

Yan Fan<sup>1</sup>, Chengyu Wang<sup>1</sup>, Guomin Zhou<sup>2</sup>, Xiaofeng He<sup>1</sup>

1. Shanghai Key Laboratory of Trustworthy Computing, School of Computer Science and Software Engineering, East China Normal University  
2. Department of Computer and Information Technology, Zhejiang Police College

Knowledge graph (KG) is a semantic network used to model entities and the relations between them. While most automatically constructed Chinese KGs are general and large-scale, they are insufficient in long-tail entities and relations in specific domains.

To meet the needs of practical applications for a certain domain, we propose a general framework to construct a Chinese domain knowledge graph (DKG). It utilizes Wikipedia pages related to the entertainment industry as data source for demonstration purpose, extracts seed entities and relations from categories and infoboxes to construct an initial DKG, and employs a word embedding based linear projection model to cover more long-tail facts from texts.

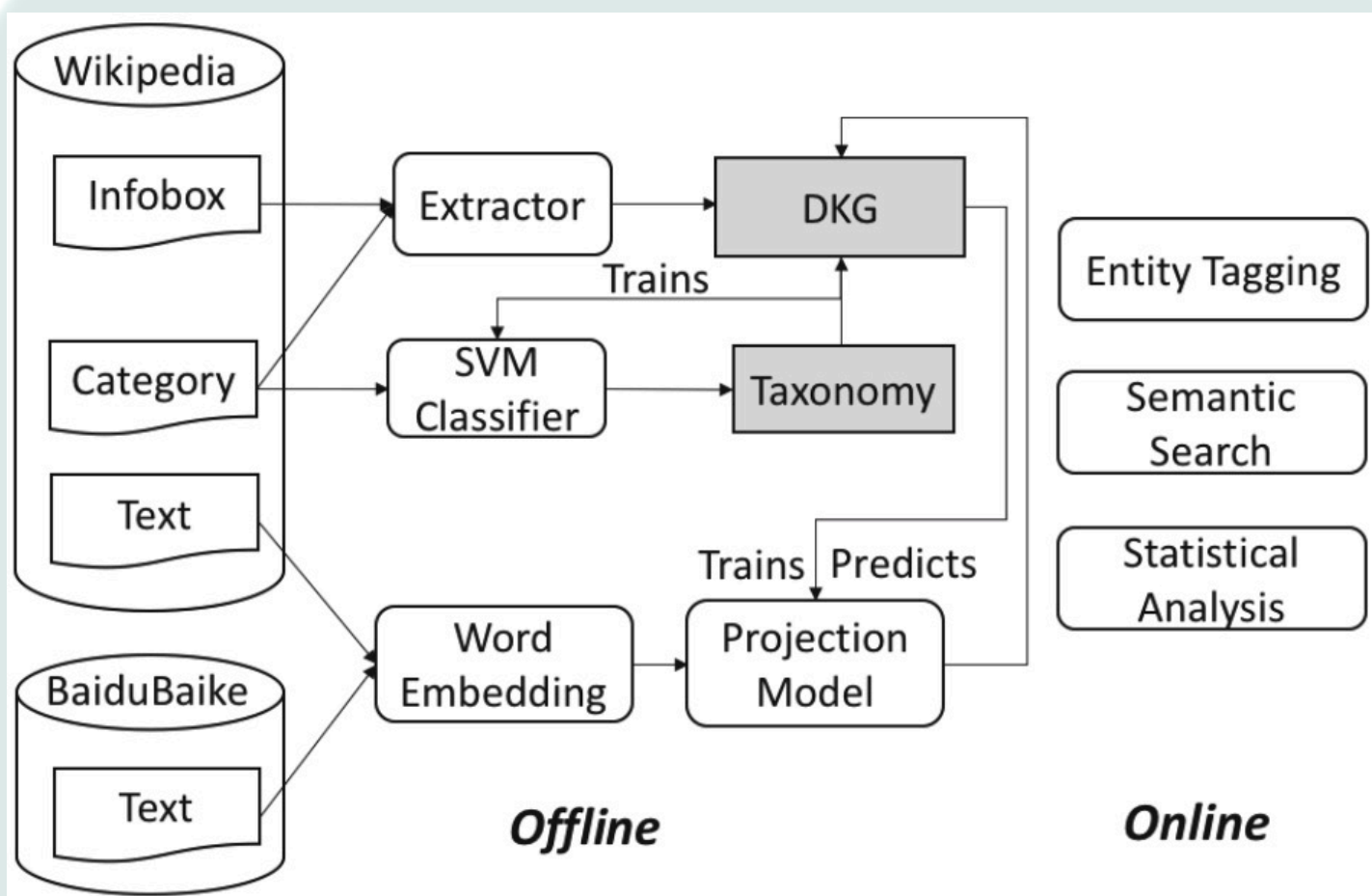


Fig. 1. System Architecture

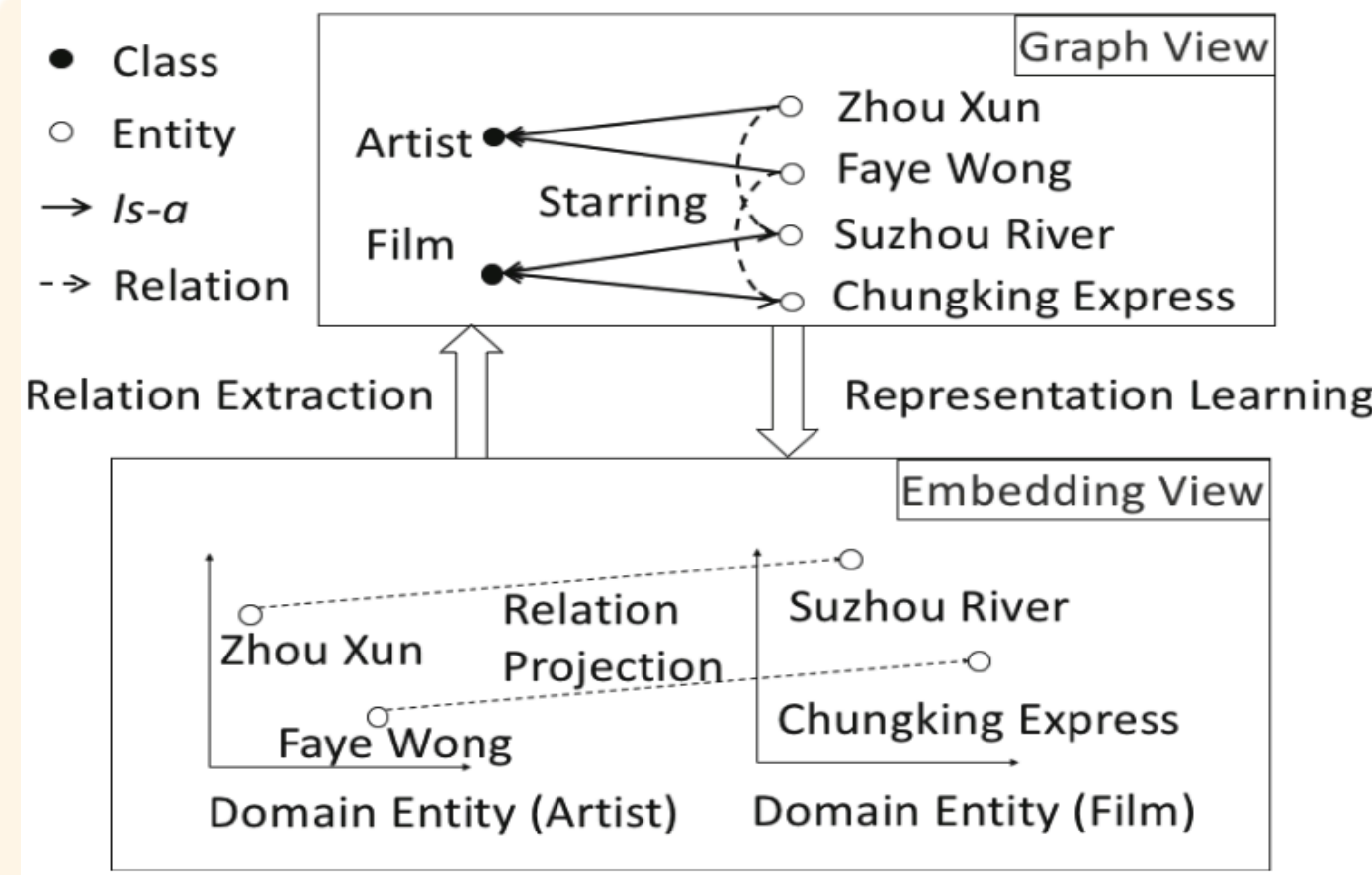


Fig. 2. Representation Learning and Relation Extraction

Table 1. Descriptions of Chinese DKG

|               |         |
|---------------|---------|
| # Entities    | 100,848 |
| # Rel. Facts  | 481,562 |
| # Attr. Facts | 251,183 |
| # Rel. Types  | 46      |
| # Attr. Types | 33      |
| Avg Acc.      | 93.1%   |

DKGBuilder has an offline module and an online module, responsible for DKG construction and demonstration respectively.

The offline module consists of three parts: 1) seed knowledge graph construction, which takes a couple of human-defined template names from Wikipedia to obtain domain entities and extracts seed attributes and relations with a simple approach of pattern matching; 2) fine-grained entity categorization, a way to construct domain taxonomy via is-a relation classification; and 3) representation learning and relation extraction, aiming at harvesting long-tail domain facts from text by a word embedding based linear projection model under distant supervision.

The online module provides services like semantic search, deep reading, etc.



Fig. 3. Deep Reading



Fig. 4. Semantic Search