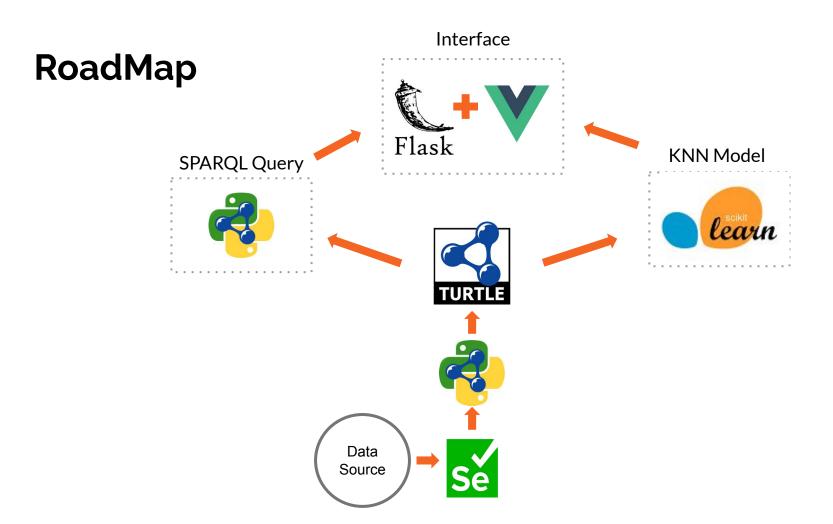
Career-Skill Knowledge Graph

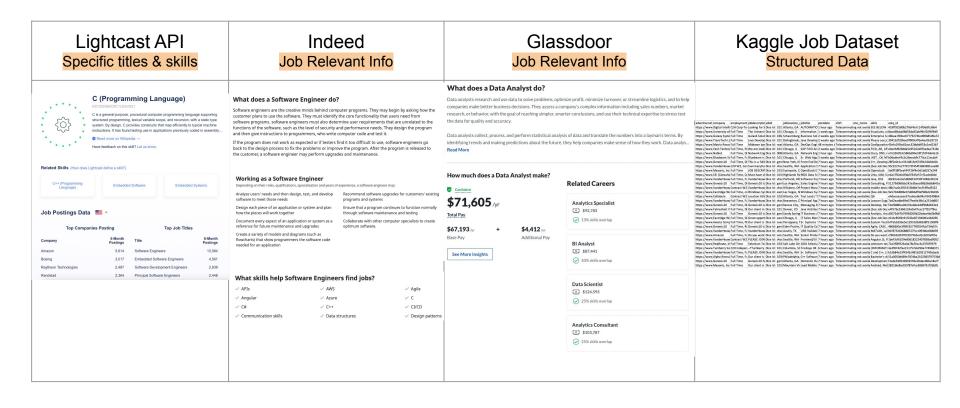
2022 Fall DS558 By Ruijie Rao, Linkai Shao

Domains & Goals





Data



Crawling

- **55,797** Titles
- **100+** Hours
- **Selenium** Headless Browser
- Python Concurrent Threading

10 Parallel Headless Selenium Drivers





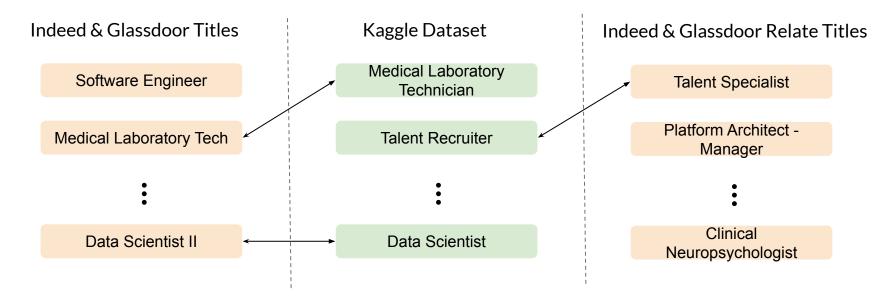




ID	titleName	salary	description	responsibility	relatedTitles	requiredSkills	
K01	Data Analyst	\$99,120	This job	You are required	Data Engineer, Applied Data	Python, Data Mining, Data Science	

__

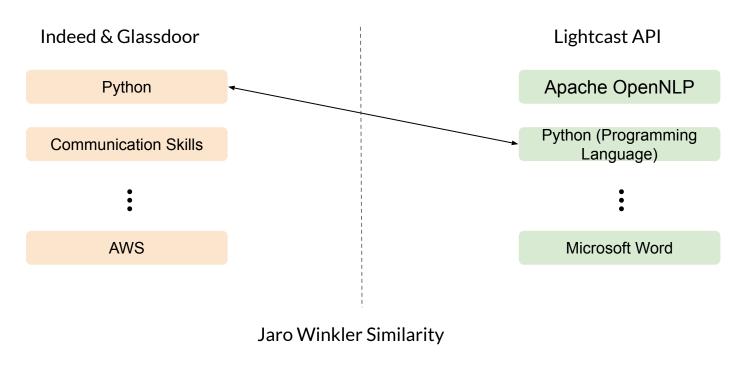
Job Title Entity Resolution

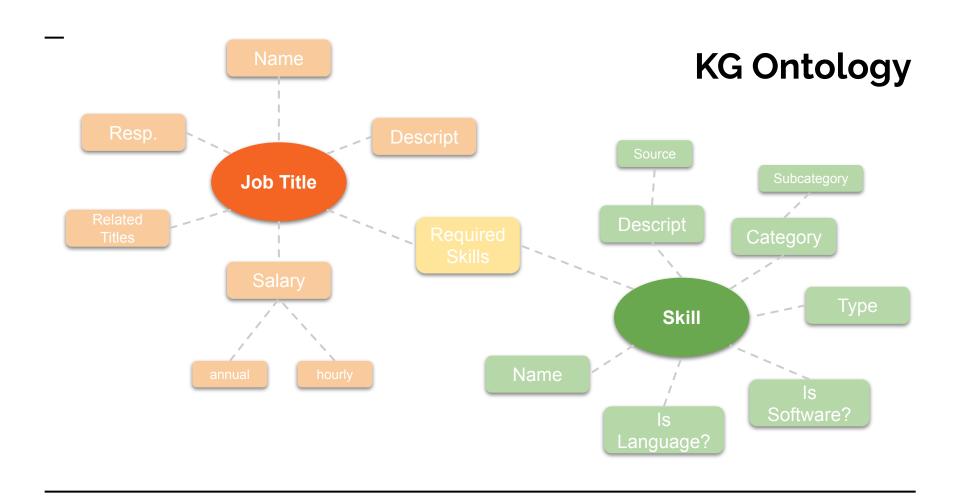


Jaro Winkler Similarity & Matched Skills

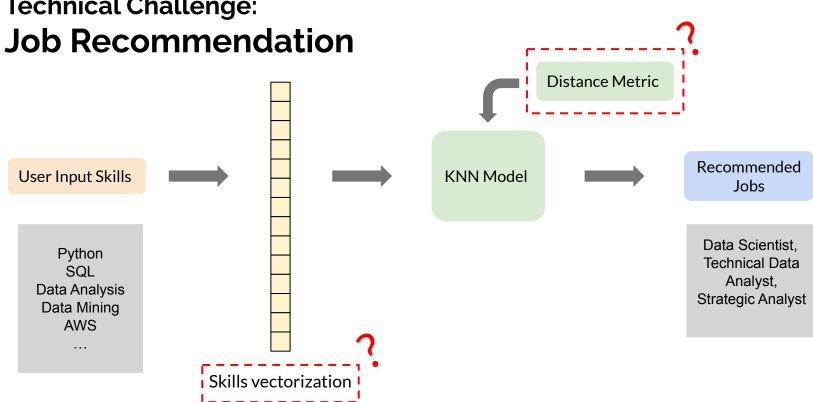
__

Job Skill Entity Resolution



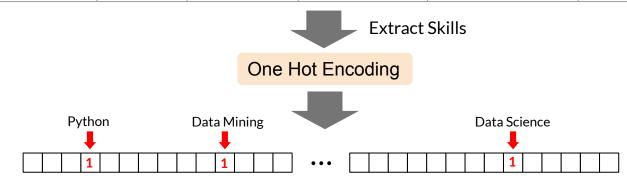


Technical Challenge:

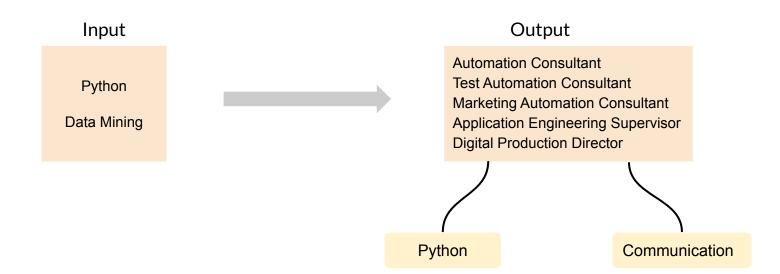


Data Preprocessing: Job Recommendation

ID	titleName	salary	description	responsibility	relatedTitles	requiredSkills	
K01	Data Analyst	\$99,120	This job	You are required	Data Engineer, Applied Data	Python, Data Mining, Data Science	

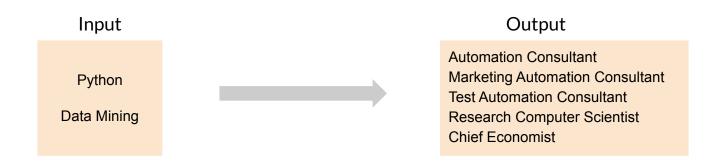


Try #1: Euclidean Distance



_

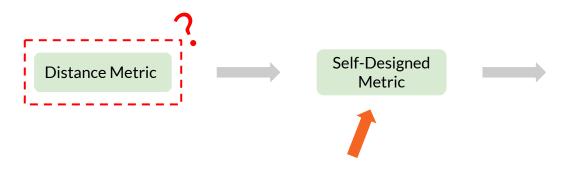
Try #2: Jaccard Distance



Penalize on jobs which have more skills when users input only a few skills

__

Final: Metric Set Up



Data Analyst/Developer Provider Data Analyst Applied Data Scientist Strategic Data Analyst

Applied "bitwise and" between two encoding vectors to compute similarity

_

Evaluation:

Job Recommendation

Approach 1:

Random combination of skills \rightarrow 100% accuracy

Approach 2:

User Research Investigation \rightarrow 20 Interviewees \rightarrow 95% accuracy

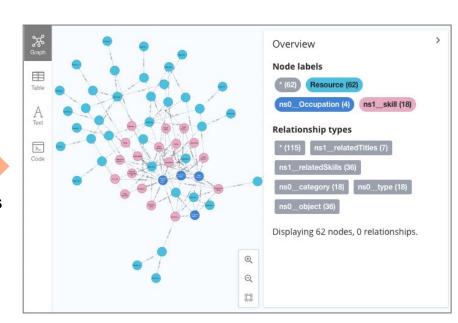
Skills	Result					
Data Modeling	Data Warehousing	Apache Hive	Big Data	-	-	Data Engineer, ETL Architect
Accounting	Computer Sales	Communication	Accounting Software	Balance Sheet	Analytical Skills	Control Accountant, Tax Accountant

Visualization

```
career_title:ET0004AFC23172193E a schema:Occupation;
   career:annualSalary "72408.0"^^xsd:float;
   career:relatedSkills emsi skill:BGS894FE73CB92723FD8.
        emsi_skill:ES8447BA722C50F8F852,
        emsi skill:ESA9372A0C2E0E70C228,
        emsi_skill:KS1218W78FGVPVP2KXPX,
        emsi_skill:KS1220H6CP1BQGH9STNK,
        emsi skill:KS122556LM0829GZCCRV.
        emsi skill:KS122ST5X31W4TVGPN0B.
        emsi skill: KS122VW78Q67XR4C3ZDW,
        emsi_skill:KS124G079QNYPLZY6TQL,
        emsi skill:KS124JB619VXG6RQ810C,
        emsi_skill:KS125025Z8ZHJCJMPF7T,
        emsi skill:KS1259Z6F890LDXJRXYW.
        emsi skill:KS125S36C8SF0912VYBS.
        emsi skill:KS1267F6MSPN366LX7ST,
        emsi skill:KS126HY6YLTB9R7XJC4Z,
        emsi skill:KS127MG77XGJ4T4R4Y44,
        emsi_skill:KS1281L64DB6GGCS37X4.
        emsi_skill:KS440F26FJ79XS9M0QW1,
        emsi skill:KS4425C63RPH46FJ9BX7.
        emsi skill:KSYIJN0ID1F7P2HPDUCJ ;
   career:relatedTitles career_title:ET5AD1CE488B4D20AA,
        career_title:ET5F8A23E938AF6E85,
        career title: ETABBAE9948F7A41C0.
        career_title:ETD3D7CCAFE65E0FFC;
    schema: name "Industrial Security Manager" .
```

Neo4j

Neosemantics



_

KG Quality Assessment

Quality dimensions

Accuracy

Intrinsic

Is data **free of error** & correctly describing facts in the real-world?

Objectivity

Is data are free of bias & impartial?

Reputation

Is data **trustworthy** & enough referenced?

Consistency

Does data form a **consistent** knowledge representation?

Timeliness

Contextual

Is data sufficiently up to date?

Completeness

Does data have **sufficient breadth**, depth, & scope?

Interpretability

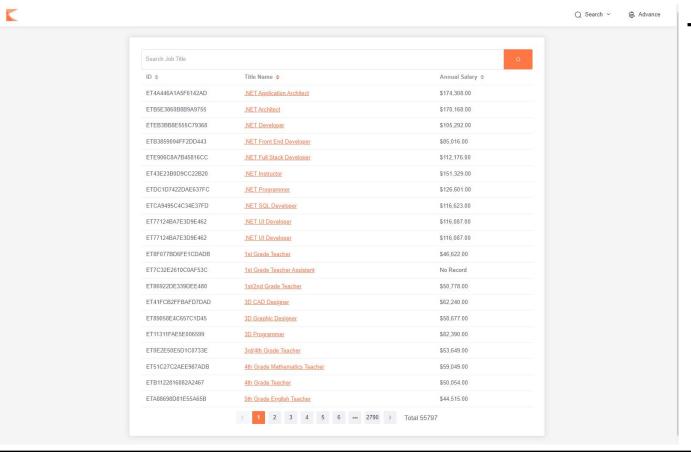
Representational

Can data can be interpreted by machines & humans without ambiguity?

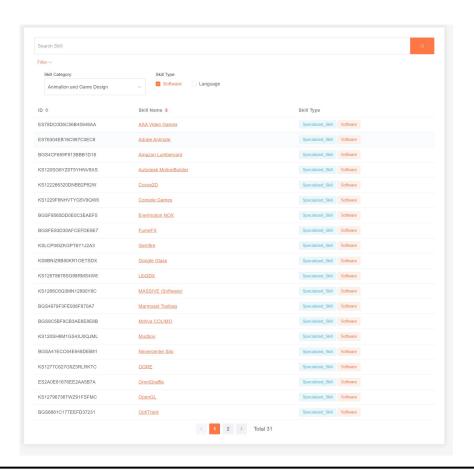
Interlinking

Accessibility

Is data **sufficiently interlinked to other resources**, either within or without Wikimedia?



Interactive App: Title Query



Interactive App: Skill Query

Job Title

.NET Application Architect

A .NET developer is an information technology professional who uses the .NET Microsoft framework to design and maintain software and applications. They collaborate with computer scientists, other web developers and clients to create a user-friendly, scalable application. They are involved with creating original code and providing consistent user support after the application is released.

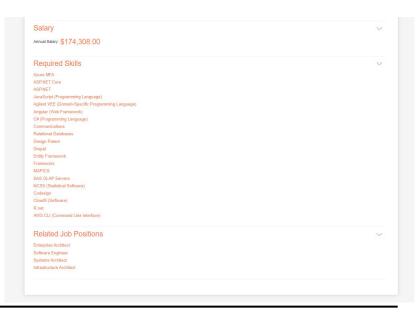
A Software Architect serves as the team leader for the innovation and creation of new company software and program features. Generally found in techheavy industries and large corporations, a Software Architect will oversee a team of programmers. They seek to manage and facilitate the proper design of large websites or programs within products. From software coding to innovative design, a Software Architect is the key to creating a seamless software experience for customers... A Software Architect will usually hold an advanced degree in Computer Science or Computer Engineering, A Software Architect works on both technical and design aspects of a project. Creativity as well as a detailed understanding of coding and leadership skills are essential. Additionally, employers may seek to hire those with specialized certifications in computer languages such as Java or PHP. A proficient Software Architect will possess leadership qualities and seek to continually improve their skill-set.

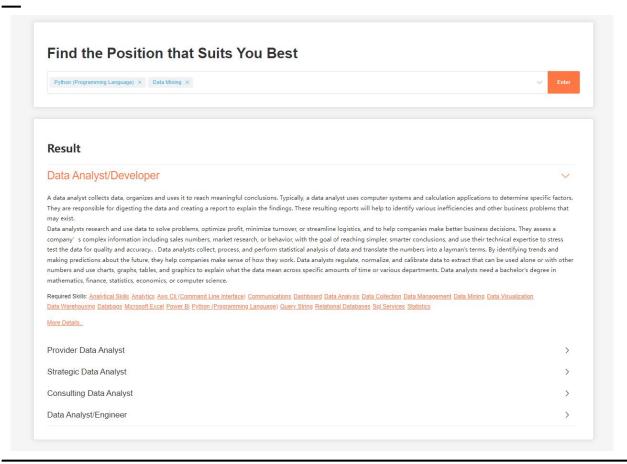
Details

Responsibility

- Designing interfaces for client use
- · Identifying the cause of application errors and coming up with solutions
- Creating code for interacting with a computer or web program
- · Providing user support and implement feedback
- Testing software and making adjustments to improve performance

Interactive App: **Details**





Interactive App: Advance Query