U.PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Retrieving and Processing Information on Company Reviews

Professor:

Sara Fernandes

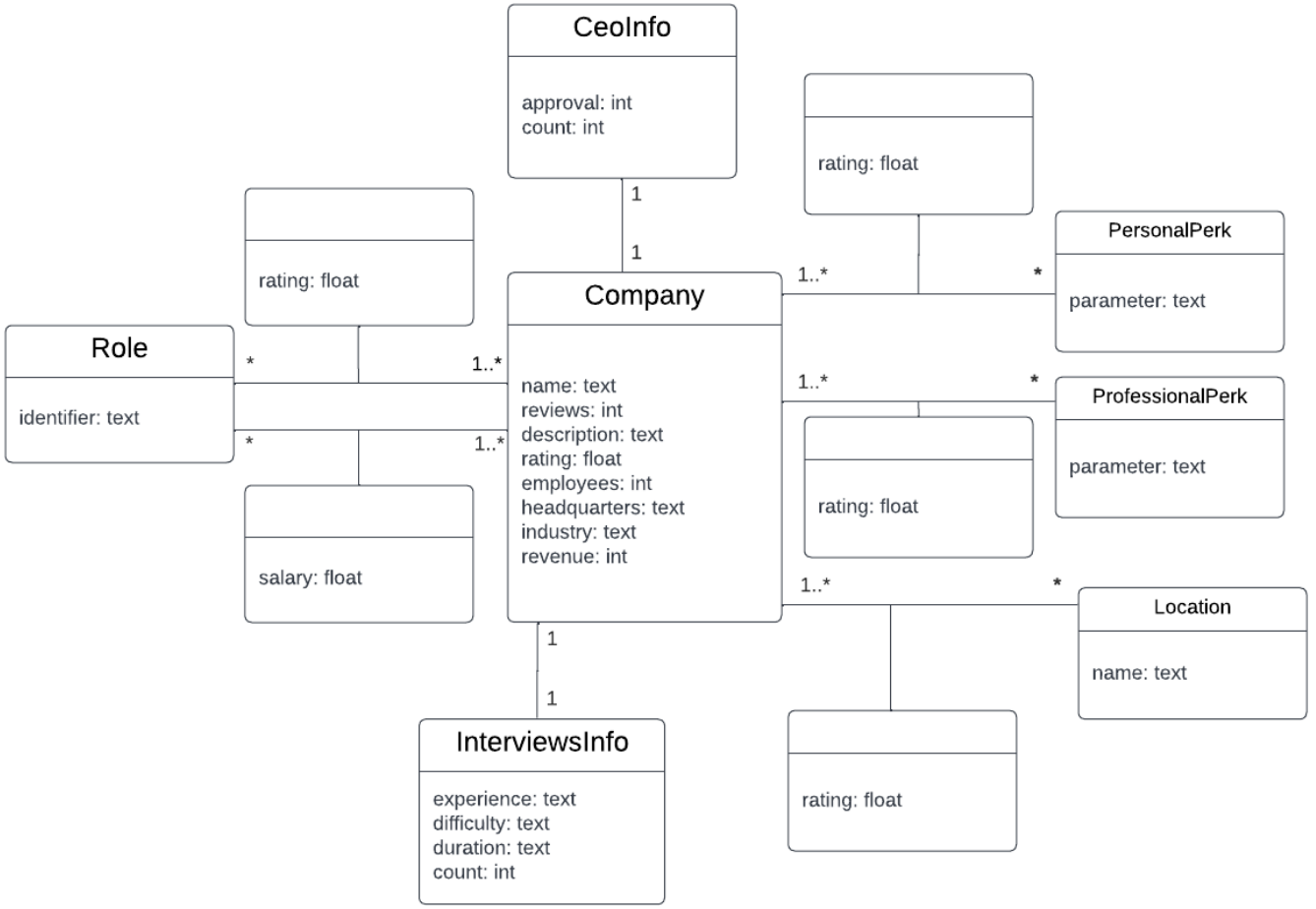
PRI, group 69

Bruno Rosendo, up201906334

João Mesquita , up201906682

Rui Alves, up201905853

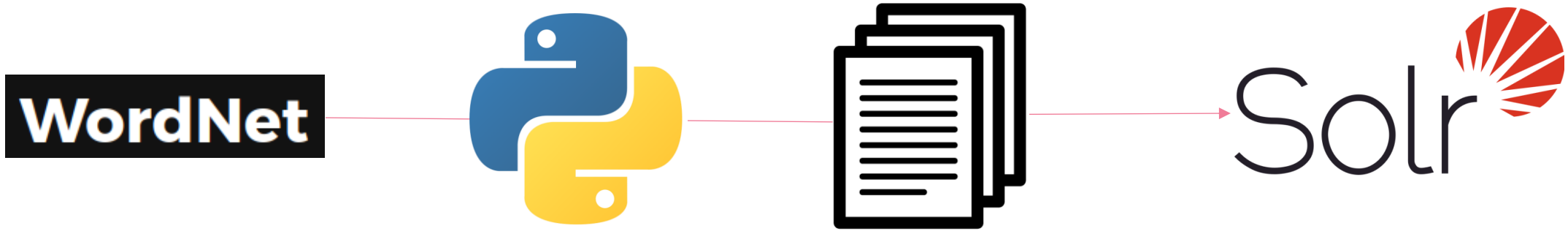
Conceptual Model of the Dataset



Search System Improvements

- Improved list of synonyms.
- Faceted search.
- Spell-checking.
- Suggestion of queries.
- Appealing user interface (frontend) to improve the experience in the system.
- Enhancements in previous iterations (data analysis, information retrieval and results evaluation).

Improved List of Synonyms



WordNet, a digital lexical database for English

Python script uses WordNet's API to generate a custom list of synonyms

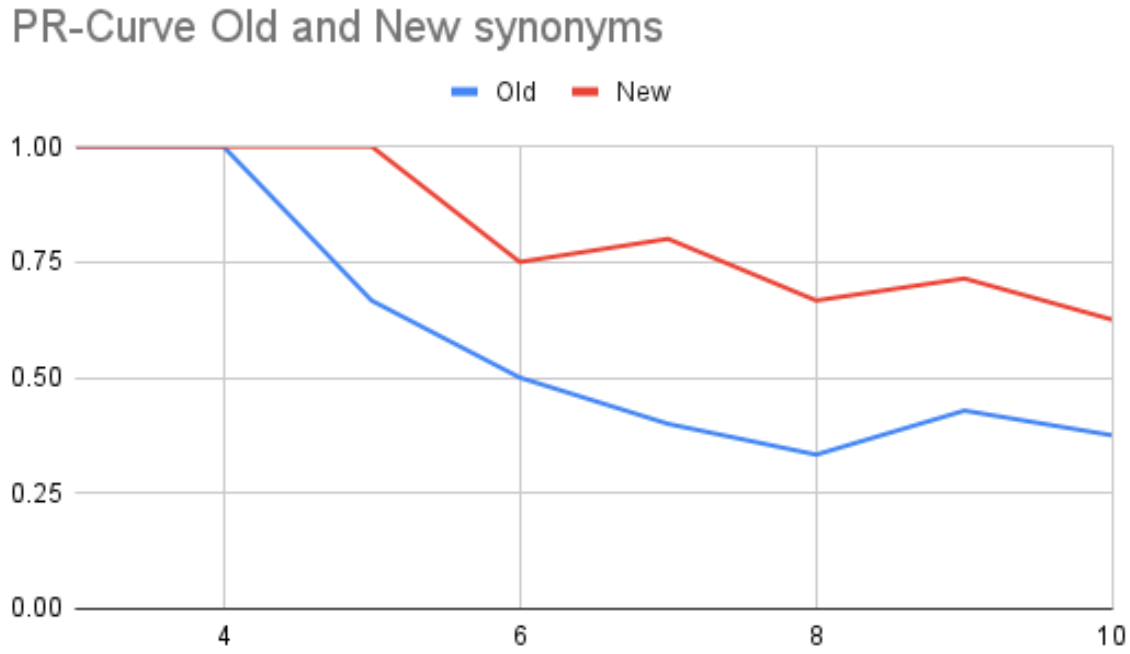
New and improved list of synonyms

List is uploaded into Solr's configuration

Improved List of Synonyms

- Let's test the new system by searching for companies involved with panic support and management (**query:** "panic").

Rank / Metric	Old Synonyms	New Synonyms
1	Y	Y
2	Y	Y
3	N	Y
4	N	N
5	N	Y
6	N	N
7	Y	Y
8	N	N
9	N	N
10	N	-
Avg. Precision	0.53	0.76
P@10	0.30	0.50



Spell-Checking

- Spell-checking allows the users to be informed of possible typos they might write in their queries.
- *IndexBasedSpellChecker* was used in the company's description, the field with the most text density.

Request:

```
/solr/reviews/spellcheck?q=tecknology
```

Response:

```
"suggestions":[
  "tecknology",{
    "numFound":1,
    "startOffset":0,
    "endOffset":10,
    "suggestion":["technology"]
  }
]
```

Query Suggestions

- Suggestions areuseful to display suggestions while the user is typing their query, identically to what happens in a Google search.
- *FreeTextLookupFactory* was used as an auto-complete that recommends words (from the *description* field) when the user starts writing them.

Request:

```
/solr/reviews/suggest?q=tech
```

Response:

```
"suggest":{"reviewsSuggester":{
  "tech":{
    "numFound":10,
    "suggestions":[{
      "term":"technology",
      "weight":10745044845232370,
      "payload":""},
    ...
    {
      "term":"techlogix",
      "weight":118948835925081,
      "payload":""}]}
  }}
```

Faceted Search

- Faceting is the arrangement of search results into categories based on indexed terms
- In this system, many categorical values benefit from this functionality: *industry*, *employees*, *revenue*, *interview experience* and *interview difficulty*.

Request:

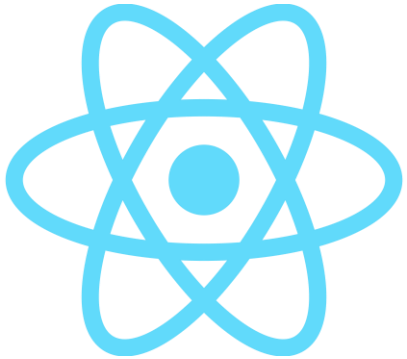
```
/solr/reviews/select?q=*:*&facet=true&
facet.field=industry&facet.limit=10
```

Response:

```
"facet_counts":{
  "facet_fields":{
    "industry":[
      "healthcare",439,
      "services",281,
      ...
      "transportation",74]},
  "facet_heatmaps":{}}
}
```

User Interface

- The final focus was on enhancing the experience for the end user so that anybody can use the system without technical background.
- The UI was developed using React and MaterialUI, with Solr as the backend.



Live Demo, let's jump into the Website!

Future Work

Enhance UI

Improve the User Interface by, for example, integrating faceted search and spell-checking.

Improve Information Retrieval

Improve the system even more by means of machine learning models that train with data from active users

Create User Accounts

Allow users to save their favorite reviews and have personalized search results.