VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

TOPIC

NAMED ENTITY EXTRACTION FOR RESUME

PROJECT REVIEW-III

CODE LINK:

https://colab.research.google.com/drive/1VLV5iXrkfkgRRrdfH9RDyk2BBdM-upZP?usp=sharing

TEAM MEMBERS:

19BCE0141 - G.GOWTHAM
19BCE0733 - ABHINAV MALLEM
19BCE2155 - CH.VISHNU VARDHAN
19BCT0252 - SHAIK AHMMED BASH

SUBMITTED TO : DR. RAJESHKANNAN

1.ABSTRACT:

Screening resumes out of bulk is a challenging task and recruiters or hiring manager wastes lot of their valuable time by searching through each and every resume. Often resumes are populated with irrelevant and unnecessary information. Therefore, the process of parsing thousands of resumes manually consumes lot of time and energy there by it makes the hiring process expensive. In traditional hiring, resume screening is a manual process which consumes a lot of time and energy. In this paper the process of screening resumes is automated by using advanced Natural Language Processing which is a field in deep Learning .Our model helps the recruiters in screening the resumes based on job description with in no time. It makes the hiring process easy and efficient by extracting the required entities automatically by using Spacy NER model from the resumes and then generates a graph displaying the score of each and every resume. Based on the scores recruiter can choose the required candidates without rummaging through piles of resumes from unqualified candidates.

2. INTRODUCTION:

Globally, companies receive resumes in large numbers that require screening. Resumes carry semi-structured text, which is difficult to parse. The difficulty arises from differences in structures, styles, formats, order, and types of information that the resumes incorporate. It usually consists of various sections that reflect the candidate’ s competency. Accurate parsing of these resume sections without manual intervention is a dire need.

A widely used technique for recognizing entities is named entity recognition (NER). NER refers to identifying all the occurrences belonging to a specific type of entity in the text. NER tasks require a large amount of annotated data that could be extremely cumbersome to produce. There exists a need for auto-annotated data, which can provide good accuracy.

One of the most common approaches used for NER is a reference from a list . This approach usually leads to better performance and depends on the entire list and, therefore, defaults. We can also perform NER tasks using various deep learning models.

In NER, the combination of word embedding , convolutional neural networks (CNN) , bidirectional long-short term memory (Bi-LSTM) and conditional random fields (CRF) is the most preferred combination . In our model, we have used the combination without the CRF layer. CRFs perform better with structured data. Since resumes have semi-structured data, we decided to skip the CRF layer .

Much research has been carried out in the field of resume parsing in recent times. Jiang et al. have used statistical and rule-based algorithms for extracting relevant information. However, this approach fails to generalize for resumes in English. Farkas et al. have devised an application where the user uploads his resume from which details are automatically extracted, and an application form is subsequently filled. The user is then allowed to edit the form, if required, and submit it. This method relies on high recall so that even if the information fetched is not precise, the user can edit the automatically extracted information in the form and submit it. A CRF-based resume miner to extract information provides a method for ranking applicants for a given job profile . These methods give low precision and low recall for institute and degree names .

3.PROBLEM STATEMENT

The problem is that the present system is not much flexible and efficient and time saving as it is not guaranteed that only eligible candidates will upload their resumes for a particular company. So out of bulk of relevant and irrelevant resumes the recruiter has to srcutinize them. Our system saves time for the recruiters to scrutinize the resumes based on job description by automatically ranking the resume .This not only helps the recruiters for srcutinization but also it helps the candidate will be get satisfied because he will get job in that company which really appreciates candidates skill and ability. Learning to rank refers to Deep learning techniques for training the model in a ranking task. Learning to rank is useful for many applications in Information Retrieval, Natural Language Processing, and Data Mining. Intensive studies have been conducted on the problem and significant progress has been made. This short paper gives an introduction to learning to rank, and it specifically explains the fundamental problems, existing approaches, and future work of learning to rank.

The major objective of our system is to automate the hiring process in order to reduce the cost of hiring and to make the hiring process more efficient.

Candidates, who has been hired:

Candidates who are searching for jobs after been graduated. Out of those, major number of candidates are so much desperate that they are ready to work on any post irrelevant to their skill set and ability. Where our algorithm will work in such a way that with the help of the previous result and previous ranking constraints, it will try to optimize the current result, which we called it deep Learning. This will make sure that the relevant candidate is been hired for that particular vacancy. You can say best possible candidate.

Client Company, who is hiring the candidates:

Like I am the owner of a particular organization, obviously my aim would be to create such a team which is the best team in the world. It is like, if there is a vacancy of a java developer in my organization. So, I won't prefer to hire a python developer and then make him learn Java. That will be pretty useless and time consuming for both that candidate and for the organization too. Where our system help the organization to make out the best possible candidates list according to their given constraints and requirement for that particular vacancy. So there would be no regrets for both the entities, client company and that hired candidate. Hence satisfaction will be achieved.

Named entity recognition using deep learning

Recently deep learning-based methods have also been explored for NER. A pre-trained word-embedding is used as an input to a neural network model and character-level features . A comparison of a **Bi-LSTM cum CRF model** with a **transition-based chunking model** with shift-reduce parsers is made for NER, concluding that the former gave a better performance . Stacking of recurrent neural network layers for a biomedical sequence was employed for classification purposes . **Bi-LSTM, CNN, and CRF layers** have been incorporated into the neural network model for unstructured data.

Yu et al have developed a cascaded information extraction (IE) framework. A **CV** is segmented into blocks with labels for different information types in the first pass using an **HMM Model**. Then in the second go, **detailed information**, like **Name**, is extracted from individual blocks instead of searching in the entire resume for it using a hybrid model consisting of **HMM and SVM**.

Maheshwary and Misra have proposed a Siamese adaptation of CNN to accomplish the task of matching resumes with a particular job opening. The model consists of a pair of identical CNN, which they propose gives them a measure of semantic relatedness of words in the resume in a controllable manner with low computational costs. They test their model against simple models like **Bag of Words, and TF-IDF** compared to which their model performed better.

Deep learning model approaches have shown a significant performance and accuracy improvement for named entity recognition task. Moreover, it can be generalized over a wide variety of data, unlike the rule-based approaches.

Chifu et al have created skills, and web crawled resumes are checked for **POS** patterns after text preprocessing using the Stanford **NLP framework**. If words are not present in the skill ontology, new skills are updated for further skill detection using algorithms trained for specific lexical patterns. Wikipedia is the primary source of ontology, and the whole system is highly dependent on the same.

Ghufran et al leverage the fact that an individual's resume contents are available on Wikipedia for automatic annotation without **POS tags**. **N-grams** are constructed from keywords in a resume and then queried to Wikipedia. Returned results are in the form of an interpretation graph, processed for disambiguation and cross-language references. Dependency on Wikipedia for information is very high, and the dataset of resumes is also limited.

Zhang et al have proposed a technique for **parsing** the **semi-structured data** of the Chinese resumes. The system consists of the following key components, firstly the set of classes used for classification of the entities in the resume, secondly the algorithm used for identification of those entities, and lastly, the system design. The entities are divided into two major subcategories, that is, simple items like name, and date of birth, and miscellaneous items like the learning experience, skills, etc., which exhaustively cover all entities. A total of 5000 resumes is used, and a system comprising of SVM, regular expression, and **vector space model**-based classifier base has been implemented and he also have proposed an analytic system for the mining and visualizing the semi-structured data in resumes. The semantic information is first extracted after which visualization, in the form of understanding the career progression of an individual, assessing the social relationships, and holding an overall view of the

resume is done to represent this collected information. Prospects include incorporating visualization of geographical dimensions.

Darshan has used Perl-based regular expressions to convert semi-structured into an ontological structure. This semantic information is further represented in XML format for information extraction from resumes.The limited literature on resume parsing has not demonstrated **very high accuracy** in identifying institutions and degrees.

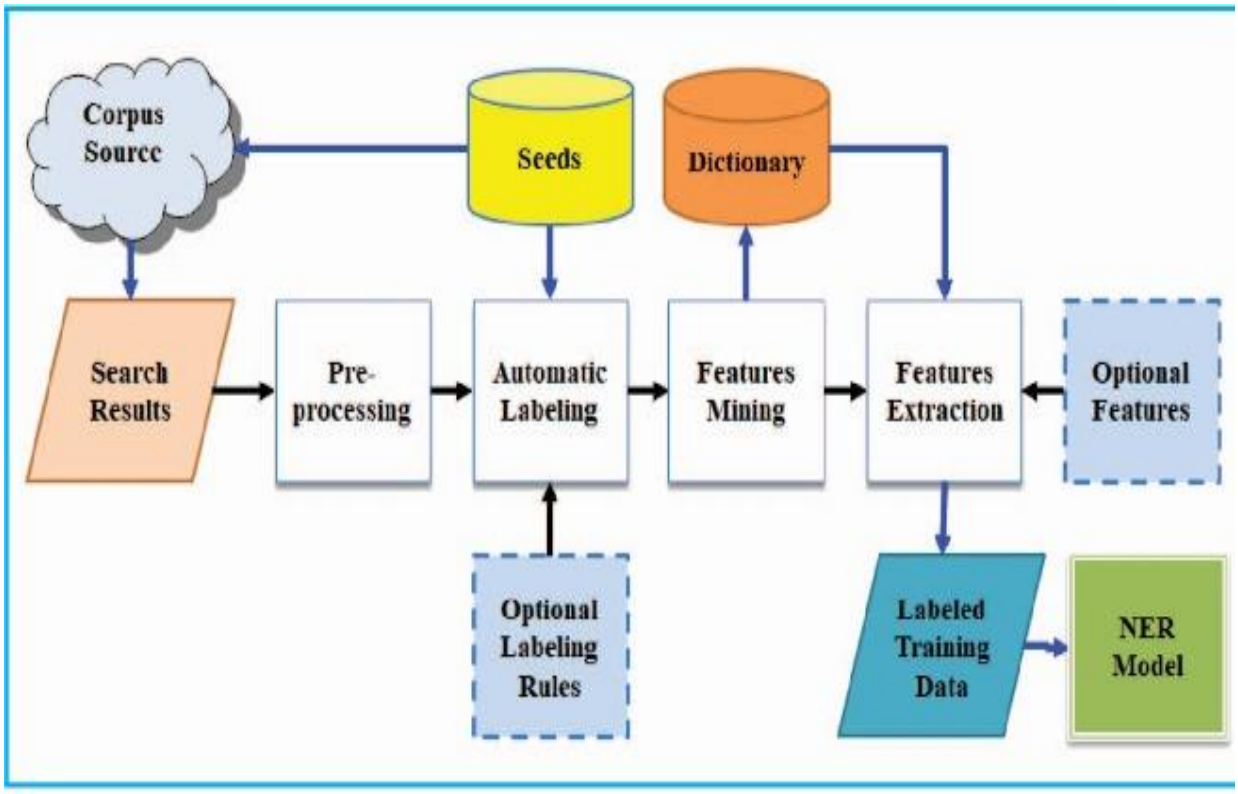
In our work, we focus on accurate identification of academic degrees and institute names in a resume’ s education section. The proposed method eases the recruiter’ s search for candidates from specific institutions or academic qualifications. It further helps in the analysis regarding recruitment trends specific to colleges, compensations, and industry exposures provided to the candidates .

The main contributions of this project are summarized as follows:

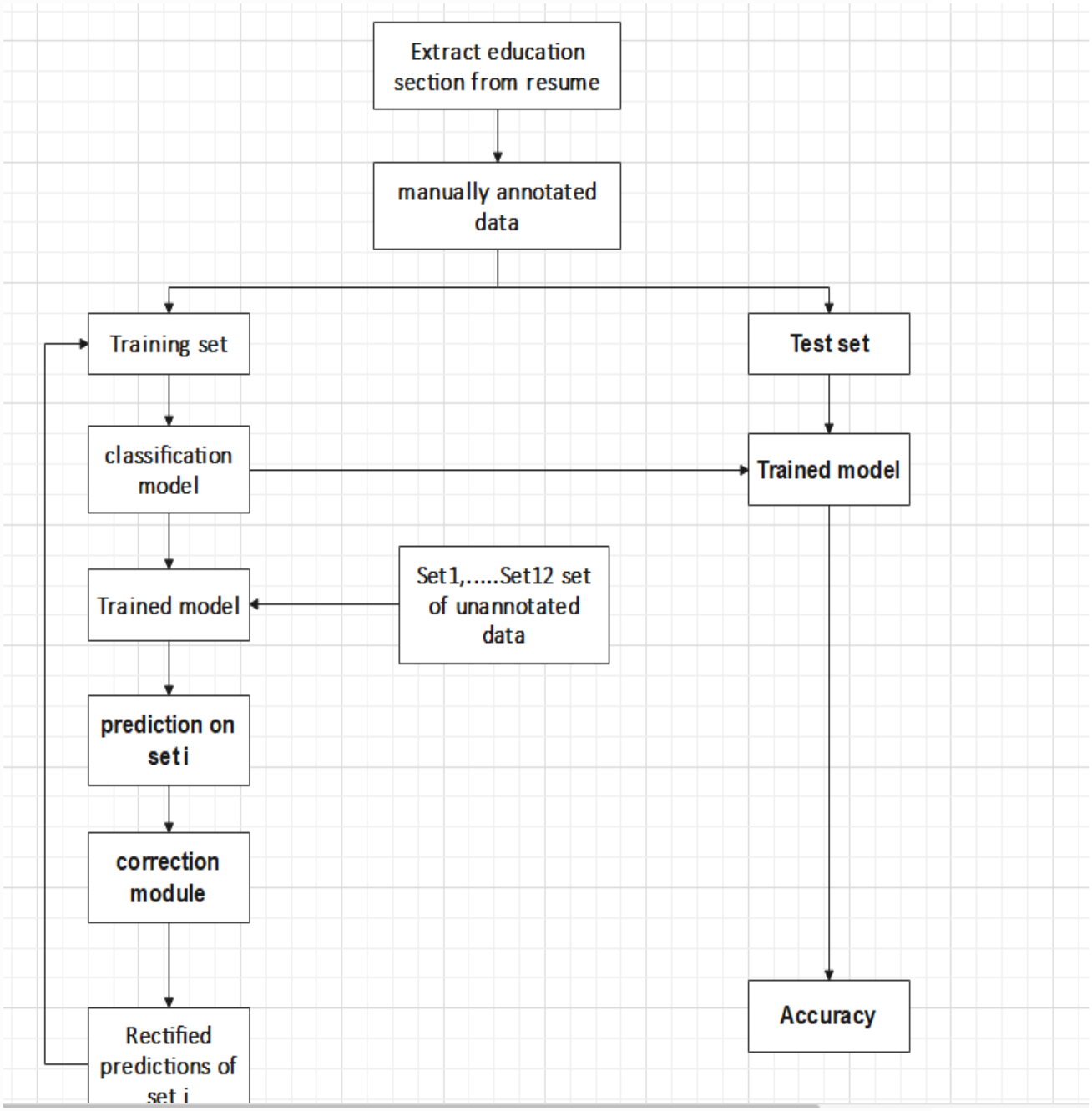
- We demonstrated the use of a modified semi-supervised technique for parsing institute and degree names. Instead of following the traditional semi-supervised approach, we introduced a correction module to rectify the predictions. We added these corrected predictions back to the original seed set, thereby increasing its size. On retraining, this procedure results in improved accuracy, precision, and recall in comparison with the previously trained model.

- We achieved high performance for recognizing degrees and institutes in a resume without large annotated data.

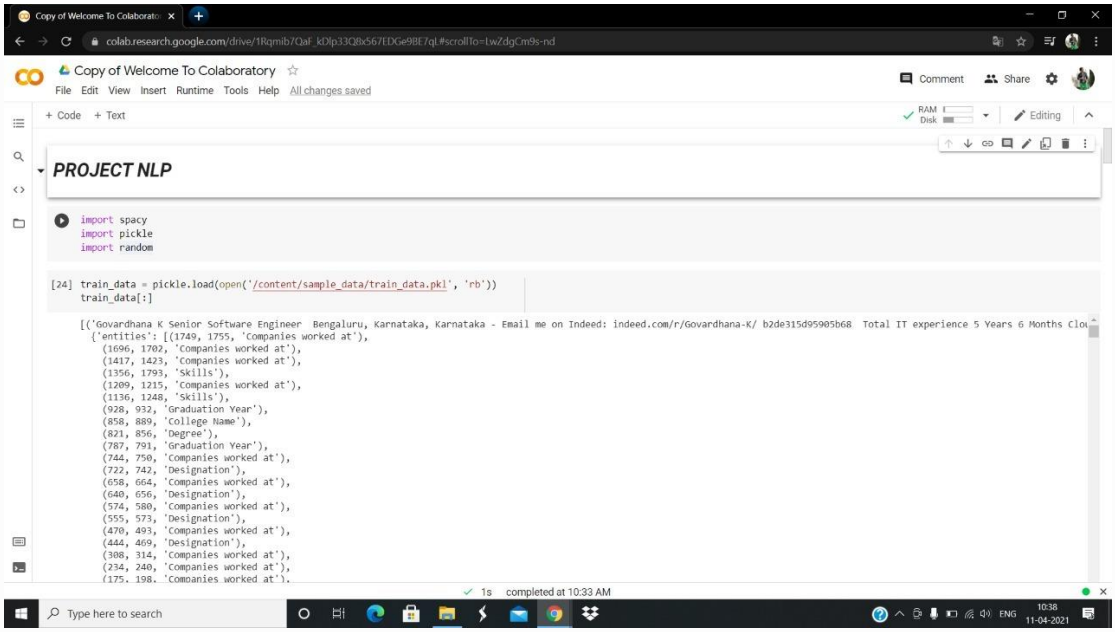
3.1ARCHITECTURE DAIGRAM

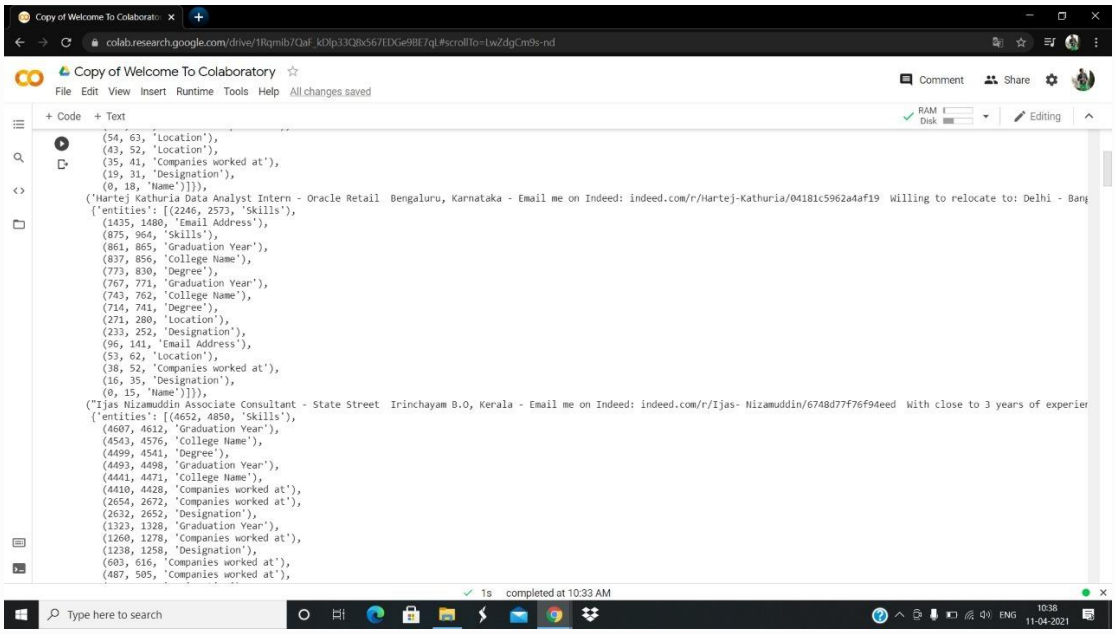


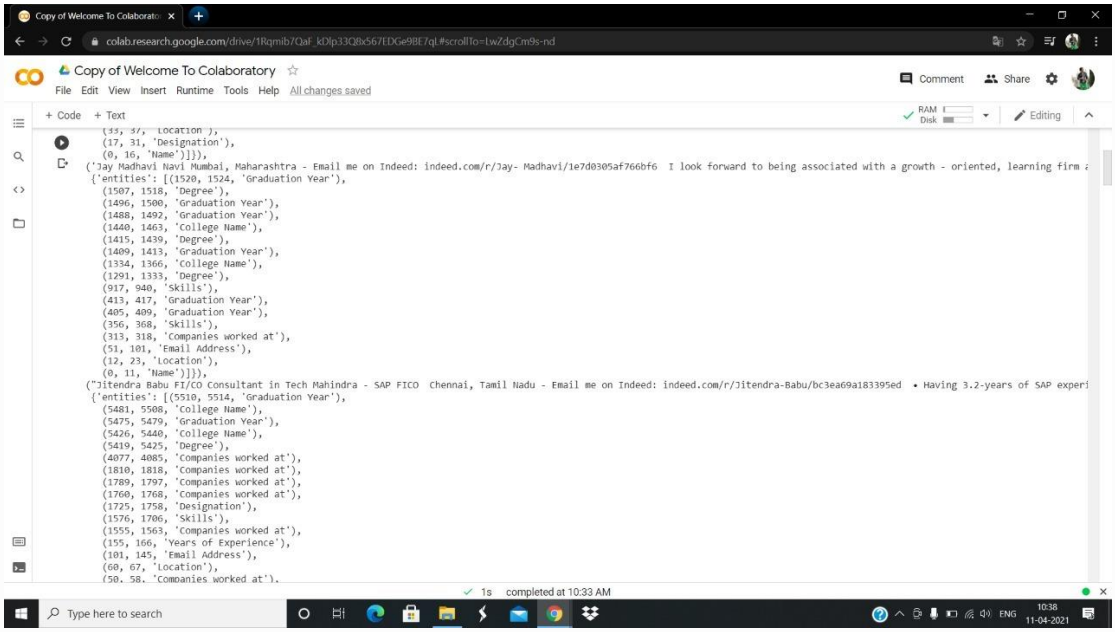
3.2 FLOWCHART



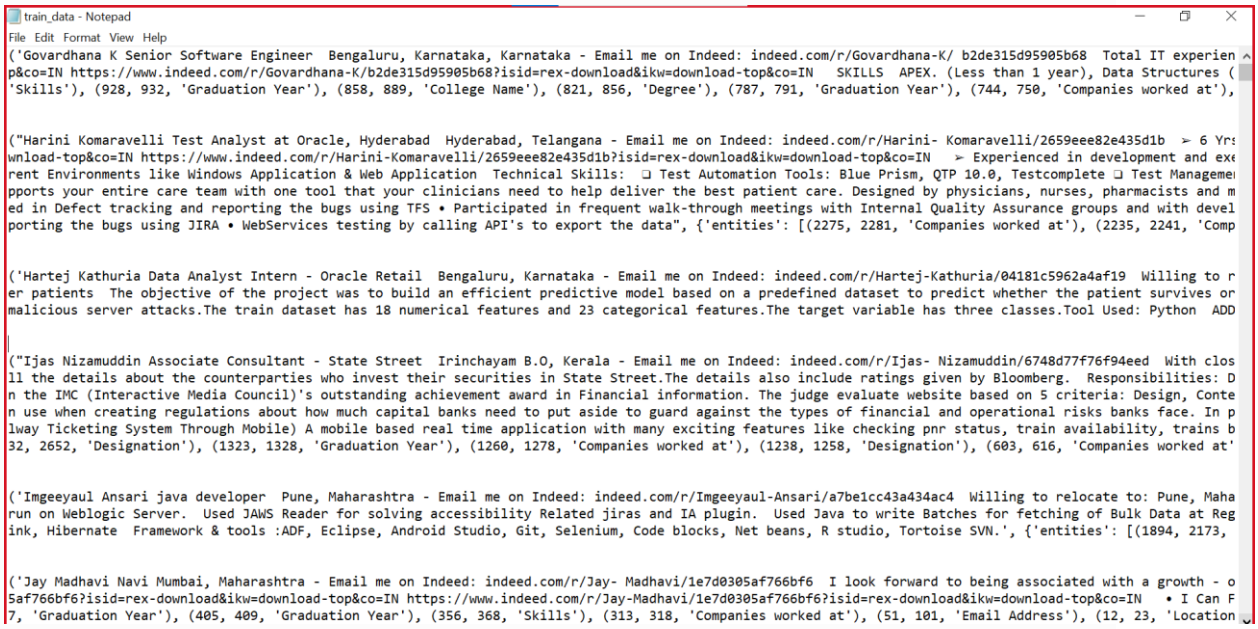
3.3 Pseudocode

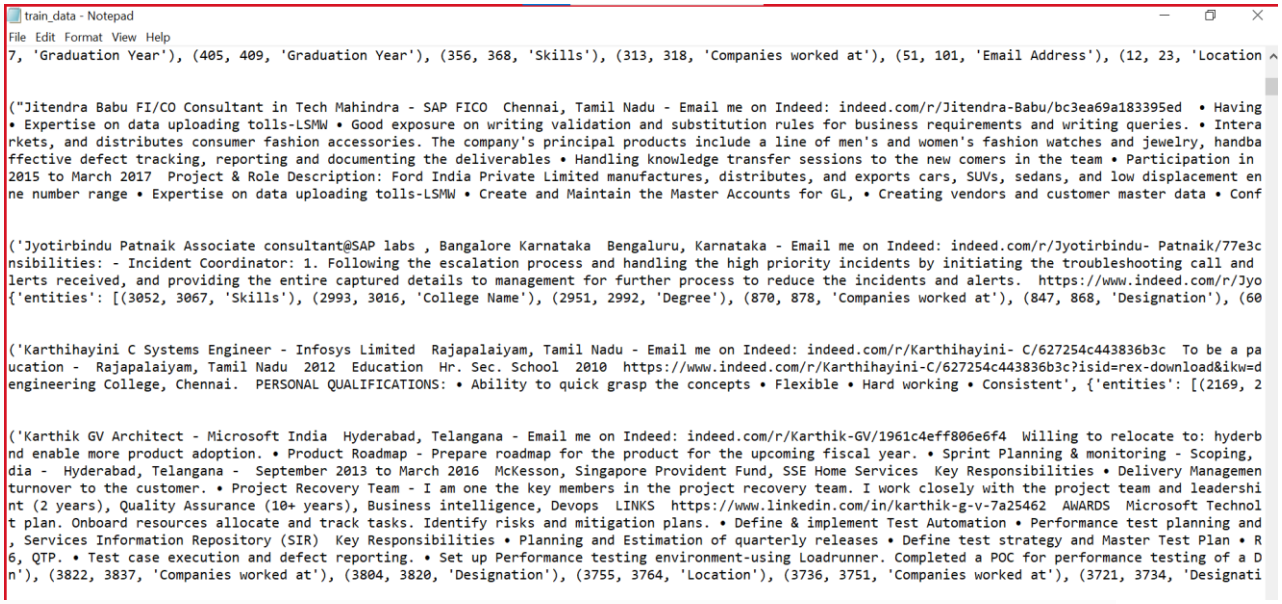





Copy of Welcome To Colaboratory
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Editing
(17, 31, 'Designation'),
(0, 16, 'Name')]}),
('Jay Madhavi Navi Mumbai, Maharashtra - Email me on Indeed: indeed.com/r/Jay- Madhavi/1e7d0305af766bf6 I look forward to being associated with a growth - oriented, learning firm
{'entities': [(1520, 1524, 'Graduation Year'),
(1507, 1518, 'Degree'),
(1496, 1500, 'Graduation Year'),
(1488, 1492, 'Graduation Year'),
(1440, 1463, 'College Name'),
(1415, 1439, 'Degree'),
(1409, 1413, 'Graduation Year'),
(1334, 1366, 'College Name'),
(1291, 1333, 'Degree'),
(917, 940, 'Skills'),
(413, 417, 'Graduation Year'),
(405, 409, 'Graduation Year'),
(356, 368, 'Skills'),
(313, 318, 'Companies worked at'),
(51, 101, 'Email Address'),
(12, 23, 'Location'),
(0, 11, 'Name')]}),
("Jitendra Babu FI/CO Consultant in Tech Mahindra - SAP FICO Chennai, Tamil Nadu - Email me on Indeed: indeed.com/r/Jitendra-Babu/bc3ea69a183395ed • Having 3.2-years of SAP experi
{'entities': [(5510, 5514, 'Graduation Year'),
(5481, 5508, 'College Name'),
(5475, 5479, 'Graduation Year'),
(5426, 5440, 'College Name'),
(5419, 5425, 'Degree'),
(4077, 4085, 'Companies worked at'),
(1810, 1818, 'Companies worked at'),
(1789, 1797, 'Companies worked at'),
(1760, 1768, 'Companies worked at'),
(1725, 1758, 'Designation'),
(1576, 1706, 'Skills'),
(1555, 1563, 'Companies worked at'),
(155, 166, 'Years of Experience'),
(101, 145, 'Email Address'),
(60, 67, 'Location'),
1s completed at 10:33 AM
Type here to search
ENG
10:38
11-04-2021

Data set:


train_data - Notepad
File Edit Format View Help
('Govardhana K Senior Software Engineer Bengaluru, Karnataka, Karnataka - Email me on Indeed: indeed.com/r/Govardhana-K/ b2de315d95905b68 Total IT experien
p&co=IN https://www.indeed.com/r/Govardhana-K/b2de315d95905b68?isid=rex-download&ikw=download-top&co=IN SKILLS APEX. (Less than 1 year), Data Structures (
'Skills'), (928, 932, 'Graduation Year'), (858, 889, 'College Name'), (821, 856, 'Degree'), (787, 791, 'Graduation Year'), (744, 750, 'Companies worked at'),
("Harini Komaravelli Test Analyst at Oracle, Hyderabad Hyderabad, Telangana - Email me on Indeed: indeed.com/r/Harini- Komaravelli/2659eee82e435d1b ➢ 6 Yrs
wnload-top&co=IN https://www.indeed.com/r/Harini-Komaravelli/2659eee82e435d1b?isid=rex-download&ikw=download-top&co=IN ➢ Experienced in development and exe
rent Environments like Windows Application & Web Application Technical Skills: ❑ Test Automation Tools: Blue Prism, QTP 10.0, Testcomplete ❑ Test Manageme
pports your entire care team with one tool that your clinicians need to help deliver the best patient care. Designed by physicians, nurses, pharmacists and m
ed in Defect tracking and reporting the bugs using TFS • Participated in frequent walk-through meetings with Internal Quality Assurance groups and with devel
porting the bugs using JIRA • WebServices testing by calling API's to export the data", {'entities': [(2275, 2281, 'Companies worked at'), (2235, 2241, 'Comp
('Hartej Kathuria Data Analyst Intern - Oracle Retail Bengaluru, Karnataka - Email me on Indeed: indeed.com/r/Hartej-Kathuria/04181c5962a4af19 Willing to r
er patients The objective of the project was to build an efficient predictive model based on a predefined dataset to predict whether the patient survives or
malicious server attacks.The train dataset has 18 numerical features and 23 categorical features.The target variable has three classes.Tool Used: Python ADD
("Ijas Nizamuddin Associate Consultant - State Street Irinchayam B.O, Kerala - Email me on Indeed: indeed.com/r/Ijas- Nizamuddin/6748d77f76f94eed With clos
ll the details about the counterparties who invest their securities in State Street.The details also include ratings given by Bloomberg. Responsibilities: D
n the IMC (Interactive Media Council)'s outstanding achievement award in Financial information. The judge evaluate website based on 5 criteria: Design, Conte
n use when creating regulations about how much capital banks need to put aside to guard against the types of financial and operational risks banks face. In p
lway Ticketing System Through Mobile) A mobile based real time application with many exciting features like checking pnr status, train availability, trains b
32, 2652, 'Designation'), (1323, 1328, 'Graduation Year'), (1260, 1278, 'Companies worked at'), (1238, 1258, 'Designation'), (603, 616, 'Companies worked at'
('Imgeeyaul Ansari java developer Pune, Maharashtra - Email me on Indeed: indeed.com/r/Imgeeyaul-Ansari/a7be1cc43a434ac4 Willing to relocate to: Pune, Maha
run on Weblogic Server. Used JAWS Reader for solving accessibility Related jiras and IA plugin. Used Java to write Batches for fetching of Bulk Data at Reg
ink, Hibernate Framework & tools :ADF, Eclipse, Android Studio, Git, Selenium, Code blocks, Net beans, R studio, Tortoise SVN.', {'entities': [(1894, 2173,
('Jay Madhavi Navi Mumbai, Maharashtra - Email me on Indeed: indeed.com/r/Jay- Madhavi/1e7d0305af766bf6 I look forward to being associated with a growth - o
5af766bf6?isid=rex-download&ikw=download-top&co=IN https://www.indeed.com/r/Jay-Madhavi/1e7d0305af766bf6?isid=rex-download&ikw=download-top&co=IN • I Can F
7, 'Graduation Year'), (405, 409, 'Graduation Year'), (356, 368, 'Skills'), (313, 318, 'Companies worked at'), (51, 101, 'Email Address'), (12, 23, 'Location



train_data - Notepad

File Edit Format View Help

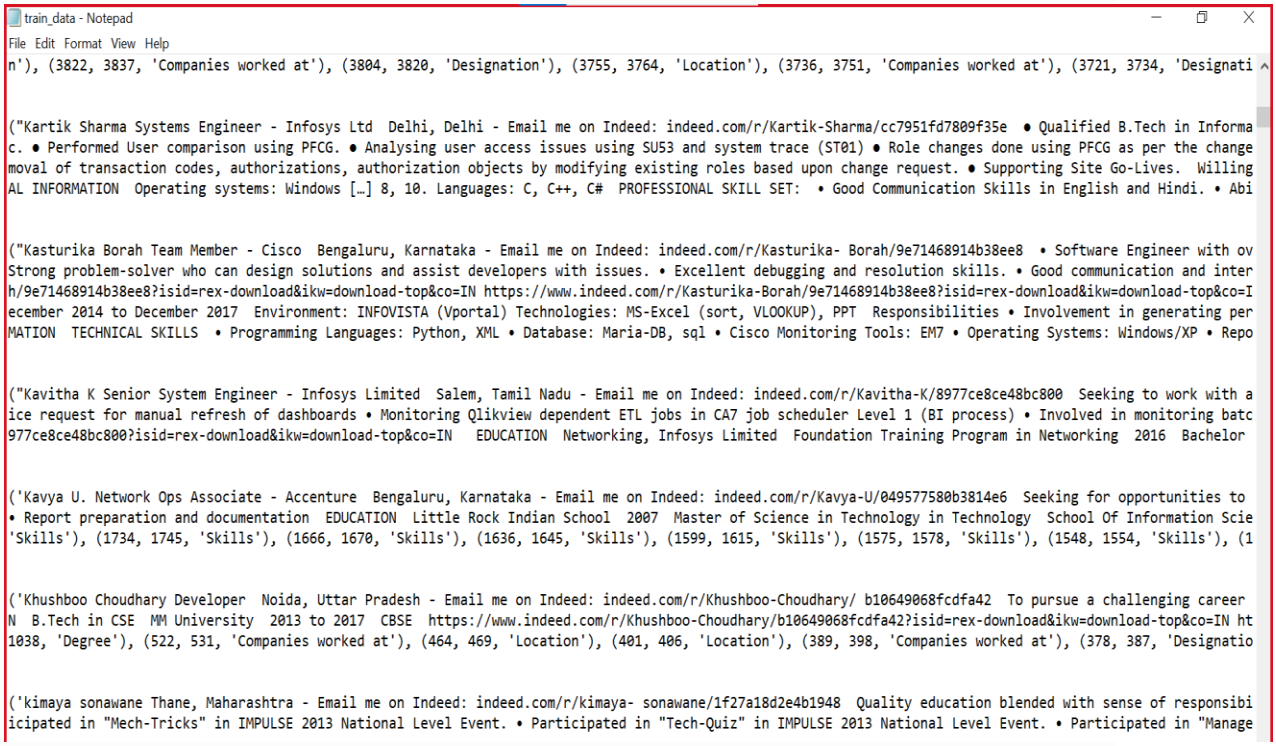
```
7, 'Graduation Year'), (405, 409, 'Graduation Year'), (356, 368, 'Skills'), (313, 318, 'Companies worked at'), (51, 101, 'Email Address'), (12, 23, 'Location


("Jitendra Babu FI/CO Consultant in Tech Mahindra - SAP FICO  Chennai, Tamil Nadu - Email me on Indeed: indeed.com/r/Jitendra-Babu/bc3ea69a183395ed  • Having
• Expertise on data uploading tolls-LSMW • Good exposure on writing validation and substitution rules for business requirements and writing queries. • Intera
rkets, and distributes consumer fashion accessories. The company's principal products include a line of men's and women's fashion watches and jewelry, handba
ffective defect tracking, reporting and documenting the deliverables • Handling knowledge transfer sessions to the new comers in the team • Participation in
2015 to March 2017  Project & Role Description: Ford India Private Limited manufactures, distributes, and exports cars, SUVs, sedans, and low displacement en
ne number range • Expertise on data uploading tolls-LSMW • Create and Maintain the Master Accounts for GL, • Creating vendors and customer master data • Conf


('Jyotirbindu Patnaik Associate consultant@SAP labs , Bangalore Karnataka  Bengaluru, Karnataka - Email me on Indeed: indeed.com/r/Jyotirbindu- Patnaik/77e3c
nsibilities: - Incident Coordinator: 1. Following the escalation process and handling the high priority incidents by initiating the troubleshooting call and
lerts received, and providing the entire captured details to management for further process to reduce the incidents and alerts.  https://www.indeed.com/r/Jyo
{'entities': [(3052, 3067, 'Skills'), (2993, 3016, 'College Name'), (2951, 2992, 'Degree'), (870, 878, 'Companies worked at'), (847, 868, 'Designation'), (60


('Karthihayini C Systems Engineer - Infosys Limited  Rajapalaiyam, Tamil Nadu - Email me on Indeed: indeed.com/r/Karthihayini- C/627254c443836b3c  To be a pa
ucation -  Rajapalaiyam, Tamil Nadu  2012  Education  Hr. Sec. School  2010  https://www.indeed.com/r/Karthihayini-C/627254c443836b3c?isid=rex-download&ikw=d
engineering College, Chennai.  PERSONAL QUALIFICATIONS: • Ability to quick grasp the concepts • Flexible • Hard working • Consistent', {'entities': [(2169, 2


('Karthik GV Architect - Microsoft India  Hyderabad, Telangana - Email me on Indeed: indeed.com/r/Karthik-GV/1961c4eff806e6f4  Willing to relocate to: hyderb
nd enable more product adoption. • Product Roadmap - Prepare roadmap for the product for the upcoming fiscal year. • Sprint Planning & monitoring - Scoping,
dia -  Hyderabad, Telangana -  September 2013 to March 2016  McKesson, Singapore Provident Fund, SSE Home Services  Key Responsibilities • Delivery Managemen
turnover to the customer. • Project Recovery Team - I am one the key members in the project recovery team. I work closely with the project team and leadershi
nt (2 years), Quality Assurance (10+ years), Business intelligence, Devops  LINKS  https://www.linkedin.com/in/karthik-g-v-7a25462  AWARDS  Microsoft Technol
t plan. Onboard resources allocate and track tasks. Identify risks and mitigation plans. • Define & implement Test Automation • Performance test planning and
, Services Information Repository (SIR)  Key Responsibilities • Planning and Estimation of quarterly releases • Define test strategy and Master Test Plan • R
6, QTP. • Test case execution and defect reporting. • Set up Performance testing environment-using Loadrunner. Completed a POC for performance testing of a D
n'), (3822, 3837, 'Companies worked at'), (3804, 3820, 'Designation'), (3755, 3764, 'Location'), (3736, 3751, 'Companies worked at'), (3721, 3734, 'Designati
```



train_data - Notepad

File Edit Format View Help

```
n'), (3822, 3837, 'Companies worked at'), (3804, 3820, 'Designation'), (3755, 3764, 'Location'), (3736, 3751, 'Companies worked at'), (3721, 3734, 'Designati


("Kartik Sharma Systems Engineer - Infosys Ltd  Delhi, Delhi - Email me on Indeed: indeed.com/r/Kartik-Sharma/cc7951fd7809f35e  ● Qualified B.Tech in Informa
c. ● Performed User comparison using PFCG. ● Analysing user access issues using SU53 and system trace (ST01) ● Role changes done using PFCG as per the change
moval of transaction codes, authorizations, authorization objects by modifying existing roles based upon change request. ● Supporting Site Go-Lives.  Willing
AL INFORMATION  Operating systems: Windows […] 8, 10. Languages: C, C++, C#  PROFESSIONAL SKILL SET:  • Good Communication Skills in English and Hindi. • Abi


("Kasturika Borah Team Member - Cisco  Bengaluru, Karnataka - Email me on Indeed: indeed.com/r/Kasturika- Borah/9e71468914b38ee8  • Software Engineer with ov
Strong problem-solver who can design solutions and assist developers with issues. • Excellent debugging and resolution skills. • Good communication and inter
h/9e71468914b38ee8?isid=rex-download&ikw=download-top&co=IN https://www.indeed.com/r/Kasturika-Borah/9e71468914b38ee8?isid=rex-download&ikw=download-top&co=I
ecember 2014 to December 2017  Environment: INFOVISTA (Vportal) Technologies: MS-Excel (sort, VLOOKUP), PPT  Responsibilities • Involvement in generating per
MATION  TECHNICAL SKILLS  • Programming Languages: Python, XML • Database: Maria-DB, sql • Cisco Monitoring Tools: EM7 • Operating Systems: Windows/XP • Repo


("Kavitha K Senior System Engineer - Infosys Limited  Salem, Tamil Nadu - Email me on Indeed: indeed.com/r/Kavitha-K/8977ce8ce48bc800  Seeking to work with a
ice request for manual refresh of dashboards • Monitoring Qlikview dependent ETL jobs in CA7 job scheduler Level 1 (BI process) • Involved in monitoring batc
977ce8ce48bc800?isid=rex-download&ikw=download-top&co=IN  EDUCATION  Networking, Infosys Limited  Foundation Training Program in Networking  2016  Bachelor


('Kavya U. Network Ops Associate - Accenture  Bengaluru, Karnataka - Email me on Indeed: indeed.com/r/Kavya-U/049577580b3814e6  Seeking for opportunities to
• Report preparation and documentation  EDUCATION  Little Rock Indian School  2007  Master of Science in Technology in Technology  School Of Information Scie
'Skills'), (1734, 1745, 'Skills'), (1666, 1670, 'Skills'), (1636, 1645, 'Skills'), (1599, 1615, 'Skills'), (1575, 1578, 'Skills'), (1548, 1554, 'Skills'), (1


('Khushboo Choudhary Developer  Noida, Uttar Pradesh - Email me on Indeed: indeed.com/r/Khushboo-Choudhary/ b10649068fcdfa42  To pursue a challenging career
N  B.Tech in CSE  MM University  2013 to 2017  CBSE  https://www.indeed.com/r/Khushboo-Choudhary/b10649068fcdfa42?isid=rex-download&ikw=download-top&co=IN ht
1038, 'Degree'), (522, 531, 'Companies worked at'), (464, 469, 'Location'), (401, 406, 'Location'), (389, 398, 'Companies worked at'), (378, 387, 'Designatio


('kimaya sonawane Thane, Maharashtra - Email me on Indeed: indeed.com/r/kimaya- sonawane/1f27a18d2e4b1948  Quality education blended with sense of responsibi
icipated in "Mech-Tricks" in IMPULSE 2013 National Level Event. • Participated in "Tech-Quiz" in IMPULSE 2013 National Level Event. • Participated in "Manage
```

5. CONCLUSION :

Our application helps the recruiters to screen the resumes more efficiently there by reducing the cost of hiring. This will provide potential candidate to the organization and the candidate will be successfully be placed in an organization which appreciate his/her skill set and ability. An analysis of various NERC tools has been presented in this study. The evaluation proposes a model that eliminates some of the competitions' limitations into assessing these tools. This model is based on the creation of a small corpus, and the adaptation of the evaluation methodology to the NERC typology of the tools, not the contrary as it is common in the major competitions. The analysis of all the identified entities and the errors committed during this process permits a study using data mining in order to discover the most frequent errors in the identification and classification of NEs.

6. REFERENCES:

1. Grishman, R., Sundheim, B.: Message Understanding Conference-6: A Brief History. In: Proc. 16th Conference on Computational Linguistics, vol. 1, pp. 466—471. ACL: NJ (USA) (1996)

2. F. N. A. Al Omran and C. Treude, "Choosing an NLP Library for Analyzing Software Documentation: A Systematic Literature Review and a Series of Experiments," 2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR), Buenos Aires, 2017, pp. 187-197.

3."NER' 19," in IEEE Pulse, vol. 9, no. 5, pp. C4-C4, Sept.- Oct. 2018.

4. Tjong Kim Sang, Erik. F.; De Meulder, F. 2003. Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. In Proc. Conference on Natural Language Learning. pp. 142-147. Edmonton, Canada (2003).

5. H. Joshi and G. Bamnote, "Distributed database: A survey," Interna-tional Journal Of Computer Science And Applications, vol. 6, no. 2, 2013.

6. R. Narasimhan and T. Bhuvaneshwari, "Big dataa brief study," Int. J. Sci. Eng. Res, vol. 5, no. 9, pp. 350 – 353, 2014

7. .Cunningham, H., Maynard D., Tab-lan V., Ursu C., Bontcheva K.: Developing Language
Processing Components with GATE. GATE v5 User Guide, http://www.gate.ac.uk/sale/tao/tao.pdf

8. Witten, I.H., Franck, E., Trigg, L., Hall, M., Holmes G., Cunningham S.J.: Weka: Practical
machine learning tools and techniques with Java implementations, Proc. ANNES'99
International Workshop on emerging Engineering and Connectionnist-based Information
Systems, pp. 192 – 196 (1999).

9. S. K. Kopparapu, "Automatic extraction of usable information from unstructured resumes to aid search," 2010 IEEE International Conference on Progress in Informatics and Computing, Shanghai, 2010, pp. 99-103

10. G. Prasad and K. K. Fousiya, "Named entity recognition approaches: A study applied to English and Hindi language," 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015], Nagercoil, 2015, pp. 1-4.

11. J. Guo, G. Xu, X. Cheng, and H. Li, "Named entity recognition in query," in *SIGIR*, 2009, pp. 267 – 274.

12. K. Balog, P. Serdyukov, and A. P. De Vries, "Overview of the trec 2010 entity track," in *TREC*, 2010.

13. L. d. Corro, A. Abujabal, R. Gemulla, and G. Weikum, "Finet: Context-aware fifine-grained named entity typing," in *EMNLP*, 2015, pp. 868 – 878.

14. Z. Ji, A. Sun, G. Cong, and J. Han, "Joint recognition and linking of fifine-grained locations from tweets," in *WWW*, 2016, pp. 1271 – 1281.

15. Q. Wei, T. Chen, R. Xu, Y. He, and L. Gui, "Disease named entity recognition by combining conditional random fifields and bidirectional recurrent neural networks," *Database*, vol. 2016, 2016.