

KnowFi: Knowledge Extraction from Long Fictional Texts

Cuong Xuan Chu, Simon Razniewski, Gerhard Weikum

Knowledge Extraction (e.g. Relation Extraction)

Real-world domain

Fictional domains

Input



Output

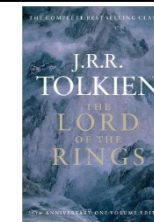
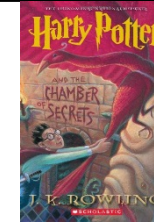
General-purpose encyclopedic knowledge

- Prominent **people, places, products, etc.**
- Basic relations of wide interest
birthplaces, spouses, writing a books, etc.

Example

Harry Potter is a series of seven fantasy novels written by British author **J.K. Rowling**. -- *Harry Potter* Wikipedia page --

Extraction: **<Harry_Potter, hasAuthor, J.K. Rowling>**



Narrative-fiction Knowledge

- Prominent **characters, battles, weapons, etc.**
- Relations: traits of characters,
key elements of the narration
allies, enemies, skills, member of clans, etc.

Harry had been a year old the night that **Voldemort** - the most powerful **Dark** wizard for a century, a wizard who had been gaining power steadily for eleven years, arrived at his house and **killed his father and mother**. **Voldemort** had then **turned his wand on Harry**; he had performed the **curse** that had **disposed** of many full-grown witches and wizards in his steady rise to power and, incredibly, it had not worked. Instead of **killing** the small boy, the curse had rebounded upon **Voldemort**. -- *Harry Potter* book --

Extraction: **<Harry, hasEnemy, Voldemort>**

Challenges

1. Lacking of training data for *non-standard relations* that have only few seed triples.
2. Working on *very long input texts*, such as entire books, where relevant cues for RE are spread across passages.

Methodology

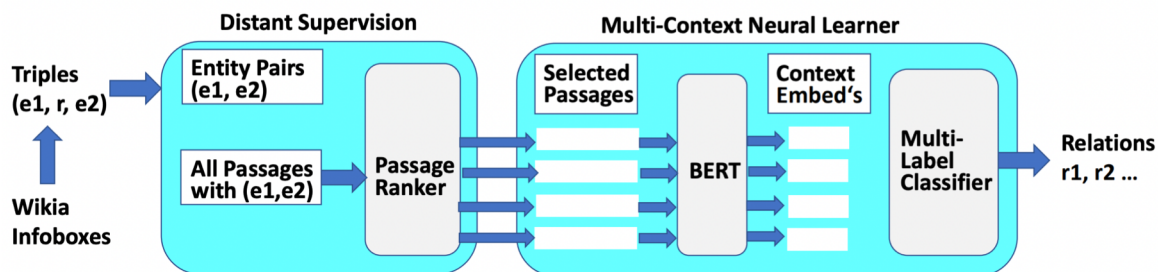


Figure 1: Overview of the KnowFi architecture.

Results

- **LoFiDo corpus** for RE in Long Fictional Documents
- Experiments on 5 diverse fictional domains

Models	Books					Wikia Texts				
	LoTR	GOT	HP	WP	Avg.	LoTR	GOT	HP	WP	Avg.
BERT-Type (Shi and Lin)	0.01	0.54	0.09	0.11	0.19	0.09	0.12	0.15	0.19	0.14
BERT-EM (Soares et al.)	0.45	0.66	0.37	0.29	0.44	0.70	0.78	0.48	0.50	0.62
GLRE (Wang et al.)	0.27	0.25	0.56	0.47	0.39	0.37	0.56	0.71	0.56	0.55
KnowFi	0.45	0.76	0.55	0.50	0.57	0.71	0.83	0.71	0.67	0.73

Homepage: <https://www.mpi-inf.mpg.de/yago-naga/knowfi>