CASIA-KB: A Multi-source Chinese Semantic Knowledge Base Built from Structured and Unstructured Web Data

Yi Zeng, Dongsheng Wang, Hao Wang, Tielin Zhang, Hongwei Hao

Institute of Automation, Chinese Academy of Sciences, Beijing, China

{yi.zeng, dongsheng.wang, hongwei.hao}@ia.ac.cn

Abstract. Knowledge bases play a crucial role in intelligent systems, especially in the Web age. Many domain dependent and general purpose knowledge bases have been developed to support various kinds of applications. In this paper, we propose the CASIA-KB, a Chinese semantic knowledge base built from various Web resources. CASIA-KB utilizes Semantic Web and Natural Language Processing techniques and mainly focuses on declarative knowledge. Most of the knowledge is textual knowledge extracted from structured and unstructured sources, such as Web-based Encyclopedias (where more formal and static knowledge comes from), Microblog posts and News (where most updated knowledge comes from). CASIA-KB also aims at bringing in images and videos (which serve as non-textual knowledge) as relevant knowledge for specific instances and concepts since they bring additional interpretation and understanding of textual knowledge. For knowledge base organization, we briefly discussed the current ontology of CASIA-KB and the entity linking efforts for linking semantically equivalent entities together. In addition, we build up a SPARQL endpoint with visualization functionality for query processing and result presentation, which can produce query output in different formats and with result visualization supports. Analysis on the entity degree distributions of each individual knowledge source and the whole CASIA-KB shows that each of the branch knowledge base follows power law distribution and the merged knowledge base keeps this structural property.

Keywords: Chinese Semantic Knowledge base; Web of Data; Semantic Web; Information Extraction.

1 Introduction

Large-scale knowledge bases are essential to support various knowledge driven applications, especially in the context of the Web. Many efforts have been made such as CYC [1], DBPedia [2], YAGO [3], ConceptNet [4], NELL [5], etc. Most of them are either based on Semantic Web techniques or Natural Language Processing. In addition, most of them are English or at least English centric semantic knowledge bases.

In this paper, we introduce the construction of CASIA-KB developed at Institute of Automation, Chinese Academy of Sciences (CASIA), which is a large-scale semantic knowledge base built from multiple structured and unstructured sources, ranging from textual contents (e.g. Web-based Encyclopedia, microblog posts, news pages) to non-textual contents (e.g. images). The whole knowledge base is constructed based on close combination of Semantic Web, Natural Language Processing, statistical analysis and network visualization techniques.

Firstly, we crawl the large volume of pages and extract structured facts in infoboxes from the three sources, including Baidu Encyclopedia, Hudong Encyclopedia and Chinese Wikipedia. The structured facts are represented in RDF N3 format and stored in Jena TDB[1], resulting in a large scale triple-based semantic knowledge base. In order to enrich the knowledge from the free texts in Web-based Encyclopedia, News Web pages, and social media such as microblog posts, we employ pattern-based methods to extract knowledge triples and enrich the original knowledge base.

Secondly, we extract the "sub class of" and "instance of" relation from the Web-based Encyclopedia category information, and construct a hierarchical ontology. In order to have cross knowledge source, we embed an entity linking algorithm into the knowledge base where entities from different sources can be dynamically linked together. Finally, we build up an SPARQL and a visualization interface for query results presentation.

The other part of this paper is organized as following. Section 2 provides a very brief introduction of related works about the development of knowledge bases. Section 3 describes the construction of CASIA-KB in detail. Section 4 presents the conclusion and future works.

2 Related Works

After successful development of state-of-the-art knowledge bases such as CYC [1], the efforts for building large-scale knowledge bases have been shifted to automatic construction of knowledge bases based on Web contents. DBpedia [2] is based on structured knowledge extraction from Wikipedia structured sources such as infoboxes. It assigns a global identification for each entity, with hierarchical ontologies managing all these triple data together. In addition, based on the multiple language versions of Wikipedia, DBPedia knowledge base also provides multiple language versions. Another well accepted semantic knowledge base YAGO [3], which is built from Wikipedia and WordNet provides temporal and spatial extensions for existing knowledge extracted from Wikipedia. Besides extraction of knowledge from structured sources, recent efforts on knowledge base construction based on information extraction as well as rule learning from unstructured texts are with great potential (e.g. NELL [5]).

As for large-scale Chinese knowledge bases, National Knowledge Infrastructure (NKI) is a pioneer work and has been applied to several domains such as call center

[1] http://jena.apache.org/documentation/tdb/

support, etc [6]. Recent efforts for building Chinese semantic knowledge bases have been focusing on extracting structured knowledge from Web-based encyclopedias and representing the knowledge using Semantic Web languages and techniques. Zhishi.me [7], which is developed by Shanghai Jiaotong University crawl facts embedded in information tables (known as infobox) from three sources including Baidu Encyclopedia, Hudong Encyclopedia and Chinese Wikipedia. Another semantic knowledge base maintained by Tsinghua University extracts and refines hierarchical ontologies for managing entities and infobox knowledge from Baidu Encyclopedia and Hudong Encyclopedia [8,9]. Nevertheless, knowledge extraction from unstructured texts has not been included, while most human knowledge on the Web is embedded in unstructured Web contents.

Early versions of CASIA-KB contains declarative knowledge from structured sources such as infobox triples and ontology from Web-based Encyclopedias [10], and extensions of representing uncertain knowledge with typicality was introduced in [11]. In this paper, we introduce the whole road map, current status and implementation details of CASIA-KB.

3 Knowledge Base Construction

This section introduces the design and implementation of the CASIA-KB in details. Section 3.1 provides the roadmap and current status of CASIA-KB. Section 3.2 illustrates the large RDF triple data construction from structured sources. Section 3.3 discusses the knowledge enrichment from unstructured sources. Section 3.4 introduces the OWL ontologies of the three resources and how to adopt entity linking algorithm for data level entity linking. Section 3.5 introduces and demonstrates the SPARQL endpoint as the interface for accessing the CASIA-KB and the visualization of the knowledge base.

3.1 The Roadmap and Current Status of CASIA-KB

CASIA-KB is a general purpose and multi-source semantic knowledge base built from various resources. CASIA-KB mainly focuses on declarative knowledge. Most of the knowledge is textual knowledge extracted from structured and unstructured sources, such as Web-based Encyclopedias (where more formal and static knowledge comes from), Microblog posts and News (where most updated knowledge comes from). CASIA-KB also aims at bringing in images and videos (which serve as non-textual knowledge) as relevant knowledge for specific instances and concepts since they bring additional interpretation and sometimes better understanding of textual knowledge [12]. The road map of CASIA-KB is illustrated as Figure 1.

The first version of CASIA-KB only contains knowledge from structured sources such as infobox triples from Web-based Encyclopedias [10]. In this paper, we provide detailed introduction of how the CASIA-KB is been built and the extraction of knowledge (including triples with Is-A, Part-Whole, and Has-Property relation) from unstructured sources (e.g. free texts in Web-based Encyclopedias, microblog

posts, and news pages) will also be discussed. For non-textual knowledge, currently, CASIA-KB contains images crawled from Baidu Images and Web-based Encyclopedias to support various applications, such as the CASIA tour recommendation system[2]. In this paper, we will mainly focus on the construction of textual knowledge base. The main resources that CASIA-KB contains are listed in Table 1.

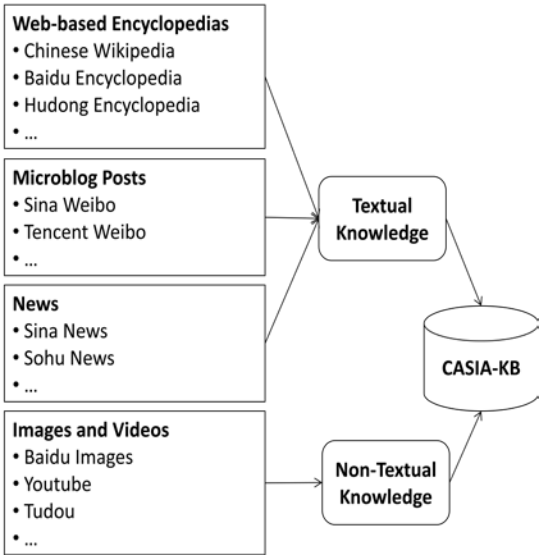


Fig. 1. The road map of CASIA-KB

Sources	**Entities**	**Triples from Structured Sources**	**Triples from Unstructured Sources**	**Triples with Typicality**
Baidu Encyclopedia	1,967,259	9,752,018	276,676	15,926,018
Hudong Encyclopedia	1,581,055	11,216,471	253,588	
Chinese Wikipedia	460,439	4,569,236	13,154	
Sina Weibo	53,172		920,823	
Sogou News	10,963		31,140	

Table 1. Current Status of CASIA-KB

All the above literal triple knowledge is transformed to RDF triples and stored in Jena TDB. Since the knowledge in CASIA-KB are all from the Web contributed by vari-

[2] CASIA Tour Recommendation System: http://imip.ia.ac.cn/TourRec

ous users and organizations with different understanding of the world knowledge. It is obvious that the extracted knowledge may be with different uncertainties. Hence, CASIA-KB is designed to be with uncertainties. In [11], we introduced a typicality approach for describing knowledge with uncertainty. How the generated typicality values are used for entity conceptualization and knowledge validation are also discussed [11]. In this paper, we focus on the knowledge extraction and ontology organization part of CASIA-KB.

3.2 Knowledge Extraction from Structured Data and Resource Organization

In this paper, we consider information tables such as infoboxes, and category information on the Web-based Encyclopedia pages as structured data sources. Firstly, we crawl the entire Web pages from Baidu Encyclopedia, Hudong Encyclopedia and Chinese Wikipedia. Secondly, we extract the entity names, their properties and property values, as well as category information to obtain the literal knowledge triple. Thirdly, all the literal triples are transformed to RDF triples.

Fig. 2. Screenshot of an page in Baidu Encyclopedia

As shown in Figure 2, the title "*中国科学院*" is extracted as the entity name, and the infobox information is represented as the properties and property values of the

entity (e.g. <中国科学院, 总部地址, 北京>). The category information in the bottom of the page is transformed to triples with the "*rdf:type*" relation. In addition, "Related Terms" in the pages are stored as triples with a general binary relation "RelatedTo".

Since the knowledge is contributed by different users in various sources, and they may mention the same entity using different names, synonyms are very important for better knowledge organization in the knowledge base in the way of linking semantically equivalent entities together. In our study, the synonym set includes possible resources from synonym labels, redirect labels, nick names, previous names, etc from the three encyclopedia sources (e.g. as shown in Figure 2, "*中科院*" is a synonym of the full title "*中国科学院*", and this piece of information is extracted to build up the synonym set). 476,086 pairs of synonyms are added in the synonym set. In order to have better supports for users to retrieve resources, we utilize the synonym set to build labels for resources to support different input alternatives.

Triples from the same knowledge source are more consistent, and well managed by the ontology designed for this specific knowledge source. Hence, triples from the same knowledge source are stored in the same named graph. Users can choose which knowledge source(s) they would like to use..

3.3 Knowledge Extraction from Unstructured Data Sources

Web-based encyclopedia Infoboxes and Tables are direct sources for extracting structured knowledge, while there are a lot more knowledge which are represented in unstructured text resources, such as free texts in Web-based encyclopedia and microblog posts. In this paper, we try to use rule based methods to extract knowledge triples to enrich the original knowledge base. We mainly focus on the extraction of three types of declarative knowledge, namely, Is-A relation, Part-Whole relation and Has-Property relation.

There are two types of binary relations which can be used to enrich the organization of the knowledge base ontology, namely, Is-A relation and Part-Whole Relation. Is-A relation represents a hierarchical binary relation either between two concepts or between an instance and a concept. Unlike extracting Is-A relation from unstructured English texts [13], there are various ways to represent Is-A relation in Chinese. Some possible (but not complete) candidates are listed in Table III. In the extraction phase, we find the subject of the triple before the possible candidate terms which represent the Is-A relation, and the object should appear after the Is-A relation. In this paper, we focus on extracting Is-A relation whose subjects and objects are all nouns.

Part-Whole relation represents the Part-Whole binary relation among two resources, in which one resource is a part of another. The most important English pattern for extracting Part-Whole relation are "part of" and "is composed of", while in Chinese, some possible patterns are listed in Table 2.

Properties are essential for describing concepts and instances in knowledge bases. In this paper, we focus on extracting property names and construct declarative

triples with Has-Property relation. In English, properties can be extracted from the pattern "the X of Y…" while "X" represents a possible property of the concept or instance "Y". The relevant pattern in Chinese is "Y的X是/为", while "X" is the property here, and the triple is represented as <Y, Has-Property, X>.

Relation Type	Possible Chinese Terms Relevant to This Relation
Is-A	是/为一(个/只/群/把/条/种/件/名/门/款/颗/根/头/尾/片/次/座/块/本/代/张/双/篇/堆/批/章/节/步/副/缕/位/台/侏), 属于, …
Part-Whole	是/为… 的一部分, 由…组成/构成
Has-Property	"…的…是/为"

Table 2. A list of Possible Chinese Terms Relevant to Specific Relations

By using the extraction strategies discussed above, Table 3 provides a list on the number of triples for each type of extraction from different sources.

Source	Is-A	Part-Whole	Has-Property
Baidu Encyclopedia	9,586	115,149	151,941
Hudong Encyclopedia	129,417	98,658	25,513
Chinese Wikipedia	3,809	6,737	2,608
Sina Weibo	40,315	29,625	850,883
Sogou News	13,962	9,806	7,372

Table 3. Unstructured Knowledge from Different Sources

It is noticed that although microblog posts and News pages are always considered as events related resources, through our experimental studies, it shows that there are also many common declarative knowledge which can be extracted from these two types of sources.

Extracted Triples	Source
<表皮层, part-whole, 皮肤> <每一个幸福的今天, part-whole, 人生> <实践部, part-whole, 团总支>	Sina Weibo
<摩托车, Is-A, 机动车> <胆碱, Is-A, 麻醉剂> <纪念碑, Is-A, 文物>	Sogou News
<郁金香, Has-Property, 原产地> <出版社, Has-Property, 负责人> <影片, Has-Property, 主角>	Baidu Encyclopedia

Table 4. Some Examples on Extracted Triples from Different Sources

Although the pattern based extraction strategy is simple, it is straightforward and practical for large-scale information extraction. Table 4 gives some illustrative example on the three types of extracted triples from different sources.

3.4 Conceptual Ontology and Entity Linking

We extract the category level information from Baidu Encyclopedia and Hudong Encyclopedia. We extract all the "*subClassOf*" relationship among classes and create an OWL file for them. We observe the file by protégé tool and find that the Baidu Encyclopedia has a relatively clean hierarchical classification system officially maintained by Baidu Inc., while the Concept-Instance relations are contributed by common users. As shown in Figure 3, there are 13 very broad categories which are with many sub-categories under each of them, and generally there are three to four hierarchical layers. While the hierarchy of Hudong Encyclopedia is with more detailed classification, but the organization of the hierarchy needs improvements (such as self-loops in the ontology [9]).

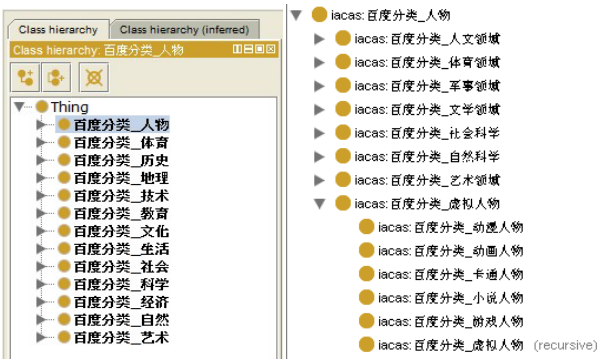


Fig. 3. Hiearchical Ontology of CASIA-KB Baidu Encyclopedia Branch

However, the OWL ontology only contains a well-structured hierarchy with "*rdf:subClassOf*" relationship among classes (concept). It is not enriched with object property, data property, or any constraint information. Therefore, we can use the Concept-Instance relations to perform some statistical analysis and conclude some constraints on the OWL ontology, such as generating Top-K attributes for a given concept or generating domain and range for a specific relation [11].

Although entities are locally managed by their own ontology, many of them appeared in different sources might be semantically equivalent. Hence, entity linking is required for better organization of CASIA-KB. The entity linking is a dynamic process, with entity disambiguation as one of the key step. For example, given a search input "*Apple*", the Baidu Encyclopedia returns several possible candidate entities (such as "Apple" in the fruit category and "Apple" in the company category); the Hudong Encyclopedia also returns several possible candidate entities which share the same semantics with the ones in Baidu Encyclopedia. So the entity linking algorithm

measures the similarity of each two entities and if the similarity value is greater than a threshold, the two are regarded to be the same and linked together. In [14], we developed the CASIA-EL subsystem and a Stepwise Bag-of-Words based entity disambiguation algorithm for CASIA-KB. The linking precision of CASIA-EL is 88.5% based on the Chinese microblog entity linking contest organized by the 2nd Conference on Natural Language Processing and Chinese Computing (CASIA-EL ranked as the 2nd team).

3.5 Knowledge Base Interface and Knowledge Network Visualization

We build a Web-based interface to support the visualization of SPARQL query. The highlighted function is the online graph visualization based on JavaScript, as well as different forms of text outputs. When we just want a text form output, we are supposed to input "*SELECT * FROM*" SPARQL format. If we want a graph based visualized result, we are required to input "*CONSTRUCT {*} FROM*" SPARQL format since *CONSTRUCT* can return a sub-model of the graph model.

Fig. 4. Screenshot of the CASIA-KB SPARQL Endpoint

Figure 4 is a snapshot of the Web interface of the SPARQL Endpoint. By using the default query as an example, if we select the "互动百科" and click "*执行搜索*" (run query), the triple format query result is listed as in Table 5.

Subject	Predicate	Object
Hudong:互动资源_龙	Hudong:名称	"龙"^^<http://www.w3.org/2001/XMLSchema#string>
Hudong:互动资源_龙	Hudong:拼音	"lóng"^^<http://www.w3.org/2001/XMLSchema#string>
Hudong:互动资源_龙	Hudong:繁体	"龍"^^<http://www.w3.org/2001/XMLSchema#string>
Hudong:互动资源_龙	Hudong:英文	"dragon"^^<http://www.w3.org/2001/XMLSchema#string>
Hudong:互动资源_龙	Hudong:笔画	"5"^^<http://www.w3.org/2001/XMLSchema#string>
Hudong:互动资源_龙	Hudong:部首	"龙"^^<http://www.w3.org/2001/XMLSchema#string>
Hudong:互动资源_龙	Hudong:笔顺	"横撇那勾撇点 "^^<http://www.w3.org/2001/XMLSchema#string>
……		

Table 5. An Example of An Query Results

The textual output may be helpful for developers but not intuitive for common users. To improve the visualztion effect, we take advantage of Javascript to demonstrate the result as visualized nodes and edges. To generate a subgraph of the knowledge base, SPARQL "*CONSTRUCT*" query is required instead of "*SELECT*" query. For instance, when we input the query as shown in Table 6:

```
prefix baidubaike:<http://www.ia.cas.cn/baike_baidu/resource/>
CONSTRUCT {?s ?p ?o}
WHERE {baidubaike:百度资源_北海公园 ?p ?o}
```

Table 6. An Example of Construct Query on CASIA-KB

Then, the corresponding subgraph is generated and presented to the end user, as shown in Figure 5.

Besides a SPARQL Construct visualization tool, we also utilze complex network visualization tools such as NodeXL to visualize and analyze the complex knowledge network of CASIA-KB. Figure 6 is a visualization of 10 thousand triples from the Hudong Branch of CASIA-KB. In this figure, we select one normal node and its two-degree adjacency nodes are shown. We can observe that a normal node in the

knowledge network does not have many connections since it is only connected to a few entities through a few triples. Figure 7 shows a pivotal node and its two-degree adjacency nodes in the knowledge network, which indicates that this node connects to many other entities, and only by two degrees, most of the entities are connected together.

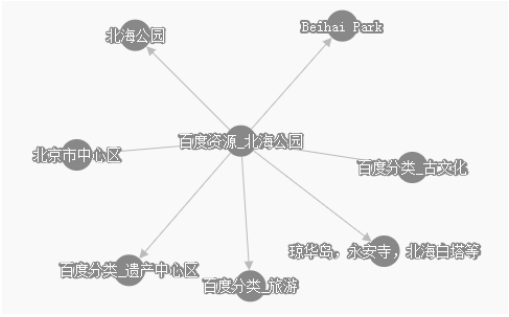


Fig. 5. An Example of a Graph Visualization

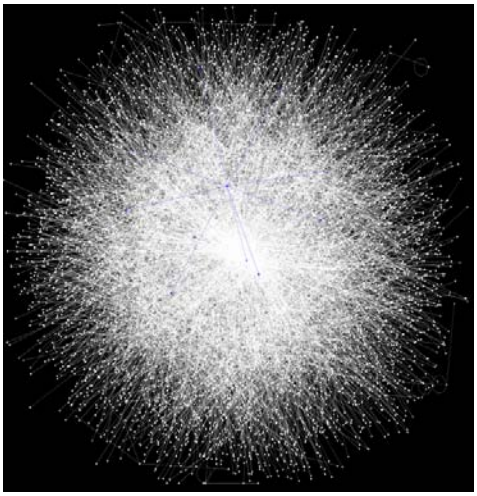


Fig. 6. Knowledge Network Visualization and A Normal Node's Two-Degree Adjacency Nodes

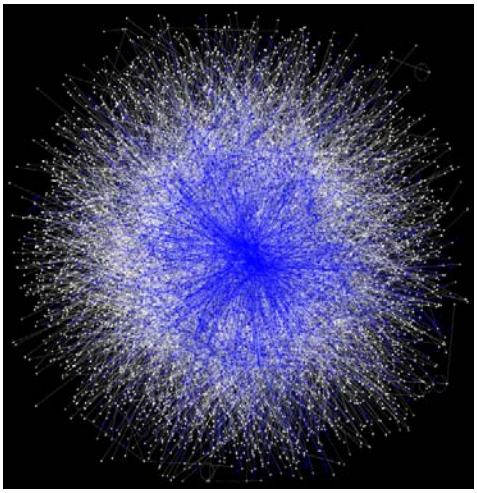
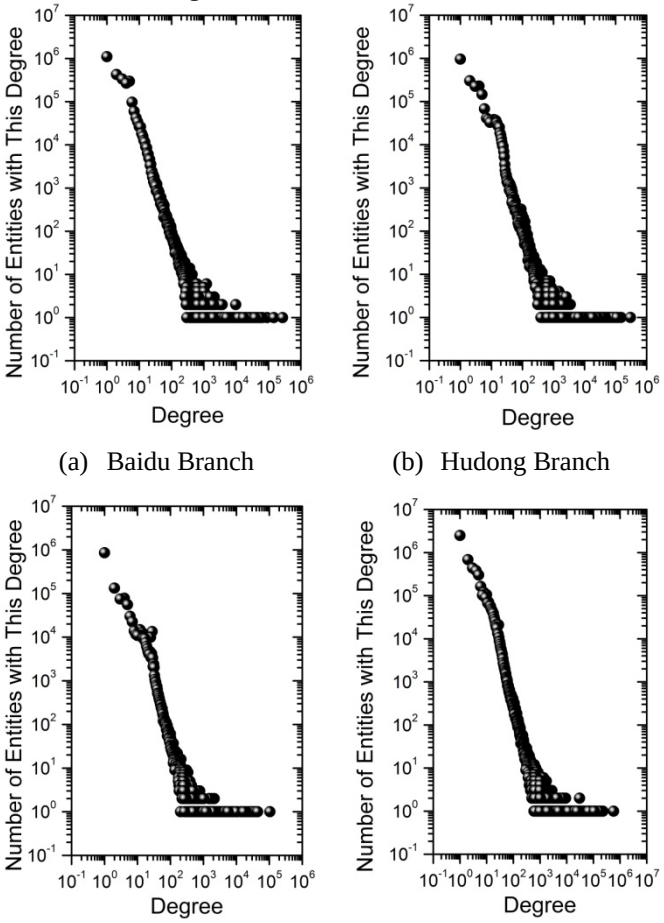


Fig. 7. Knowledge Network Visualization and A Pivotal Node's Two-Degree Adjacency Nodes

In [10], we analyzed the structural properties on the Baidu branch of the CASIA-KB. From the degree distribution perspective, it yields a power law distribution. As a step forward, it is not surprising to see that the Hudong branch and the Chinese Wikipedia Branch also follow power law distributions, as shown in Figure8. When we merge these branches together as one, the whole CASIA-KB also follows a power law distribution (This is consistent with previous findings for structural properties preservation in the network merged from multiple complex networks [15]). This observations show that from a global point of view, although the entities and relevant rela-

tions are from different resources on the Web, there are always several pivotal entities (but not many) which serve as the hub of the whole knowledge base and connect the rest of the entities together. This structure keeps the entire knowledge base as an interconnected Web of knowledge.



(a) Baidu Branch (b) Hudong Branch

(c) Chinese Wikipedia Branch (d) The Whole CASIA-KB

Fig. 8. Degree Distribution of CASIA-KB

4 Conclusion and Future Works

Many semantic knowledge bases are being developed in a variety of areas and multiple languages such as YAGO and Zhishi.me. Also the Chinese knowledge base is constructed before from multiple sources. However, they just extract structured information from Encyclopedia pages.

This paper introduces CASIA-KB, a general purpose Chinese knowledge base built from structured and unstructured resources, including Web-based Encyclopedias, microblog posts and Web News pages, etc. CASIA-KB mainly focuses on declarative knowledge and most of the knowledge is textual knowledge extracted from structured and unstructured sources, such as Web-based Encyclopedias (where more formal and static knowledge comes from), Microblog posts and News (where most updated knowledge comes from). CASIA-KB also aims at bringing in images and videos (which serve as non-textual knowledge) as relevant knowledge for specific instances and concepts since they bring additional interpretation and understanding of textual knowledge. The structured facts are represented in RDF N3 format and stored in Jena TDB, resulting in a large scale of triple-based semantic knowledge base. For triple extraction from unstructured texts, we focus on three types of declarative knowledge, namely, those with Is-A, Part-Whole, or Has-Property relations. The extractions are based on patterns, and further investigations on the extraction quality need to be improved. For this reason, currently, we store knowledge triples from unstructured texts as independent resources. After quality evaluation and improvements, they should be linked to knowledge from structured resources such as Baidu, Hudong, Wikipedia infobox knowledge and corresponding ontologies. In this paper, we focus on common knowledge.

As a step forward, events related knowledge base should be considered as an important branch for CASIA-KB since they can be considered as an important source where most updated knowledge is from. Inspired by YAGO's efforts on extraction of time and location from Wikipedia, each event in CASIA-KB can be with more provenance relation extracted from News contents and microblog posts, such as event time, location, source of the News, etc..

Acknowledgement

The set of news Web pages are from Sogou Lab[3]. The set of Sina Weibo microblog posts are crawled by Bo Xu and Feng Wang from the Computational Brain Group at Institute of Automation, Chinese Academy of Sciences. This study is supported by the Young Scientists Fund of the National Natural Science Foundation of China (61100128).

References

1. Douglas B. Lenat: CYC: A Large-Scale Investment in Knowledge Infrastructure. Commun. ACM 38(11): 32-38 (1995)
2. Christian, B., Jens, L., Georgi, K. Soren, A., Christian, B., Richard, C., and Sebastian, H. 2009. DBpedia - A crystallization point for the Web of Data. Web Semantics: Science, Services and Agents on the World Wide Web. 7, 154-165.

[3] Sogou Lab News Data <http://www.sogou.com/labs/dl/ca.html>

3. Fabian, M.S., Gjergji, K., and Gerhard, W. 2008. YAGO: A Large Ontology from Wikipedia and WordNet. Journal of Web Semantics. 6, 3 (Sep. 2008), 203-217.
4. Catherine Havasi, Robert Speer, and Jason Alonso. ConceptNet: A lexical resource for common sense knowledge. In the book, Recent Advances in Natural Language Processing, Volume 5. John Bernjamins Publishers, Amsterdam & Philadelphia, 2009.
5. Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka Jr., Tom M. Mitchell: Toward an Architecture for Never-Ending Language Learning. AAAI 2010
6. Cungen Cao, Qiangze Feng, Ying Gao, Fang Gu, Jinxin Si, Yuefei Sui, Wen Tian, Haitao Wang, Lili Wang, Qingtian Zeng, Chunxia Zhang, Yufei Zheng, Xiaobin Zhou: Progress in the Development of National Knowledge Infrastructure. Journal of Computer Science and Technology. 17(5): 523-534 (2002)
7. Xing, N., Xinruo, S., Haofen W., Shu, R., Guilin, Q., and Yong, Y. 2011. Zhishi.me - Weaving Chinese Linking Open Data. ISWC 2011. 205-220.
8. Zhichun Wang, Zhigang, Wang, Juanzi, Li, and Jeff Pan. Knowledge extraction from Chinese wiki encyclopedias. Journal of Zhejiang University-SCIENCE C (Computer & Electronics), 1869-1951.
9. Zhichun Wang, Juanzi Li, Zhigang Wang, Jie Tang: Cross-lingual knowledge linking across wiki knowledge bases. WWW 2012: 459-468
10. Yi Zeng, Hao Wang, Hongwei Hao, and Bo Xu. Statistical and Structural Analysis of Web-based Collaborative Knowledge Bases Generated from Wiki Encyclopedia. In Proceedings of the 2012 IEEE/WIC/ACM International Conference on Web Intelligence, Macau, China, December. 4-7.
11. Yi Zeng, Hongwei Hao, and Bo Xu. Entity Conceptualization and Understanding based on Web-scale Knowledge Bases. In: Proceedings of the 2013 IEEE International Conference on System, Man, and Cybernetics, Manchester, the United Kingdom, 2013.
12. Jill Larkin, Herbert Simon. Why a Diagram is (Sometimes) Worth Ten Thousand Words. Cognitive Science, 11, 65-99, 1987.
13. Yangqiu, S., Haixun, W., Zhongyuan, W., Hongsong, L., and Weizhu, C. 2011. Short Text Conceptualization Using a Probabilistic Knowledgebase. In: Proceedings of the 22nd International Joint Conference on Artificial Intelligence. Barcelona, Spain, (Jul. 2011), 2330-2336.
14. Yi Zeng, Dongsheng Wang, Tielin Zhang, Hao Wang, Hongwei Hao. Linking Entities in Short Texts based on a Chinese Semantic Knowledge Base. Proceedings of the 2nd Conference on Natural Language Processing and Chinese Computing, 2013.
15. Xiaoping Sun, Hai Zhuge. Merging Complex Networks. Proceedings of the Seventh International Conference on Semantics Knowledge and Grid (SKG 2011), Beijing, China, October 24-26, 2011.