Burning Glass Technologies: Skill Extraction and Building a Skill Knowledge Graph

Bhuvana Surapaneni
Ly (Harriet) Bui
Rishi Mody
Rishikesh Jha

OUTLINE

- Extract new skills using Wikipedia
- Knowledge base population and visualization using Wikipedia and Wikidata

METHODS

I. Extracting New Skills:

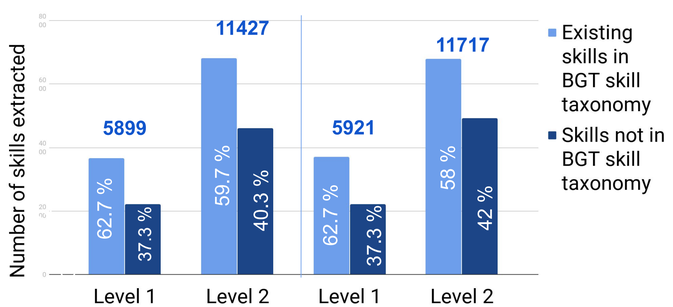
SpaCy Named Entity Recognition Model

Convolutional layers with residual connections, layer normalization and maxout non-linearity, use an imitation learning objective

Experiment 1 Trained on job posting data

Experiment 2 Updated model with Wikipedia training data: 36772 Wikipedia articles

Results

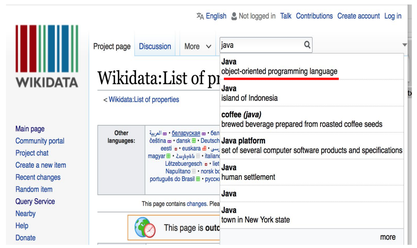


Accuracy for skills extracted	Experiment 1	Experiment 2
Existing skill extracted	82.3%	83.1%
New skill extracted	75.1%	76%

II. Constructing Knowledge Graph:

Dataset: Wikidata RDF triplets of selected properties

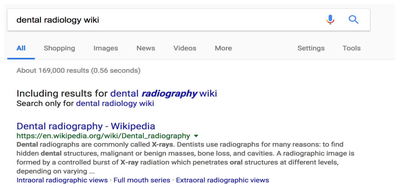
JAVA	java
JAVA	java software
JAVA 1.2	java 1 . 2
Java Applets	java applets
Java Card	java card
Java Concurrent Programmr	concurrent programming
Java Enterprise Edition	enterprise edition

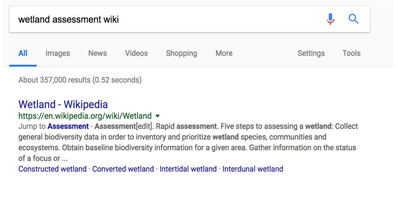


BGT Taxonomy and Wikidata intersection

- Found Wikidata entries for 7121(out of 16K) 43.67% coverage on Canonical skills
- 1814 matches are scientific journals

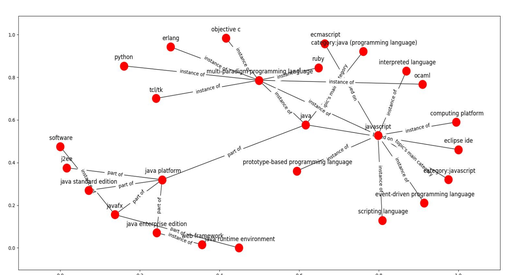
Improving and increasing intersection using google search API





Building Knowledge Base

- Extract RDF triplets of shortlisted properties corresponding to the mapped skills
- Used Apache Jena to create a database of RDF Triples using its Graph interface
- Created API to find relationship between two skills



III. Knowledge Base Population:

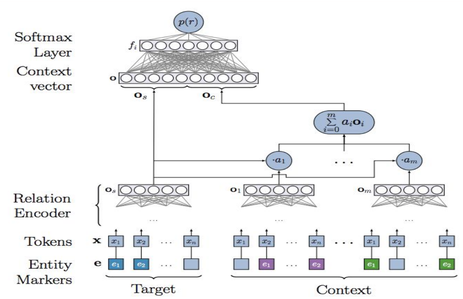
Dataset: Wikidata and Wikipedia

- Used Distant Supervision to create dataset for relation extraction
- Created feature vector by concatenating GloVe word embeddings with marker embeddings

Results using LSTM(Baseline)

Relation Type	Precision	Recall
Part of	0.589	0.503
Instance of	0.761	0.837
Subclass of	0.513	0.468

Context Aware LSTM Model



Future Work

- Deduplication of the extracted skills
- Creating interactive UI for exploring skill relations from knowledge graph
- Exploring different approaches for improving relation extraction model

REFERENCES

- Zhao et al. SKILL: A System for Skill Identification and Normalization. AAAI 2015.
- https://engineering.linkedin.com/blog/2016/10/building-the-linkedin-knowledge-graph
- Sorokin.et al. Context-Aware Representations for Knowledge Base Relation Extraction. Proceedings of the 2017 Conference on Empirical Methods in NLP, pages 1784–1789
- Traylor et al. Learning String Alignments for Entity Aliases. NIPS 2017