

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

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 2

Trained on job posting data

Updated model with Wikipedia training data: 36772 Wikipedia articles

Results



II. Constructing Knowledge Graph: Dataset: Wikidata RDF triplets of selected properties

JAVA JAVA	java java software	WIKIDATA	Project page C Wikidata < Wikidata:List	Discussion ata:Lis	More ~	java Q Java object-oriented programming language Java island of Indonesia
Java Applets Java Card Java Concurrent Program	java applets java card	Main page Community portal Project chat Create a new item Recent changes Random item	Other languages:	العربية <u>беларуская</u> m čeština m • dansk Ħ • D eesti n • euskara d magyar Ħ • Հայերեն Ħ •	<u>مع • قع • قع • قع مع مع</u>	coffee (java) brewed beverage prepared from roasted coffee seeds set Java platform brewed several computer software products and specifications land
lava Enterprise Edition	enterprise edition			Lëtzel Napul português do	Lêtzebuergesch • • liei Napulitano 🗄 • norsk b português do Brasil 🗄 • русск	Java human settlement Java
		Query Service Nearby Help Donate		This page conta	ins changes. Plea page is out d	Java town in New York state more



- 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

_		wetland assessment wiki	୍ ପ୍			
dental radiology wiki	الب م					
All Shopping Images News Videos More	Settings Tools	All Images News Videos Shopping More	Settings Tools			
About 169,000 results (0.56 seconds)		About 357,000 results (0.52 seconds)				
Including results for dental <i>radiography</i> wiki Search only for dental radiology wiki		Wetland - Wikipedia https://en.wikipedia.org/wiki/Wetland Jump to Assessment - Assessment[edit]. Rapid assessment. Five steps to assessing a wetland: Collect general biodiversity data in order to inventory and prioritize wetland species communities and				
Dental radiography - Wikipedia https://en.wikipedia.org/wiki/Dental_radiography - Dental radiographs are commonly called X-rays. Dentists use radiographs fo hidden dental structures, malignant or benign masses, bone loss, and cavitie formed by a controlled burst of X-ray radiation which penetrates are structure	or many reasons: to find es. A radiographic image is	ecosystems. Obtain baseline biodiversity information for a given area. Gather information on the status of a focus or Constructed wetland · Converted wetland · Intertidal wetland · Interdunal wetland				
depending on varying Intraoral radiographic views · Full mouth series · Extraoral radiographic views						

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



Bhuvana Surapaneni Ly (Harriet) Bui Rishi Mody Rishikesh Jha

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-knowled ge-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