Warning: It's a scam!! Towards understanding the Employment Scams using Knowledge Graphs

Nidhi Goyal
IIIT-Delhi
New Delhi, India
nidhig@iiitd.ac.in

Radhika Mamidi
IIIT-Hyderabad
Hyderabad, India
radhika.mamidi@iiit.ac.in

Niharika Sachdeva
InfoEdge India Limited
Noida, India
niharika.sachdeva@infoedge.com

Ponnurangam Kumaraguru
IIIT-Hyderabad
Hyderabad, India
pk.guru@iiit.ac.in

ABSTRACT

Employment scams, such as scapegoat positions, clickbait and non-existing jobs, etc., are among the top five scams registered over online platforms.[1] Generally, scam complaints contain heterogeneous information (money, location, employment type, organization, email, and phone number), which can provide critical insights for appropriate interventions to avoid scams. Despite substantial efforts to analyze employment scams, integrating relevant scam-related information in structured form remains unexplored. In this work, we extract this information and construct a large-scale Employment Scam Knowledge Graph consisting of 0.1M entities and 0.2M relationships. Our findings include discovering different modes of employment scams, entities, and relationships among entities to alert job seekers. We plan to extend this work by utilizing a knowledge graph to identify and avoid potential scams in the future.

CCS CONCEPTS

• **Information systems → Spam detection**.

KEYWORDS

Employment Scams, Fraudulent, Information Extraction, Imposters, complaints, Knowledge Graphs

ACM Reference Format:

Nidhi Goyal, Radhika Mamidi, Niharika Sachdeva, and Ponnurangam Kumaraguru. 2023. Warning: It's a scam!! Towards understanding the Employment Scams using Knowledge Graphs. In *6th Joint International Conference on Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD) (CODS-COMAD 2023), January 4–7, 2023, Mumbai, India.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3570991.3571053

[1]https://www.bbbmarketplacetrust.org/

CODS-COMAD 2023, January 4–7, 2023, Mumbai, India

ACM ISBN 978-1-4503-9797-1/23/01.
https://doi.org/10.1145/3570991.3571053

1 INTRODUCTION

Employment scams are among the top five scams registered over consumer complaint platforms. These scams typically include jobs from fake recruiters, fake check scams[2] asking for money or offering attractive non-existing jobs. Figure 1 shows a sample Employment scam complaint from a popular platform. According to report[2] of employment scams, 14 million people are exposed to employment scams with more than $2 billion lost per year. These reports focus

"I applied for an admin job on Indeed. I received an interview over the phone and did not think anything of it because of the Coronavirus. I was informed that I had the job and it was a work from home. I was issued a check in the amount of $4,800 off of a M&T business account. I finally got an information that the check was fraudulent."

Figure 1: A sample Employment scam complaint from BBB platform. Entities present in text are shown in color.

on understanding the overall impact of these employment scams, providing prevention tips and structural interventions to help people avoid losing money. Researchers [4, 5] have previously analyzed, normalized, and detected online job scams.

However, there are a few limitations to these approaches, a) most of these qualitative studies are survey-based and are costly, b) these methods also ignore the heterogeneous information about money, location, employment type, organization, email, and phone number in the unstructured text that can help prevent consumers from losing money, c) understanding the victim's perspective (reported scams) and providing them objective information about scams is still unexplored in the domain. To address these issues, Knowledge graphs [3] are considered suitable to capture heterogeneous information in form of entities and their relationships. A knowledge graph is a multi-relational graph that stores the information in triples, consisting of entities as nodes and relations as different types of edges. Many research teams have organized the knowledge in their domain into a structured format, e.g., YAGO [6], NELL [1], and DBpedia [8].

Towards this end, we collect a dataset of six years consisting of 17K scam complaints from the BBB website.[3] Further, we investigate and extract the entities (money, location, date, email, phone

[2]https://bbbfoundation.images.worldnow.com/library/d8707e47-c886-48ec-b143-7b3db2806658.pdf
[3]https://www.bbbmarketplacetrust.org/story/39089075/bbb-scam-tracker

number, employment type, and organization) from employment scam complaints. Then, we construct an Employment Scam Knowledge Graph consisting of 0.1M entities and 0.2M relationships to help workers protect themselves before falling for scams. To the best of our knowledge, this is the first work to leverage and obtain structured information in knowledge graphs from scam complaints. This is our preliminary work, and we plan to extend this work by leveraging graph representations to identify and avoid potential scams using this Employment scam knowledge graph.

2 DATASET

The dataset source for this research is BBB Scam Tracker[1], an online platform where consumers and businesses report scams. BBB scam tracker website was accessible at the time of data collection. We collected around 17K employment scam complaints of USA and Canada from August 2015 to April 2021. We consider the employment scam complaints, which have a non-empty description field. The dataset consists of the scam's reported date, textual descriptions, the dollar value of any loss, zipcode, and business name (imposter).

3 APPROACH

The approach consists of three components a) Understanding the complaints, b) Entity Extraction, and c) Knowledge integration. We extract the mode of approaching the victim from textual descriptions using a set of keywords such as Email, mailbox, inbox, spammed by inbox, text message, Facebook, phone call, call, called me up, social media, Instagram, Twitter, LinkedIn, personal message, WhatsApp, message, telegram, phone message, and contact number. We categorize them into six significant ways, i.e., *'email'*, *'call'*, *'online social networks'*, *'online professional networks'*, *'text message'* and others. These are the most common modes scammers use to approach victims online [2]. We use rule-based heuristics and Spacy NER [7] to identify these entities from text. There are 2.2K unique money values, 1.9K unique dates, 1.1K unique URLs, 4.6K unique organisations, 0.2K unique phone numbers, and 55 individual states from employment scam complaints content. We define seven types of relationships (*'reported_from'*, *'reported_date'*, *'job_type'*, *'contact'*, *'mentions_money'*, *'mentions_org'*, and *'email'*) from these complaints. To integrate the knowledge, we extract these entities and relationships and store them efficiently in graphical storage, i.e., Neo4j Cypher Query Framework [9]. Figure 2 shows a sample snapshot of an Employment Scam Knowledge Graph.

4 DISCUSSION

Our findings show that 61% of the complaints mentioned *'email'* and 12% are engaged through *'online professional networks'*. 10% scammed through *'online social networks'*, usually Facebook chats. At least 500 scam complaints mentions *'work from home'*, *'Covid'* or *'Corona'*. For scammers, email is the most common method of approaching the victim. We observe that scammers impersonated well-known retailers like *Amazon* and *Walmart* to seem legitimate, posting scam jobs on major online employment platforms. We find that the money or amount mentioned in the complaint text was mostly $32). Interestingly, there was a surge in scams during late 2018, where scammers impersonated popular business names using

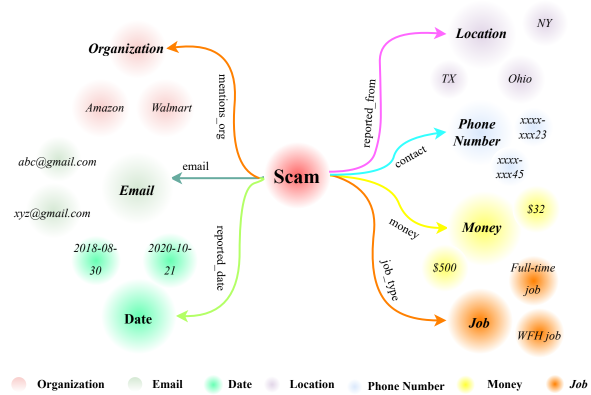

Figure 2: A sample snapshot of Employment Scam Knowledge Graph. All the entity types are shown in bold. For better visualization, we show edges among the entity types in snapshot. For instance, a fact in the KG is of form (Scam, mentions_org, Amazon).

amazon jobs. Such information, when converted to KG, is helpful for victims to protect themselves from potential scams. This is our preliminary work, and we aim to extend this work by proposing the strategies that precarious job seekers can utilize to identify and avoid scams for future research directions in the area.

ACKNOWLEDGMENTS

This work is done as a part of ACM Anveshan Setu Programme. We would like to thank the LTRC, IIIT-H, and Precog Research Group, who gave us continued support throughout the project.

REFERENCES

[1] Dieter Fensel, Umutcan Şimşek, Kevin Angele, Elwin Huaman, Elias Kärle, Oleksandra Panasiuk, Ioan Toma, Jürgen Umbrich, and Alexander Wahler. 2020. Introduction: what is a knowledge graph? In *Knowledge Graphs*. Springer, 1–10.
[2] Deanna Grant-Smith, Alicia Feldman, and Cassandra Cross. 2022. Key trends in employment scams in Australia: What are the gaps in knowledge about recruitment fraud. *QUT Centre for Justice Briefing Papers* 21 (2022).
[3] Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard De Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, et al. 2021. Knowledge graphs. *ACM Computing Surveys (CSUR)* 54, 4 (2021), 1–37.
[4] Syed Mahbub, Eric Pardede, and ASM Kayes. 2022. Online Recruitment Fraud Detection: A Study on Contextual Features in Australian Job Industries. *IEEE Access* 10 (2022), 82776–82787.
[5] Alexandrea J Ravenelle, Erica Janko, and Ken Cai Kowalski. 2022. Good jobs, scam jobs: Detecting, normalizing, and internalizing online job scams during the COVID-19 pandemic. *New Media & Society* 24, 7 (2022), 1591–1610.
[6] Thomas Rebele, Fabian Suchanek, Johannes Hoffart, Joanna Biega, Erdal Kuzey, and Gerhard Weikum. 2016. YAGO: A multilingual knowledge base from wikipedia, wordnet, and geonames. In *International semantic web conference*. Springer, 177–185.
[7] Xavier Schmitt, Sylvain Kubler, Jérémy Robert, Mike Papadakis, and Yves LeTraon. 2019. A replicable comparison study of NER software: StanfordNLP, NLTK, OpenNLP, SpaCy, Gate. In *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. IEEE, 338–343.
[8] Kuldeep Singh, Ioanna Lytra, Arun Sethupat Radhakrishna, Saeedeh Shekarpour, Maria-Esther Vidal, and Jens Lehmann. 2020. No one is perfect: Analysing the performance of question answering components over the dbpedia knowledge graph. *Journal of Web Semantics* 65 (2020), 100594.
[9] Jim Webber and Ian Robinson. 2018. *A programmatic introduction to neo4j*. Addison-Wesley Professional.