



# Skill Me

Extract Skills from Job Descriptions

## Yue Teng

Insight Data Science Fellow



# How to best allocate skills training dollars?

- **40%** employers cannot find the talents they need
- Finding **personalized** career and learning pathways is hard
- **AI-powered** platform that matches people, business and learning resources





# Skill Me



NLP-based model that extracts **job skills** from **course descriptions**

# Data

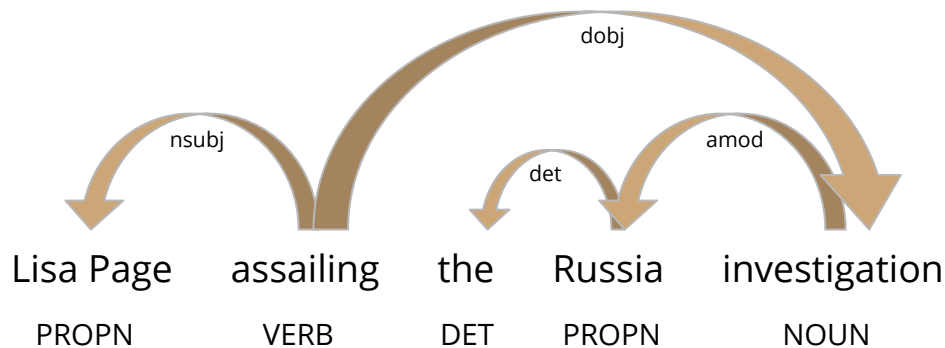
- 3164 courses with title and descriptions
- 853 (26.9%) labeled with skill tags
- 224 unique predefined skill tags

Title	Description	Skill
Energy Industry Fundamentals Certificate	The Energy Industry Fundamentals course, for which...	Energy Industry
Principles of Health Law and Regulatory Issues	This course is part of the Certified Lifestyle Medicine Executive ....	Legal Industry Knowledge
...	...	...

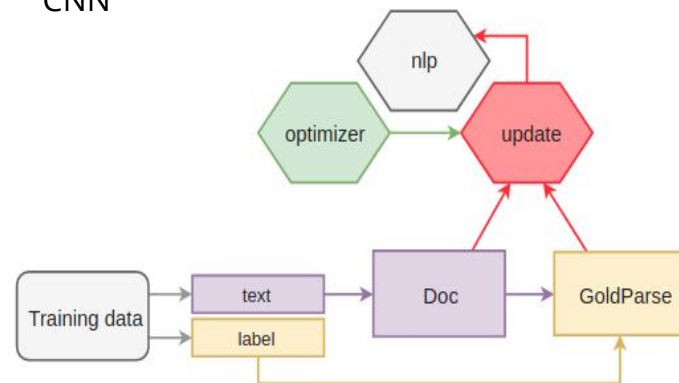
# Named Entity Recognition (NER)

Lisa Page — in **PERSON** assailing the **Russia GPE** investigation as an illegitimate "witch hunt." Mr. **Strzok PERSON**, who rose over **20 years DATE** at the **F.B.I. GPE** to become one of its most experienced counterintelligence agents, was a key figure in **the early months DATE** of the

## Semantic Dependency Parsing



## CNN



# Training NER

- Exact matches
  - Train NER on skill labels
  - 497 (58%) courses
- Non-exact matches
  - Generated keywords by [TextRank](#)
  - Train NER on keywords
  - Map keywords back to skill labels
  - 356 (42%) courses

Title	Description	Skill
Energy Industry Fundamentals Certificate	The Energy Industry Fundamentals course, for which...	Energy Industry

Title	Description	Skill
Principles of Health Law and Regulatory Issues	This course is part of the Certified Lifestyle Medicine Executive ....	Legal Industry Knowledge

# Validation

- ◆ Semantic distance between model **predicted skill** and true **skill tag**
- ◆ Word embeddings: **word2vec** pretrained model (enwiki + news)
- ◆ Training
  - Mean cosine similarity: .90
  - Standard deviation: .21
- ◆ Test
  - Mean cosine similarity: .73
  - Standard deviation: .30

For example...

Predicted Skill	Skill Tag	Similarity
Programming	Software Engineering	0.92



# SKILL ME

## Extract Skills from Course Descriptions

Go!

Input



# SKILL ME

## Extract Skills from Course Descriptions

Go!

Input

The non-technical Soft Skills course consists of online learning, including practice scenarios. Five technical courses consist of hands-on classroom training followed by online practice scenarios. Technical Courses    Electrical Maintenance and Repair    Plumbing Maintenance and Repair    Heating, Ventilation and Air Conditioning (HVAC) Maintenance and Repair    Appliance Maintenance and Repair    Interior and Exterior Maintenance and Repair Non-Technical Online Course Modules: Welcome to the Industry Customer Service Teamwork Time & Project Management Money Matters Documentation & Paperwork Maintenance & Emergencies Safety First Compliance Enhanced with FranklinCovey Insights on Demand video-based online courses Programs may be taken as standalone seminars or in full to earn the industry credential.

Found Something 😊

Skill



# SKILL ME

## Extract Skills from Course Descriptions

INPUT: course title / description

Go!

Input

Discover key moments in American history after the Civil WarAmerican history is vast and complex. Through this course you will explore some of its key moments - examining domestic history and foreign relations.You will investigate the background to the modern history of the United States, including the Civil War, Reconstruction, segregation and the black experience, and US expansion and foreign policy into the 19th century.This course is for anyone interested in American history. You don't need any previous experience.

Found Nothing

It seems that the course description you just provided is about skills that are out of the realms the model was trained for. The model was trained on predefined skills determined by FutureFit AI's database.

Skill

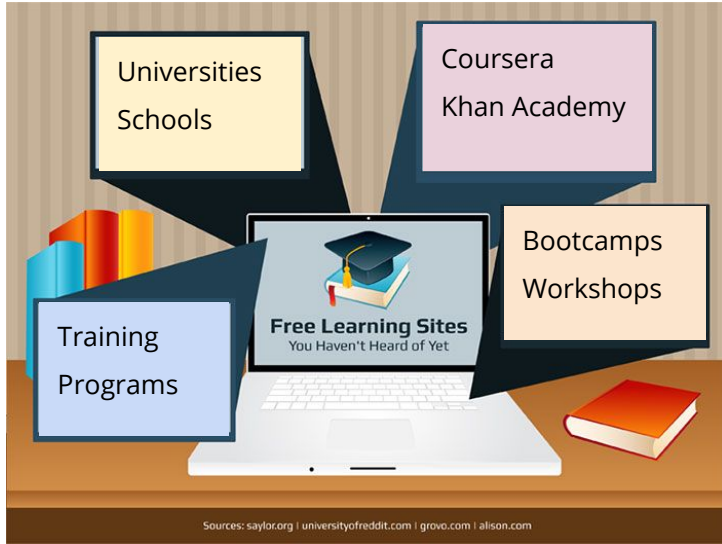
[']

# Skill Me



NLP-based model that extracts **job skills** from **course descriptions**

# What's Next



AI-powered platform that matches people, business and learning resources

# Yue Teng



B. A. in English Linguistics and Literature  
Tsinghua University



M.Phil in Psychology  
Chinese University of Hong Kong



Ph.D in Psychometrics and Quantitative  
Psychology  
Fordham University

$$P_i(X_i = 1 | \theta_j) = c_i + (1 - c_i) \frac{\exp[Da_i(\theta_j - b_i)]}{1 + \exp[Da_i(\theta_j - b_i)]}$$

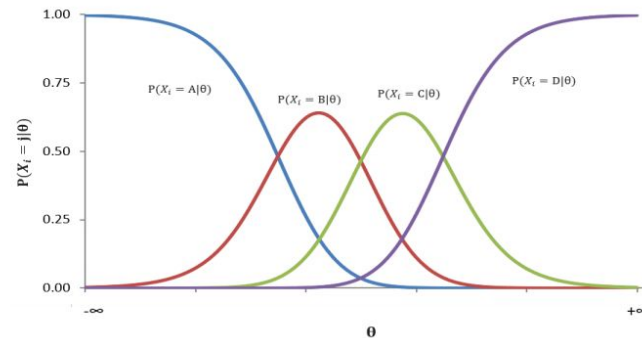
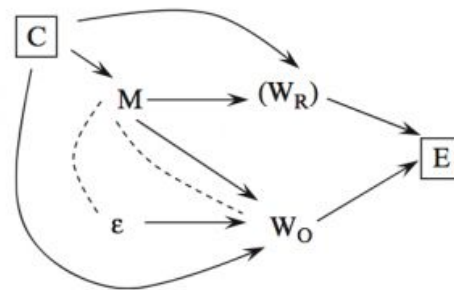


Figure 1: Item Characteristic Curves for a hypothetical item with four answer options, calibrated with the Partial Credit Model.

Thank You

# spaCy NER

- Training data

```
[('Energy Industry Fundamentals Certificate',  
 {'entities': [(0, 15, 'SKILL')]}),  
 ('NSF Certificate in Food Safety Programs',  
 {'entities': [(19, 30, 'SKILL')]}),  
 ('Certified Food Safety Manager Training', {'entities': [(10, 21, 'SKILL')]}),  
 ('Certified Logistics Technician Courses', {'entities': [(10, 19, 'SKILL')]}),  
 ('Certificate for Apartment Maintenance Technicians Courses',  
 {'entities': [(16, 37, 'SKILL')]})]
```

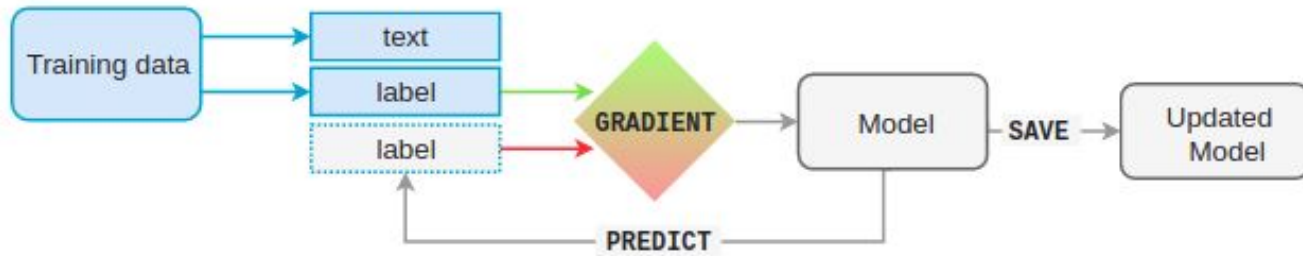
- Exact matches of skill label and words in title (n=497)

	title	description	skill
5	Energy Industry Fundamentals Certificate	The Energy Industry Fundamentals course, for w...	Energy Industry
6	Insurance Studies Certificate	The Certificate of Completion (CCL) in Insuran...	Insurance Industry Knowledge
10	BOC Level I Training Certificate of Completion	Level I classes support the BOC operator skill...	HVAC
11	Official (ISC)² CISSP Training	Our training helps you fully prepare for this ...	Cybersecurity Strategy
12	NSF Certificate in Food Safety Programs	NSF International, NSF Training & Education Se...	Food Safety
13	Certified Food Safety Manager Training	Certified Managers have a responsibility to sh...	Food Safety
42	Professional Development Certificate in SQL Se...	An 11-week in-depth program designed to provid...	Microsoft SQL



# spaCy NER

- spaCy v2.0's Named Entity Recognition system features a sophisticated word embedding strategy using subword features and "Bloom" embeddings, a deep convolutional neural network with residual connections, and a novel transition-based approach to named entity parsing.



# Challenges

- Choose which NLP approach
- Keywords generation: TextRank vs manual
- Validation
  - Similarity in meaning but not exactly the same words
  - Non-overlapping between training and test set
- Generalizability
  - Expand training data

# Model: Input / Output

```
test_text = '''
The non-technical Soft Skills course consists of online learning, including practice scenarios.

Technical Courses

    Electrical Maintenance and Repair
    Plumbing Maintenance and Repair
    Heating, Ventilation and Air Conditioning (HVAC) Maintenance and Repair
    Appliance Maintenance and Repair
    Interior and Exterior Maintenance and Repair

Non-Technical Online Course Modules:

Welcome to the Industry
Customer Service
Teamwork
Time & Project Management
Money Matters
Documentation & Paperwork
Maintenance & Emergencies
Safety First
Compliance

Enhanced with FranklinCovey Insights on Demand video-based online courses
Programs may be taken as standalone seminars or in full to earn the industry credential.

'''
```

```
(base) → insight python test_model.py
Loading from skill_recognition/models
SKILL Exterior Maintenance
SKILL Customer Service
(base) → insight
```

[Apartment Maintenance]

<https://skillfinder.appspot.com/>

The Energy Industry Fundamentals course, for which the credential is based, takes a comprehensive look at the energy industry, including nuclear, natural gas, and renewable. Instructor guides and student materials are provided as part of the program, focusing on experiential learning techniques with a comprehensive online assessment at the end of the course. Students will be required to pass the assessment to receive the certificate. Since CEWD is a coalition made up of energy companies, this credential is being created by the industry for the industry. This ensures that students are learning the right material that will help them succeed at their jobs. Energy Industry Fundamentals is divided into modules. For those who want to offer the opportunity for their students to earn the credentials, all modules must be completed. This may not be the case in all situations, so schools or other training organizations can just use specific modules to fill in what they are already doing.

[Energy Industry]

NSF Certificate in Food Safety Programs

[Food Safety]

Acquire the skills and knowledge necessary to influence the hyperconnected digital landscape. Become fluent in programming languages. Learn about data structures and algorithms, networks and operating systems. Study software engineering, computer vision, robotics, artificial intelligence and platform-based development. Explore human-computer interaction, computer graphics and virtual reality.