

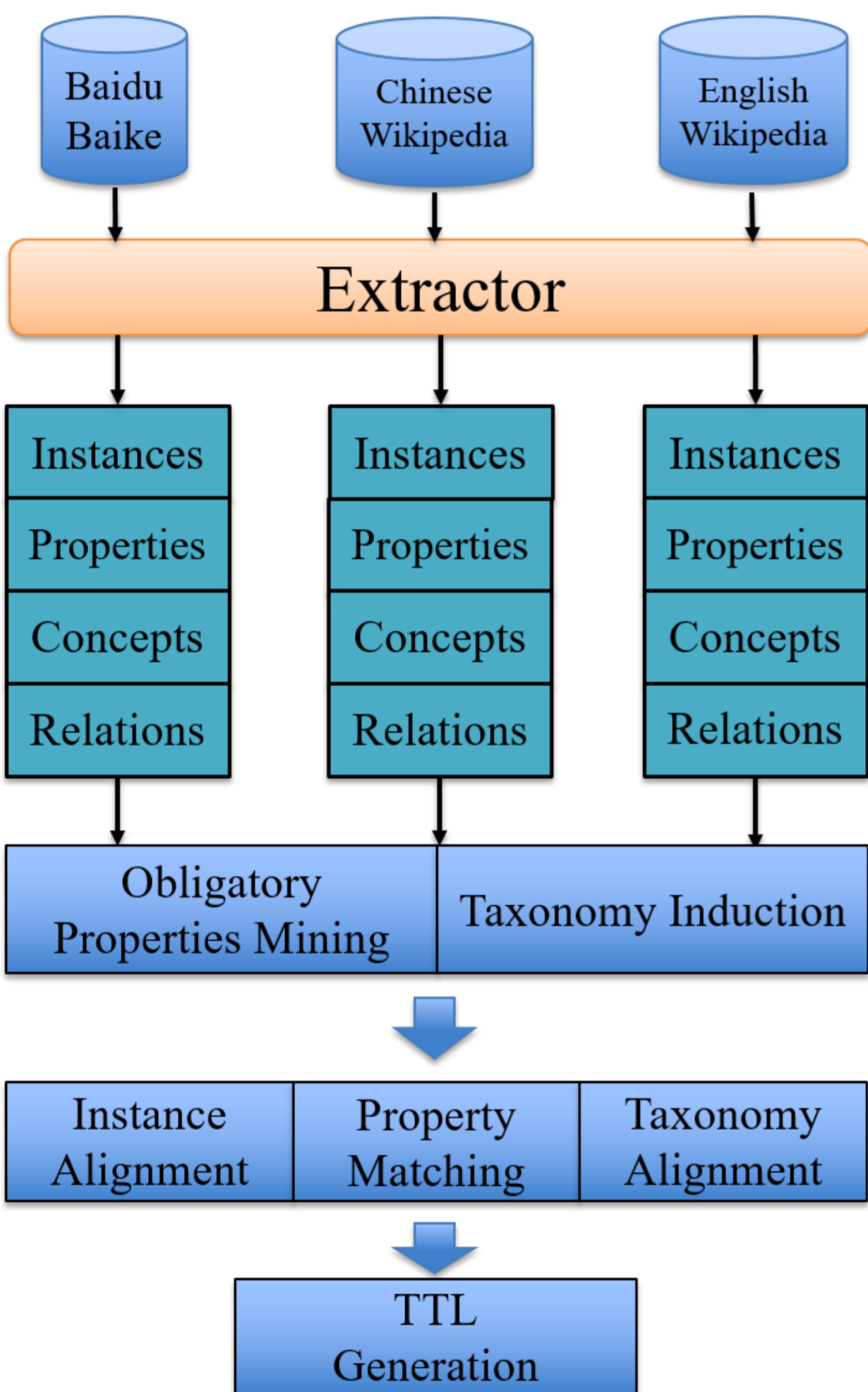
Hailong Jin, Chengjiang Li, Xin Lv, Fangwei Zhu, Shulin Cao, Lei Hou and Juanzi Li
 {jinh15, licj17, lv-x18, zfw15} @mails.tsinghua.edu.cn

◆ Introduction

- Multilingual \mathcal{KB} s are important for the globalization of knowledge sharing. **XLORE** is a large-scale multilingual \mathcal{KB} with balanced Chinese and English knowledge.
- However, there are still three problems in the previous vision of **XLORE**:
 - How to discover more cross-lingual links between instances in different languages?
 - How to correct mistakenly imported *isA* relations (*instanceOf* and *subClassOf*) in \mathcal{KB} s?
 - How to determine obligatory properties in \mathcal{KB} and remove redundant properties?
- We solve the above-mentioned problems in the new vision of **XLORE**, and present it in this poster.

◆ Approach

Framework of XLORE



Cross-lingual Link Discovery

Embedding Based – We propose a heterogeneous network embedding method to project instances into low-dimensional spaces, and then link the possibly same instances based on the distances between them in the unified vector space. We also study Graph Convolutional Network based method (on going).

Translation Based – We translate instance names via Google API from two directions, then calculate the similarities between instances.

Obligatory Properties Mining

Heuristic Rules – We determine obligatory properties in \mathcal{KB} according to their frequency (in template) and attribute value completeness automatically.

Manual Annotation – We employ 20 experienced researchers to manually annotate properties filtered by heuristic rules.

Taxonomy Induction

Heuristic Rules – Guided by multiple heuristics proposed by (Gupta et al., 2016), we induce an unified taxonomy of instances and concepts distilled from the Wikipedia category network. It exploits syntactic evidence in category titles to connect instances with increasingly more general categories.

◆ Experiments

Data Statistics

XLORE harvests **16,284,901 instances**, **2,466,956 concepts** and **446,236 properties** across English Wikipedia, France Wikipedia, Chinese Wikipedia and Baidu Baike. XLORE is available at <http://xlore.org>.

Table1. The number of instances, concepts and properties.

| | Enwiki | Zhwiki | Baidu |
|-----------------|-----------|---------|-----------|
| Instance | 5,300,338 | 887,814 | 8,278,684 |
| Concept | 1,592,543 | 410,022 | 33,127 |
| Property | 31,050 | 11,689 | 400,298 |

Table2. Cross-lingual link between instances in different languages.

| | Enwiki | Zhwiki | Baidu |
|---------------|---------|---------|---------|
| Enwiki | -- | 454,579 | 303,108 |
| Zhwiki | 454,579 | -- | 336,890 |
| Baidu | 303,108 | 336,890 | -- |