articles with about 9 million experts' contribution until 2015. Meanwhile, Baidu Baike maintains more than 11 million articles.

Here, we consider an encyclopedia wiki as a collection of articles, category system, which can be defined as: $W =< A, C >$, where A denotes articles, C denotes categories in W.

## 2.2 Wiki Pages

Articles from the four sources are similar in structure. Usually they provide two important elements with potential semantic information, category system and articles. A category system represents the relations between categories as a tree by the relation *subCategoryOf*. Fig.1 shows a screenshot of Hudong Taxonomy. An article describes an entity with rich information created and modified by several verified editors. Besides, each article is linked to one or more categories by *articleOf* relation. In general, there are five elements can be exploited in each article page:


**Fig. 1.** Taxonomy in Hudong

- Title: A Title is the label of an entity, which is unique so that it can be used to distinguish entities.
- Abstract: An abstract is a brief summary of the entity. It's always the first paragraph of an article.
- Infobox: Most of articles contain infobox. An infobox maintains structured data which are subject-attribute-value triples formalized as a table. Information in this table includes important properties of an entity.
- Links: Links are entries to other articles within the encyclopedia. They lead readers to reference articles. Actually, they represent the relations between the current article and other articles.
- Category: The categories that an article belongs to are usually listed at the bottom of article page, shown as tags. An article has *articleOf* relation with its categories.
- URL: Each article has an HTTP url to identify itself on web.

Fig. 2 shows a snap of an article in Chinese Wikipedia.


**Fig. 2.** A snap of Interstellar(Film) article in Chinese Wikipedia

The elements of each article $a$ can be defined as follow:

$$a =< Ti(a), Ab(a), Li(a), In(a), C(a), U(a) > \quad (1)$$

where $Ti(a), Ab(a), Li(a), In(a), C(a), U(a)$ denotes title, abstract, links, infobox, category tags, url of article $a$.

Notably, articles in Wikipedia follow templates specified by Wikipedia when being edited. A template defines items that a group of article should fill. Besides, infoboxes are also generated based on certain templates recommended by Wikipedia. For example, The infobox in film (Interstellar) is edited according to the Template *Infobox film*, which maintains a property set of films. An Infobox $In(a)$ contains a set of attribute-value pairs $p_1$, $p_2$,.... We denote infobox template as $T(a)$. Templates specify certain attributes, which are usually different

from those displayed on the webpage. Thus, we define an attribute-value pair as a triple $p =< tl, dl, v >$, where $tl$ is attribute label in template, $dl$ is displayed label in web page and $v$ is the corresponding value. The value maybe a text or a reference to another entity.

Moreover, in Wikipedia, some article pages have language links which help readers switch to other language-version within the same article. Fig. 2shows language links of *Interstellar* on the right column of the Wikipedia page. As to an article $a$ containing multi-language content in Wikipedia, $L_e$ and $L_z$ denotes its article links, usually titles, in English and Chinese separately. Thus, to a $cl$ in $CL$, $cl(a) =< L_e(a), L_z(a) >$.

### 2.3 Ontology and Knowledge Base

An ontology is a formal specification of a group of entities. In our work, an ontology is described as a 4-tuple:

$$O =< C, I, P, H^C > \tag{2}$$

where $C$, $I$, $P$ are the sets of concepts, instances, and properties, respectively. $H^C$ represents the hierarchical relationships of concepts. A Taxonomy includes two types of relationships, which are *subClassOf* of concept-concept and *instanceOf* of concept-instance.

A cross-lingual knowledge base is a database conform to a cross-lingual ontology. Taking advantage of language links in $CL$, several monolingual-ontologies generated from various sources can be merge into one cross-lingual ontology. Thus a knowledge base can be defined as:

$$KB =< O_i, CL > \tag{3}$$

where $O_i$ denotes the ith monolingual-ontology, $CL$ represents the cross-lingual link set.

### 2.4 Cross-lingual Knowledge Base Building

In this paper, our task is to build an Cross-lingual Knowledge Base(CLKB) assembling knowledge from several English or Chinese encyclopedia sources. Given four encyclopedia datasets $W_1, W_2, W_3, W_4$, we get $O_i$ including concept list $C_i$, instance list $I_i$, property list $P_i$, taxonomy $H_i^C$ of each dataset $W_i$ by extracting. We then enrich the extracted language-link set $CL$ using an link-discovery extension method. We also refining taxonomy by checking if an *articleOf* or *subCategoryOf* is really an *instanceOf* or *subClassOf* relationship. Our final output is an English-Chinese ontology with SPARQL endpoint though process of combine $O_i$ by utilizing language links in $CL$.

The whole building procedure is shown in Fig. 3 We extract information from four sources, Baidu Baike, Hudong Baike, Chinese Wikipedia and English Wikipedia. Considering the different data format of each source, that is,

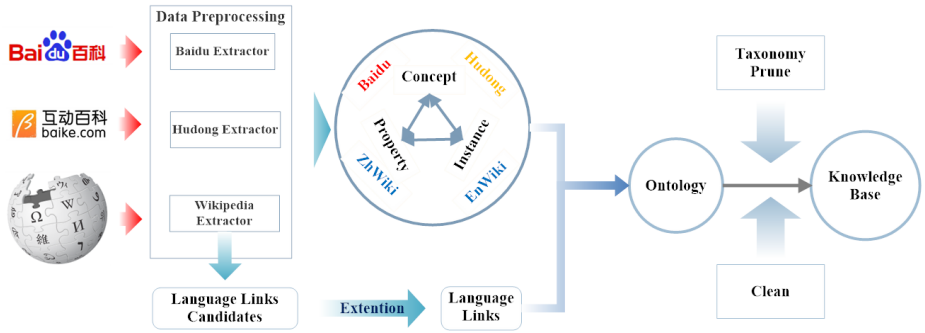


**Fig. 3.** Procedure of building our cross-lingual knowledge base

html code page of Baidu and Hudong with different layout, and XML format of Wikipedia dump file, various extractors should be employed. After data parsing, we get four ontologies each source. Going through the process of taxonomy pruning, we delete some incorrect relations. At last Using the extended language links, we combine the four ontologies into a cross-lingual knowledge base.

## 3 Semantic Data Extraction

Semantic data extraction aims to achieve a structured dataset from the input encyclopedias, preparing for the knowledge base construction. Specifically, we extracts concepts from category taxonomy, instances according to articles, and properties based on infoboxes and its templates.

### 3.1 Concept Extraction

A concept is defined as a type of similar instances. For example, the concept of instance *Interstellar* is *Movie*. In general, a concept has super classes and sub classes. Concepts comprise a taxonomy which presents a backbone of an ontology. In an encyclopedia, a category groups several articles and also has super-categories and sub-categories. Therefore we can extract concepts based on existing category system.

However, the whole taxonomy can not directly transform from category system because of the following problems:

- There are auxiliary categories in Wikipedia, which help arrange specific articles or category pages. For example, *Lists of artists* or *Food templates*.
- Some categories relate to only one or two articles. According to the definition of concept, such categories are less representative to a group of instances, therefore it's unwise to retain it as concept.

To obtain a more precise $H^C$ in a $W$, we remove such categories in all encyclopedias, then build the original concept hierarchy $H^C$ using the remaining categories.

However, among the relations there are still incorrect samples. For example, *Tsinghua University* is not an entity of *Haidian District*, but relates to. Thus we will prune the taxonomy later.

### 3.2 Property Extraction

A property is defined as an attribute of an entity. It represents the relation between two instances or an instance and its value. We divided properties into two types: object property, whose value is an individual, such as *directed by*; datatype property, whose value is a literal text, such as *birth date*. Considering both content and infobox of an article, we extract two kinds of properties, general-properties and Infobox-properties.

**General-properties** Characteristics of an entity are regarded as general-properties, including label, abstract, and url. General-properties describe general information of an entity. We define three datatype properties as general-properties for a given article $a$: (1) label property; (2) abstract property; (3) URL property.

**Infobox-properties** Attributes acquired from infobox are considered as Infobox-properties, such as (release date), (directed by) in a movie's infobox. The type of a property, datatype or object, depends on the type of the value. Ordinarily, a plain text value marks the property as datatype while an entity reference determines the property as objecttype. For example, the attribute (release date) can be defined as a datatype property as its value is a datetime string. Meanwhile, (directed by) can be an object property because its value points to a person who directed the movie.

We are challenged when extracting properties from infoboxes:

- In Wikipedia, the attribute label displayed in the webpage infobox is inconsistent with it in the published dump file. Fig.4 gives a mapping result of display labels and dump labels in *Interstellar*' infobox. The left is infobox, the middle is a snap from dump file in Wikipedia. As we've seen, attribute label in infobox is different from it in dump file. We explore the display labels as property labels in Wikipedia rather than dump labels extracted from raw data.
- There are special characters in labels. Wikipedia usually uses hyphen "-" or dot "" to mark sublabels. For example, *population* property has subproperties "-Density" and "-Urban". In addition, odd signs, such as colon or asterisk, may occur in Baidu or Hudong property labels by mistake.

To solve the problems above, we take advantage of template information. Specifically, Wikipedia institutes rules of rendering label in templates. For example, movie infobox follows template *Infobox film*, which is shown on the right

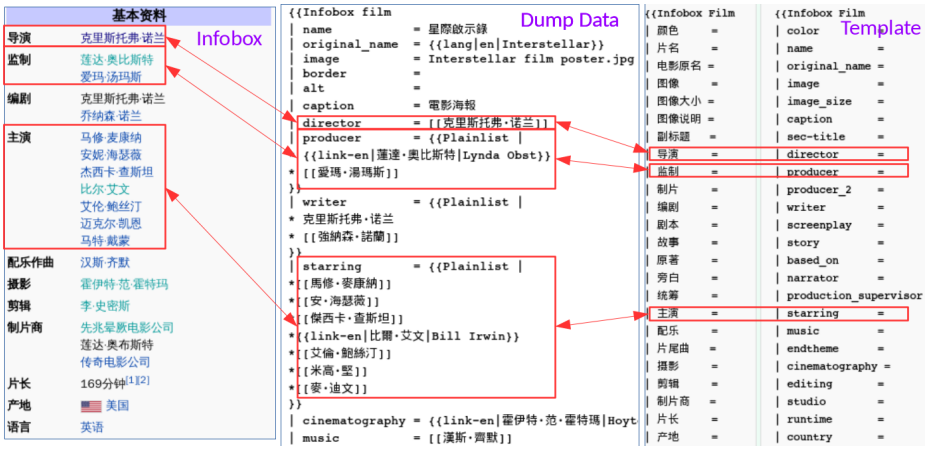


**Fig. 4.** Comparison of display label and dump label in *Interstellar* infobox

of Fig.4, where both display labels and dump labels come from. When a dump label in triple-bracket occurs in dump file, we replace it by its mapping display label. After convert all the dump labels, we make a filter to redress the label text.

### 3.3 Instance Extraction

In encyclopedia, an article describes unique entity in the world. Therefore we can extract an article as an instance. During the extraction, illustrative or structure-related articles in Wikipedia are deleted, including category list pages and template documentations.

We harvest four types of information during this stage. (1) General-properties of instance, including title as label property value, first paragraph as abstract property value and HTTP URL as URL property value. (2) Infobox-properties which are acquired via extracting from the infobox in the article; (3) $articleOf$ relation with categories listed at the bottom of article page. For example, (American science fiction films) is a category of Interstellar; (4) Reference relation with other instances according to links in the content. We gain the reference $Li(a)$ between the current instance and others, such as (Warner Bros.).

## 4 Cross-lingual Integration

To construct a cross-lingual knowledge base with existing structured data, firstly we gather cross-lingual links which can help match the same entity in two languages and extend the link set. Secondly, we integrate this four encyclopedias, which is to say, respectively merging concepts, instances and properties from the four sources if they represent the same thing. Thirdly, we prune the taxonomy

generated from concept relationships to make it more accurate. At last, we make instances and properties attached to complete the cross-lingual knowledge base construction.

### 4.1 Cross-lingual Linking

There are number cross-lingual links between English and Chinese, which constitute the initial cross-lingual link set of concepts and instances. Moreover, we utilize the language-independent method in [16] to extend the language-link set. With the linkage factor graph model, we harvest a cross-lingual links extension as many as 20 thousands with an ideal precision 85.5% and a recall of 88.1% between English Wikipedia and Baidu Baike.

However, due to using templates, Infobox-properties have no obvious cross-lingual links. To acquire such links, we take the following steps:

1) Given a matched template, which means $T_e$ and $T_z$ are cross-lingual pairs, find the display labels mapping the same dump label. That is to say, to two pairs, $p_e$ in $T_e$ and $p_z$ in $T_z$, if $tl_e$ is equal to $tl_z$, $< dl_e, dl_z >$ are cross-lingual properties;
2) Given the English and Chinese infoboxes of a matched instance, compare their templates, which are English template $T_e(a_e)$ and a Chinese template $T_z(a_z)$ in which $a_e$ and $a_z$ direct to the same entity, find the matched display labels mapping to the same dump label;
3) Given the English and Chinese infoboxes of a matched instance, for datatype properties, compare the similarity of literal value; to object properties, check whether the value refer to the same entity.

In order to make all these encyclopedias link to each other, we unify the same concept, instance and property from four sources, and give them unique identifiers. For instance, we merge instances by the following method: (1) Merge all instances extracted from Chinese encyclopedias by instance title. (2) To a $L_z(a)$, find whether there is an English cross-lingual link in $CL < L_e, L_z >$. If exists, make the two as one instance and identify it using an ID, else number it with an new ID.

After the above steps, we acquire a list of instances and their unique IDs. Some of them contain cross-lingual information while some contain just monolingual information.

The process of unify concept and property is the same as instance. Meanwhile, all the relations in all sources are kept to prevent loss of information.

### 4.2 Taxonomy Prune

As a result of combining multi-source information without verifying, the taxonomy of concept system is mussy. For example, example. Therefore, we introduce the method from[18] to detect the correct *subClassOf* and *instanceOf* relations from *subCategoryOf* and *articleOf*. Table. **??** shows some examples of

correct relations. In particular, some language-dependent literal and language-independent structural features are defined to vectorize each concept or instance. Employing these features, a Yes-or-Not binary-classification model is trained based on Logistic Regression. The whole process is iterative by retraining model with assured result to get higher precision. Confirming a right *subClassOf* or *instanceOf* not only depends on the prediction result of classification, but also cross-lingual knowledge validation, which ensures correctness if both the mapped English and Chinese relations are correct.

The ideal result of pruning is a tree, whose edges, nodes, and leaves separately denote relations, concepts and instances. However, since getting rid of incorrect entity relations without consideration of integrity, a forest result is inevitable.

## 5 Result

### 5.1 Extracted Knowledge Base

We collect the resources from 4 wiki sites, English and Chinese Wikipedia dump files in April, 2015, Hudong html pages until May, 2014, and Baidu html pages until September, 2014. Each of the wikis has three types of information, which can be utilized for constructing our ontology, namely, specific articles, classification system, and attribute of articles. We extract each raw data, then form the extracted information into well-structured data. Table 1 the result we get after elementary extraction in 4 different wiki sources.

**Table 1.** Statistics of elementary extraction result

| | Enwiki | Zhwiki | Hudong | Baidu |
|---|---|---|---|---|
| #Instance | 4304113 | 662650 | 5590751 | 5622404 |
| #Concept | 982432 | 159705 | 31802 | 1300 |
| #Property | 43976 | 18842 | 1187 | 139634 |

To build a knowledge base, we create URIs to identify each element and provide corresponding information if users look up elements over HTTP protocol to achieve the knowledge base. Table 2 lists the defined URIs.

**Table 2.** URIs for concept, instance, property in our knowledge base

| Type | URI |
|---|---|
| Concept | http://clkb/concept/id |
| Instance | http://clkb/instance/id |
| Property | http://clkb/property/id |

After fusing the heterogeneous resources, we harvest a cross-lingual graph with 1,093,855 classes, 200,223 properties, and 11,721,238 instances respectively.

With different methods of extraction and language link discovery, these three kinds of entries show different results in languages. We give a breakdown of both Chinese knowledge and English knowledge in Table 3.

**Table 3.** Statistics of our knowledge base

| | Concepts | | Instance | | Property | |
|---|---|---|---|---|---|---|
| English | 982,982 | 89.86% | 4,311,719 | 37.79% | 54,802 | 27.37% |
| Chinese | 176,648 | 16.15% | 7,842,117 | 66.9% | 167,083 | 83.45% |
| Cross-lingual | 65,775 | 6.01% | 432,598 | 3.7% | 21,662 | 10.82% |
| Only English | 917,207 | 83.9% | - | - | 33140 | 16.6% |
| Only Chinese | 110,873 | 10.1% | - | - | 145,421 | 72.6% |
| Sum | 1,093,855 | - | 11,721,238 | - | 200,223 | - |

Our knowledge graph is organized in Openlink Virtuoso, which is a data management platform covering various server, including triple store.

### 5.2 Web Access to Knowledge Base

We provide a website to present an intuitive graph of our knowledge in the forms of class, instance and property. Fig.5 shows sample pages of the integrated data. Users can choose language preference which is convenient for both English speaker and Chinese speaker. In the class webpage, we exhibit the label, super classes, subclasses, related topics, properties and instances of the specific class in bilingual way on condition that the corresponding English entries or Chinese entries exist. In the instance webpage we display bilingual label, super classes, related classes, abstract, property key value pairs, images and references. In the property webpage, we present bilingual label, domains, ranges, and related instances of each property. Beside these user-friendly pages, we provide two ways to access our knowledge base: via the search engine shown in Fig.**??**, or via SPARQL endpoint shown in Fig.**??**. For users who know little about semantic web, they can query by inputing related text into searchbox and search to get entities with similar label. To present practicable result, An index is generated over all entities. We as well provide SPARQL interface for professional users to query our knowledge graph. Users can choose the language tags of their desired results by **"filter(langMatches(?label),"en"))"** or **"filter(langMatches (?label),"zh"))"**.

## 6 Related Work

In this section, we introduce some related knowledge bases and cross-lingual knowledge linking methods.

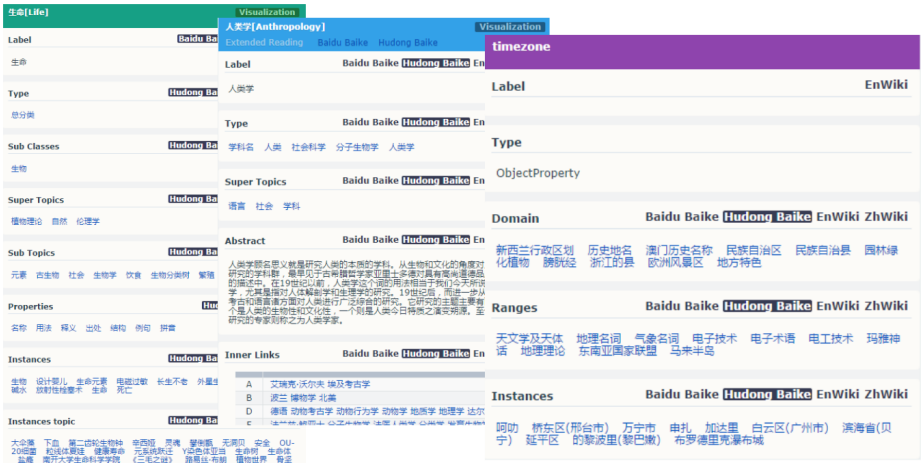


**Fig. 5.** Sample Pages of Class, Instance and Property

### 6.1 Chinese Knowledge Bases

Currently, several large-scale Chinese knowledge bases have been generated. Zhishi.me[11, 15] is the first published Chinese large-scale Linking Open Data. It acquires structural information from three original sources, Chinese Wikipedia, Baidu Baike and Hudong Baike and gains more than 5 million distinct entities. Zhishi.me helps generate knowledge base focused on relations in Junfeng Pans work[12]. Similar with Zhishi.me, CKB[17] is created from Hudong Baike. It first learns an ontology based on category system and properties, and then collects 19542 concepts, 2381 properties, 802593 instances. Besides using existing encyclopedias, CASIA-KB employs other types of sources(e.g. microblog posts, news pages, images) to enrich the structured knowledge.

### 6.2 Cross-lingual Knowledge Bases

DBpedia [2, 10] is one of the most used cross-lingual knowledge base in the world. It's extracts various kinds of structured information from Wikipedia and employ the multilingual characteristic of Wikipedia to generate 97 language versions of content. This knowledge base is widely applied in many domains, including media recommendation [13, 5, 6], entity linking[9] and information extraction [4]. Universal WordNet(UWN)[8] is a large multilingual lexical knowledge base which is build from WordNet and enriched its entities from Wikipedia. It is constructed using sophisticated knowledge extraction, link prediction, information integration, and taxonomy induction methods. The API is available to over 200 languages and more than 16 million words and names. UWN provides semantic relationship of list of word meanings for Aya's work on conceptual search [1]

## 7 Conclusion

This paper presents a procedure of building a Chinese-English cross-lingual knowledge base from four encyclopedia sources. At first, we extract structured information and unify data format. Then a cross-lingual language link set is generated to help combine the bilingual sources. Besides, an extension method is applied to enrich the links. To refine our dataset, we also conduct pruning work on taxonomy. Finally, we acquire a knowledge base containing XXX concepts, XXX instances and XXX properties. Currently, a SPARQL query interface is provided to access the knowledge base.

## References

1. Al-Zoghby, A.M., Shaalan, K.: Conceptual search for arabic web content. In: Computational Linguistics and Intelligent Text Processing, pp. 405–416. Springer (2015)
2. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: Dbpedia: A nucleus for a web of open data. Springer (2007)
3. Bollacker, K., Evans, C., Paritosh, P., Sturge, T., Taylor, J.: Freebase: a collaboratively created graph database for structuring human knowledge. In: Proceedings of the 2008 ACM SIGMOD international conference on Management of data. pp. 1247–1250. ACM (2008)
4. Dutta, A., Meilicke, C., Niepert, M., Ponzetto, S.P.: Integrating open and closed information extraction: Challenges and first steps. In: NLP-DBPEDIA@ ISWC (2013)
5. Fernández-Tobías, I., Cantador, I., Kaminskas, M., Ricci, F.: A generic semantic-based framework for cross-domain recommendation. In: Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems. pp. 25–32. ACM (2011)
6. Kaminskas, M., Fernández-Tobías, I., Ricci, F., Cantador, I.: Knowledge-based music retrieval for places of interest. In: Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies. pp. 19–24. ACM (2012)
7. Mahdisoltani, F., Biega, J., Suchanek, F.: Yago3: A knowledge base from multilingual wikipedias. In: 7th Biennial Conference on Innovative Data Systems Research. CIDR 2015 (2014)
8. de Melo, G., Weikum, G.: Uwn: A large multilingual lexical knowledge base. In: Proceedings of the ACL 2012 System Demonstrations. pp. 151–156. Association for Computational Linguistics (2012)
9. Mendes, P.N., Daiber, J., Jakob, M., Bizer, C.: Evaluating dbpedia spotlight for the tac-kbp entity linking task. In: Proceedings of the TACKBP 2011 Workshop. vol. 116, pp. 118–120 (2011)
10. Mendes, P.N., Jakob, M., Bizer, C.: Dbpedia: A multilingual cross-domain knowledge base. In: LREC. pp. 1813–1817 (2012)
11. Niu, X., Sun, X., Wang, H., Rong, S., Qi, G., Yu, Y.: Zhishi. me-weaving chinese linking open data. In: The Semantic Web–ISWC 2011, pp. 205–220. Springer (2011)
12. Pan, J., Wang, H., Yu, Y.: Building large scale relation kb from text. In: 11th International Semantic Web Conference ISWC 2012. p. 93. Citeseer (2012)

13. Passant, A.: dbrecmusic recommendations using dbpedia. In: The Semantic Web–ISWC 2010, pp. 209–224. Springer (2010)
14. Shen, W., Wang, J., Luo, P., Wang, M.: Linden: linking named entities with knowledge base via semantic knowledge. In: Proceedings of the 21st international conference on World Wide Web. pp. 449–458. ACM (2012)
15. Wang, H., Wu, T., Qi, G., Ruan, T.: On publishing chinese linked open schema. In: The Semantic Web–ISWC 2014, pp. 293–308. Springer (2014)
16. Wang, Z., Li, J., Wang, Z., Tang, J.: Cross-lingual knowledge linking across wiki knowledge bases. In: Proceedings of the 21st international conference on World Wide Web. pp. 459–468. ACM (2012)
17. Wang, Z., Wang, Z., Li, J., Pan, J.Z.: Building a large scale knowledge base from chinese wiki encyclopedia. In: The Semantic Web, pp. 80–95. Springer (2012)
18. Wang, Z., Li, J., Li, S., Li, M., Tang, J., Zhang, K., Zhang, K.: Cross-lingual knowledge validation based taxonomy derivation from heterogeneous online wikis. In: Twenty-Eighth AAAI Conference on Artificial Intelligence (2014)