A hybrid decision support system for employee recruitment in employment oriented social networks

Robert Aduviri, Juanjosé Tenorio, Oscar Vega, André Chávez
Pontificia Universidad Católica del Perú

Abstract

A fundamental part in the operation of organizations is the recruitment of human resources. There are currently many platforms that are used as job boards, as well as employment oriented social networks, which have become a key source of information for recruiters. However, this high volume of information makes recognizing the best candidates for a job more challenging and expensive. Nonetheless, at the same time, it turns out that there are different patterns that are associated to the match between job requirements and the profile of a potential candidate. This opens the possibility of automating a large part of this process through the creation of a decision support system based on pattern recognition techniques, such as natural language processing, both classical and based on deep learning, and recommendation systems, which are widely used in domains such as movies, music, news, books and products in general. In this paper, we present an approach based on the conjunction of such techniques, in order to help recruiters find the optimal candidate for a specific job. Using web pages of jobs, we extract the most important descriptive texts of job offers and user profiles to generate features through word embeddings that allow us to homologate them in a vector space. Thanks to these features, we can measure the skills of an applicant and then make a ranking of the most suitable candidates for each job offer.

Keywords: Job Recruitment, Decision Support Systems, Recommender Systems, Natural Language Processing, Word Embeddings

Methodology

- Laborum + LinkedIn web scraping
- Stopword removal, lowercase conversion, length limit, unicode character decoding and special characters removal
- Word2vec/Doc2vec skip-gram model with negative sampling, context window of size 10 and 20 iterations
- Kmeans clustering with Kmeans++ initialization and cosine distance evaluation through normalized euclidean distance
- t-SNE with perplexity of 30

Conclusions and Future Work

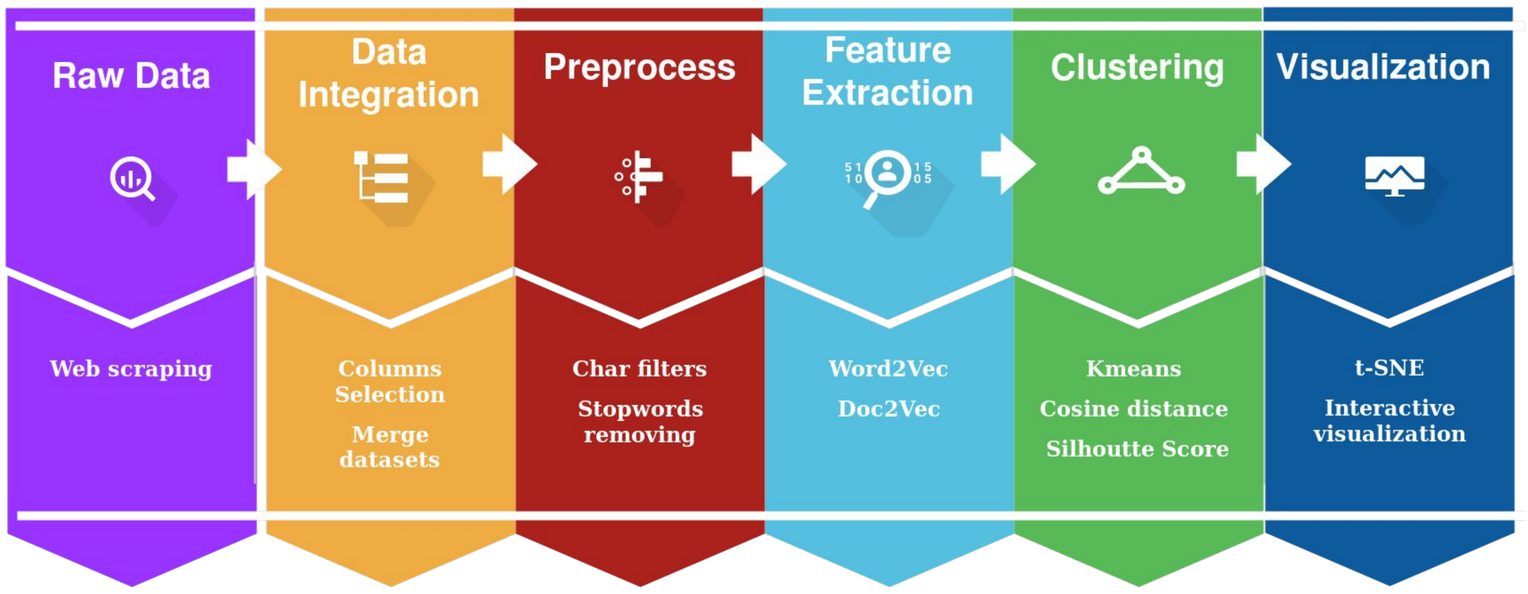
By using word embeddings considering different aspects of candidates and job offers, we were able to generate latent feature representations of both of them, which allowed us to compare their similarities in a vector space and match them.

Our next steps will be the following:

- Train our model with even more data also considering different sources for job offers, not only Laborum.
- Apply information retrieval techniques for filtering by other fields like years of experience.
- Homologate several fields of the data, like universities, companies and job position hierarchy.
- Analyze the semantics and syntactics of the text using more NLP techniques (Syntactics trees, Lemmatization, Topic Modeling etc).

We believe that by doing this, we can adapt this approach into a company's recruitment pipeline in conjunction with information retrieval tools like Elastic Search.

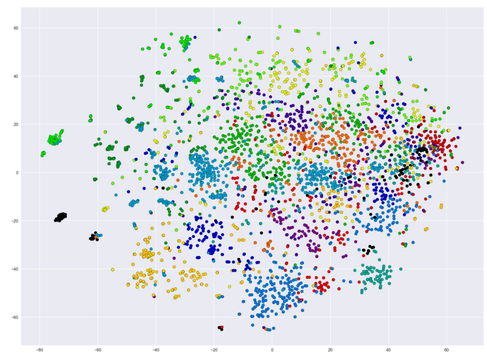
Pipeline



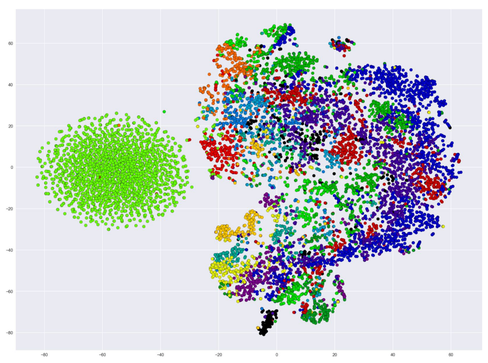
- Full name
- Headline
- Summary
- Education
- Experience
- Industry
- Languages
- Skills

- Company name
- Company type
- Job position
- Job description
- Job function
- Job requirements

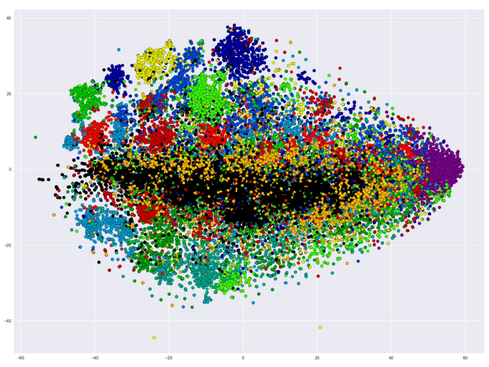
Results



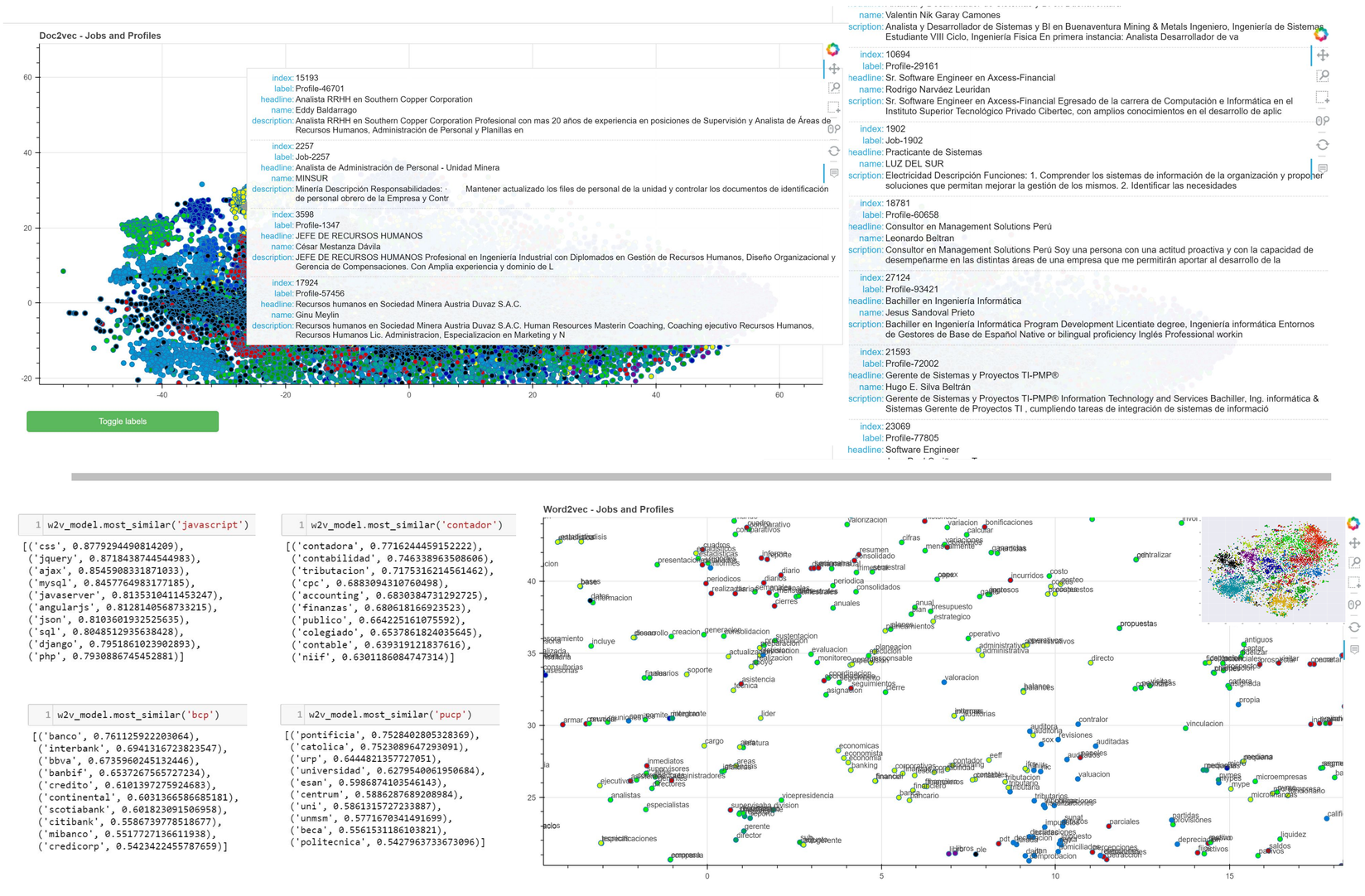
Job offers



Candidates profiles



Job offers and candidates profiles




```
1 w2v_model.most_similar('javascript')
```

```
[('css', 0.8779294490814209),
 ('jquery', 0.8718438744544983),
 ('ajax', 0.8545908331871033),
 ('mysql', 0.8457764983177185),
 ('javaserver', 0.8135310411453247),
 ('angularjs', 0.8128140568733215),
 ('json', 0.8103601932525635),
 ('sql', 0.8048512935638428),
 ('django', 0.7951861023902893),
 ('php', 0.7930886745452881)]
```

```
1 w2v_model.most_similar('contador')
```

```
[('contadora', 0.7716244459152222),
 ('contabilidad', 0.7463389635308606),
 ('tributacion', 0.7175316214561462),
 ('cpc', 0.6883094310760498),
 ('accounting', 0.6830384731292725),
 ('finanzas', 0.680618166923523),
 ('publico', 0.664225161075592),
 ('colegiado', 0.6537861824035645),
 ('contable', 0.639319121837616),
 ('niif', 0.6301186084747314)]
```

```
1 w2v_model.most_similar('bcp')
```

```
[('banco', 0.761125922203064),
 ('interbank', 0.6941316723823547),
 ('bbva', 0.6735960245132446),
 ('banbif', 0.6537267565727234),
 ('credito', 0.6101397275924683),
 ('continental', 0.6031366586685181),
 ('scotiabank', 0.601823091506958),
 ('citibank', 0.5586739778518677),
 ('mibanco', 0.5517727136611938),
 ('credicorp', 0.5423422455787659)]
```

```
1 w2v_model.most_similar('pucp')
```

```
[('pontificia', 0.7528402805328369),
 ('catolica', 0.7523089647293091),
 ('urp', 0.6444821357727051),
 ('universidad', 0.6279540061950684),
 ('esan', 0.5986874103546143),
 ('centrum', 0.5886287689208984),
 ('uni', 0.5861315727233887),
 ('unmsm', 0.5771670341491699),
 ('beca', 0.5561531186103821),
 ('politecnica', 0.5427963733673096)]
```