SKILL AND JOB RECOMMENDER

SUBMITTED BY

1	Nanjundeshwari M	61771931037
2	Buvanaa B	61771931010
3	Madhumitha M	61771931029
4	Sneka S	61771931046

BACHELOR OF ENGINEERING -ELECTRONICS AND COMMUNICATION

LITERATURE SURVEY:

[1] In this section, we describe our framework for job recommendation. We narrow down the scope and focus on recommendation of job vacancies for Information Technology (IT) professionals acting in the Brazilian market. The proposed framework is composed by three stages: data collection, data preparation and recommendation.

[2] we select a group of the nearest job offers based on the distance to that profile (job matching). In the case of TF-IDF representation, we use the cosine distance while for word embeddings, we use the relatively new Word Mover's Distance (WMD) [Kus15]. Once retrieved the top "k" job offers for the profile, we sort them in descending order based on the inverse of this distance (ranking)

[3] To perform job offers scraping, we created a list of keywords from the IT industry and used them as search terms. For each keyword, we search all the related job offers using Catho's search engine and save the retrieved results in our database; thus, the content's quality is highly related to the quality of the Catho's search engine.

[4] we retrieved data from job search sites using only IT keywords, there were still some job offers that do not correspond to this field, then, the first step in this phase is filtering out job offers that do not belong to the IT field. To achieve this, we use a dictionary of weighted IT terms to match each job offer in its document-like format.

[5] Once job offers and profiles are filtered, the second step is text preprocessing. In this task, we perform stop words removal, tokenization and lemmatization for the Portuguese language.

[6] The third step, feature representation, aims to represent these documents (job offers and profiles) as vector space models. For this purpose, we adopted two approaches: word embeddings and TF-IDF. The latter technique does not require so much effort to be implemented unlike the former.

REFERENCE:

[1] Shaha T Al-Otaibi and Mourad Ykhlef. "A survey of job recommender systems". In: International Journal of the Physical Sciences 7.29 (2012), pp. 5127–5142. issn: 19921950. doi:10.5897/IJPS12.482.

[2] M Diaby and E Viennet. "Taxonomy-based job recommender systems on Facebook and LinkedIn profiles". In: Proc. of Int. Conf. on Research Challenges in Information Science (2014), pp. 1–6. issn: 21511357. doi:10.1109/RCIS.2014.6861048.

[3] T Mikolov et al. "Efficient estimation of word representations in vector space". In: arXiv preprintarXiv:1301.3781 (2013)

[4] T Mikolov et al. "Distributed Representations of Words and Phrases and Their Compositionality".In: Proc. of the 26th Int. Conf. on Neural Information Processing Systems - Volume 2. NIPS'13.Lake Tahoe, Nevada, 2013, pp. 3111–3119. url:http://dl.acm.org/citation.cfm?id=2999792.2999959.

[5] M Kusner et al. "From word embeddings to document distances". In: Proc. of the 32nd Int. Conf. on Machine Learning, ICML'15. 2015, pp. 957–966.

[6] G Salton and C Buckley. "Term-weighting approaches in automatic text retrieval". In: InformationProcessing and Management 24.5 (1988), pp. 513–523. issn: 0306-4573. doi:https://doi.org/10.1016 /0306 - 4573(88) 90021 - 0.url:http: / / www . sciencedirect . com / science / article / pii /0306457388900210.