Identify Relevant Entities Through Text Understanding

Pooja Oza
pho1003@wildcats.unh.edu
University of New Hampshire
Supervised by: Laura Dietz

ABSTRACT

An Entity Retrieval system is a fundamental task of Information Retrieval that provides direct answer to an information need of user. Prior work of entity retrieval utilizes either the Knowledge Graph fields or the text relevant to the query via pseudo-relevance feedback [11] to improve the performance. Recently, Knowledge Graph embeddings or other entity representations, which capture the entity information from a knowledge graph are shown to be beneficial for entity retrieval. However, such embeddings are query-agnostic. In this dissertation work, we aim to improve entity retrieval by exploring the pseudo-relevance feedback to generate entity representations that capture query-aware entity information to determine the relevance of entities. We study the effectiveness of pseudo-relevance feedback against Knowledge Graph fields and investigate the efficacy of the Knowledge Graph embeddings for entity retrieval. We aim to understand the importance of utilization of query-aware signals and modeling of such signals with Knowledge Graph embeddings. Our results show that pseudo-relevance feedback is more effective than the Knowledge Graph fields by 30%.

KEYWORDS

Entity Retrieval, Knowledge Graph embeddings

ACM Reference Format:
Pooja Oza. 2022. Identify Relevant Entities Through Text Understanding. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22), October 17–21, 2022, Atlanta, GA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3511808.3557819

1 INTRODUCTION

An entity retrieval system responds to the information needs of the users which can range from keywords such as "Coffee" to natural language questions such as "Who is the driver who has won the most races in F1?" either by list of entities or a particular entity. Pound et al. [17] finds that a great fraction of users' queries are entity-oriented, 40% of the queries are for a particular entity, 12% are for a list of entities and 5% are for a particular attribute of an entity. Recent work in Entity Retrieval [8] utilizes Knowledge Graph embeddings to identify relevant entities for an entity-oriented query. These embeddings are static embeddings and do not contain any query-aware entity information that is beneficial to determine the relevance of entities effectively. In entity-oriented queries, the relevance of the entities depends on the information need. In this dissertation, we improve the performance of entity retrieval system by addressing two aspects of the search: (1) **Entity Representations:** generate entity embeddings that are more relevant to the query than the general-purpose knowledge graph embeddings; (2) **Entity Retrieval:** identify relevant entities to a query through relevant text and entity representations.

To improve the performance of an entity retrieval system, we propose to explore the text corpora, specifically the text that is relevant to the query via, pseudo-relevance feedback. We study the text retrieved by pseudo-relevance feedback to reveal influential signals that can effectively identify the relevance of entities for the query. We further examine whether these influential signals can be used to obtain better entity representations, by integrating them as a complementary source of entity information to the existing Knowledge Graph embeddings. In summary, our goal is to introduce query-aware entity information from a textual information source to generate query-aware entity representation. Our novelty lies in using influential signals from pseudo-relevance feedback that integrates entity information for a query, such as a different entity relatedness measure and entity relevance, to the static knowledge graph embeddings. Specifically, we aim to address the following research questions: ***RQ1)*** Is entity information derived from pseudo-relevance feedback more influential than Knowledge Graph entity information at identifying the entity relevance? ***RQ2)*** Is the entity relatedness captured in the Knowledge Graph embeddings sufficient to determine the relevance of entities? ***RQ3)*** How can we leverage the entity information derived from pseudo-relevance feedback to generate better entity representations?

2 RELATED WORK

2.1 Knowledge Graph Embeddings

Knowledge graph embeddings are vector representations of entities and their properties in a continuous vector space. Different approaches are used to generate knowledge graph embeddings. One of the well-known embeddings is Wikipedia2Vec [23] that jointly learns entity and word embeddings in the same vector space by capturing both Wikipedia link structure and text. Likewise, TransE [2], is a translation-based embedding model. It captures the relations between entities through triplets *(head, label, tail)* where the *label* corresponds to a translation between *head* and *tail*. However, TransE is based on 1-to-1 relation triples and does not work well for 1-to-N, N-to-N, N-to-1 relationships. To overcome this, several other Knowledge Graph embeddings are proposed such as TransH [7] which projects head and tail entities in relation-specific hyperplane or TransR [13] which projects each relationship in its specific space while learning embeddings.

CIKM '22, October 17–21, 2022, Atlanta, GA, USA

ACM ISBN 978-1-4503-9236-5/22/10…$15.00
https://doi.org/10.1145/3511808.3557819

2.2 Knowledge-enhanced BERT

Recently, various Pre-trained Knowledge-enhanced Language Models (PKLMs) infuse Knowledge Graph embeddings of Wikipedia2Vec [23] and TransE [2] in Language Models (LMs) to incorporate LMs with additional entity information. E-BERT[16] fine-tunes BERT [4] with Wikipedia2Vec [23] entity and word embeddings, ERNIE [24] integrates the knowledge entity-wise through TransE [2] entity embeddings while KEPLER [21] first converts the relational triplets to text and incorporates the text into BERT [4] embeddings. KELM [14] endows LMs with the multi-relational subgraph extracted from the Knowledge Graph. Most of the PKLMs are re-pretrained on a large corpus to infuse the knowledge into LMs. While PKLMs aims to enrich LMs with knowledge, our work focuses on enhancing the Knowledge Graph embeddings with query-aware information to generate entity representations.

2.3 Entity Retrieval

Previous work of Entity Retrieval uses fields of the knowledge graph such as BM25F [19] which uses term frequencies of different fields in documents. One well-known method [18] uses the sequential dependence model (SDM) to estimate the relevance of different entity fields, names, documents, and types. Gerritse et al. [8] uses entity representation of Wikipedia2Vec to re-rank the initial candidate set of the entities. The scoring function is based on the cosine similarity of entity embeddings of entities mentioned in the query and document.

On the other hand, several works use feedback runs of the query for the entity retrieval task. Dalton et al. [3] use entity-linked feedback runs and other sources such as entity links and candidate sets from entity rankings to compute an expansion model for document retrieval. In recent work, Dietz [5] retrieves various features from entity and text feedback runs such as entity co-occurrence, entity-mention features, etc., and combine these features in the Learning-To-Rank framework to rank the entities.

Our work differs from the previous work in our approach of using the context of the entities provided by the feedback runs to generate better entity representations. We combine the entity information of Knowledge Graph entity representations with the entity information from feedback runs.

3 PROPOSED APPROACH

The focus of this dissertation is to exploit entity information in text, in particular pseudo-relevance feedback, and utilize it to generate entity representations with the goal to satisfy complex information needs of the users.

3.1 Published - Entity Information from Pseudo-Relevance Feedback

3.1.1 ***Approach:*** Our work [15] finds that pseudo-relevance feedback is more effective source of information than the entity information present in KG. While previous work uses pseudo-relevance feedback for Entity Retrieval, they focus on the structured annotations such as entity names, types and categories in KG or graph walks on related entities. Our approach focuses on the text surrounding the entity mentions. In our analysis, we observe entity information which serves as signals to identify entity relevance as follows:

(1) Entities that frequently appear together i.e., mentioned together, in the text of pseudo-relevance feedback are more relevant to the query. For example, if the entities *Coffee bean* and *Coffee house* appear frequently together in the pseudo-relevance feedback text, they are more relevant. The feature *Co-occurrence Count* in Table 1 reflects this finding.
(2) We infer connections between entities expressed in text and find that entities that are mentioned together in the pseudo-relevance feedback indicate relevant connections. Feature *Co-occurrence Relevance* in Table 1.
(3) Entities that appear frequently, feature *Mention-Freq*, in pseudo-relevance feedback text are more relevant to the query. For example, if an entity such as *Starbucks* is appearing frequently in the pseudo-relevance feedback, it is more relevant.
(4) Entity relatedness measured through Knowledge Graph links is less effective to determine the relevance of entities. This finding is reflected in Table 1 by features *Outlinks, Inlinks, KG-Entity*.

Table 1: Overall performance results from our work [15]. The features Co-occurrence Relevance, Co-occurrence Count, Mention-Freq are the signals from pseudo-relevance feedback of queries. The features Outlinks, Inlinks and KG-Entity are based on Wikipedia's link graph and entity description text. As seen in the results, influential signals from pseudo-relevance feedback outperform the Knowledge Graph fields. For more results, please refer to our work [15]

Feature	MAP	Recall	F1
Co-occurrence Relevance	0.2140	0.3092	0.3106
Co-occurrence Count	0.1669	0.2636	0.2714
Mention-Freq	0.1971	0.3032	0.3189
Outlinks	0.1509	0.2568	0.2722
Inlinks	0.1566	0.2630	0.2839
KG-Entity	0.0172	0.0519	0.0499

Most of the prior work focuses on identifying a small high precision list of correct entities, whereas our approach aims for high recall sets of relevant entities of varying degree of relevance to the query.

These findings leads us to conclude that the text relevant to a query provides distinguishable entity information which is different for every query while the general-purpose Knowledge Graphs provides static entity information for any query. This distinction is important because the entity information such as inlinks, outlinks, entity description of any entity retrieved from the Knowledge Graph would contain high volume of irrelevant data to a query.

3.1.2 ***Methodology:*** We retrieve the ranking of the entity-linked pseudo-relevance feedback text and infer the connections between entities via co-occurrence. We model the relevance of co-occurrence connections through frequency and reciprocal rank of the feedback runs. The fields of the Knowledge Graph considered are outlinks, inlinks, and entity descriptions.

3.1.3 ***Evaluation:*** We use TREC Complex Answer Retrieval (CAR) dataset [6] which consists of keyword queries such as "Chocolate". For evaluation, we use MAP, Recall, and F1 scores. The results are shown in Table 1.

3.2 Ongoing - Entity representation from Pseudo-Relevance Feedback

3.2.1 ***Approach:*** Knowledge Graph embeddings are entity representations that capture the relatedness of the entities through different measures such as Wikipedia link graph or relation triples. We explore these entity representations to understand how these different measures determine the relevance of the entities to a query in an entity retrieval system. To this end, we attempt to address our next two research questions:

RQ2: Are Knowledge Graph embeddings sufficient? In our analysis of the static Knowledge Graph embeddings, we observe that they are less effective at identifying relevant entities. We further investigate through KPLMs whether the lacking performance of the entity representations is due to its static nature. We find a similar pattern in contextualized entity representations generated via KPLMs. These entity representations, unlike the static Knowledge Graph embeddings, contain contextualized information through BERT [4] embeddings.

This leads us to conclude that the general-purpose entity representations that are generated from large text corpora, which contain a mix of multiple diverse topics, lose distinguishable characteristics of a query thus falling short for entity retrieval. This prompts us to ask our next research question: *How can we leverage the query-aware entity information to generate better entity representations?*

3.2.2 ***Methodology:*** We follow the re-ranking framework of Gerritse et al. [8] where we retrieve a candidate set of entities for queries and use entity embeddings of Wikipedia2Vec [23], ERNIE [24] and E-BERT [16]. We entity link the queries and take cosine similarity between entity embeddings and query embeddings to rank the entities. We further use the Learning-To-Rank framework to get the final ranking of entities.

3.2.3 ***Evaluation:*** We use two datasets, TREC CAR [6] and DBpediaV2-CAR version of DBpedia Entity-V2 [9]. We provide the results in Table 2. We use the best entity ranking feature of Dietz [5] for TREC CAR and BM25F-CA [9] as baselines. L2R-* are the interpolated results of the model entity ranking with the baseline. We take cosine similarity between the entity embeddings and query embedding for the models Wikipedia2Vec, ERNIE, and E-BERT.

RQ3: Entity Information from Text for Entity representations.

3.2.4 ***Approach:*** Our idea is to integrate query-aware entity information with the general-purpose Knowledge Graph entity information to generate entity representations that are query-aware. This also gives us an advantage of combining entity information from two different information sources i.e., the general-purpose Knowledge Graphs and text of pseudo-relevance feedback.

We introduce query-aware entity information through two different components: **(1) entity relatedness (2) relevance**. Based on our observations from our previous work [15], we model the entity relatedness via entity co-occurrence connections expressed in the pseudo-relevance feedback text. The rationale behind this approach, is the observation in our previous work [15] that entities that are mentioned together in the pseudo-relevance feedback text reveal relevant information to the query.

Table 2: Overall Performance Evaluation. As seen in the results below, the Knowledge Graph embeddings perform worse than the baselines. We display results of TREC CAR and only one type of query for DBpedia-v2 here due to space constraint.

Dataset	TREC-CAR		ListSearch	
Model	MAP	Recall	MAP	Recall
Baseline	0.1573	0.2226	0.4406	0.4269
Wikipedia2Vec	0.0838	0.1293	0.3971	0.4453
ERNIE	0.061	0.2268	0.3277	0.3959
E-BERT	0.0752	0.219	0.3813	0.4206
L2R-Wikipedia2Vec	0.1661	0.2275	0.4528	0.4358
L2R-ERNIE	0.1625	0.2268	0.4543	0.4391
L2R-E-BERT	0.1599	0.219	0.447	0.4317

One way we model the relevance is through frequency, i.e., the entity that co-occurs frequently with other entities in the text, because in our previous work [15] we found that the entities that co-occur frequently are relevant entities. The idea is that the entity that has a high degree of connections is more relevant to the query. As an alternative, we integrate the relevance is through similarity between the entities that co-occur in the pseudo-relevance feedback. This gives us an insight about how similar the entities connected through the relevant connections of a query are?

3.2.5 ***Methodology:*** We retrieve an entity-linked candidate set of passages as pseudo-relevance feedback for queries. We construct an entity co-occurrence graph for each query of the dataset and use Graph Convolutional Network [10] and Graph Attention Network [20] to model the relevance. We evaluate them on the entity retrieval task. We use the evaluation metrics MAP and Recall.

3.3 Future - Contextualized Entity Descriptions and Entity Relations for Entity Retrieval

3.3.1 ***Approach:*** To further improve the entity retrieval, we propose to enrich the Knowledge Graph embeddings as follows:

Contextualized Entity Descriptions. Recently, research has been focused on enhancing the conventional Knowledge Graph embeddings with textual descriptions [1, 22]. However, these entity descriptions (1) pertain to the relation (predicate) between the two entities and (2) are the generic descriptions of the entities found in knowledge bases. The generic entity descriptions from the knowledge bases, may contain information that would be irrelevant to the query. Moreover, they do not capture all different aspects in which an entity might appear. For example, the query "economic policies of Barack Obama" does not benefit from the generic description of the entity "Barack Obama" in the knowledge base. Our idea is to enrich the entity representation with the entity descriptions which are relevant to the query. We leverage text analysis on top of pseudo relevance feedback to extract the context in which the entities are mentioned with respect to the query.

Entity Relations. Conventional Knowledge Graph embeddings such as TransE [2] are based on 1-to-1 relations. PtransE [12] proposes to use the multi-step relational path, consisting of sequence of relations, to model complex relations between the entities. PtransE [12] considers the relational paths between two entities based on Path Constrained Resource Allocation [25] algorithm. Inspired by [12], we propose the idea to consider the relations between the entities based on the query. For example, the query "Barack Obama Office" should consider the relation "Barack Obama $\xrightarrow{\text{VicePresident}}$ Joe Biden" and the query "Presidents after Barack Obama" should consider the relation "Barack Obama $\xrightarrow{\text{SucceededBy}}$ Donald Trump $\xrightarrow{\text{SucceededBy}}$ Joe Biden".

Integration of Text and Entities: Previous work finds that relevant text can be found by combining the signals from text and entities. We propose the idea to infuse the entity descriptions and entity connections through a bi-partite graph of entity descriptions and entity relevant connections.

4 CONCLUSION

The goal of this dissertation is to improve entity retrieval systems by developing text analysis methods that can extract query-aware entity representations from pseudo-relevance feedback. We study the efficacy of entity representations from Knowledge Graph embeddings and describe approaches to infuse query-aware entity relevance into entity embeddings which are beneficial for downstream IR tasks. While we study the entity representations in the context of entity retrieval system, other applications such as question-answering, conversational search, and recommender systems would also benefit.

5 ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1846017. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Bo An, Bo Chen, Xianpei Han, and Le Sun. 2018. Accurate text-enhanced knowledge graph representation learning. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. 745–755.

[2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-Relational Data. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2* (Lake Tahoe, Nevada) *(NIPS'13)*. Curran Associates Inc., Red Hook, NY, USA, 2787–2795.

[3] Jeffrey Dalton, Laura Dietz, and James Allan. 2014. Entity Query Feature Expansion Using Knowledge Base Links. In *Proceedings of the 37th International ACM SIGIR Conference on Research amp; Development in Information Retrieval* (Gold Coast, Queensland, Australia) *(SIGIR '14)*. Association for Computing Machinery, New York, NY, USA, 365–374. https://doi.org/10.1145/2600428.2609628

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[5] Laura Dietz. 2019. ENT Rank: Retrieving Entities for Topical Information Needs through Entity-Neighbor-Text Relations. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Paris, France) *(SIGIR'19)*. Association for Computing Machinery, New York, NY, USA, 215–224. https://doi.org/10.1145/3331184.3331257

[6] Laura Dietz and Ben Gamari. 2018. TREC Complex Answer Retrieval. *Proceedings of Text REtrieval Conference (TREC)* (2018).

[7] Jianlin Feng. 2014. Knowledge Graph Embedding by Translating on Hyperplanes.

[8] Emma J Gerritse, Faegheh Hasibi, and Arjen P de Vries. 2020. Graph-Embedding Empowered Entity Retrieval. In *Advances in Information Retrieval, Proceedings of the 42nd European Conference on Information Retrieval (ECIR 2020)* (Lisbon, Portugal) *(Lecture Notes in Computer Science)*. Springer, Cham, 97–110. https://doi.org/10.1007/978-3-030-45439-5_7

[9] Faegheh Hasibi, Fedor Nikolaev, Chenyan Xiong, Krisztian Balog, Svein Erik Bratsberg, Alexander Kotov, and Jamie Callan. 2017. DBpedia-Entity v2: A Test Collection for Entity Search. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Shinjuku, Tokyo, Japan) *(SIGIR '17)*. Association for Computing Machinery, New York, NY, USA, 1265–1268. https://doi.org/10.1145/3077136.3080751

[10] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[11] Victor Lavrenko and W Bruce Croft. 2009. *A generative theory of relevance*. Vol. 26. Springer.

[12] Yankai Lin, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, and Song Liu. 2015. Modeling relation paths for representation learning of knowledge bases. *arXiv preprint arXiv:1506.00379* (2015).

[13] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning Entity and Relation Embeddings for Knowledge Graph Completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (Austin, Texas) *(AAAI'15)*. AAAI Press, 2181–2187.

[14] Yinquan Lu, Haonan Lu, Guirong Fu, and Qun Liu. 2021. KELM: Knowledge Enhanced Pre-Trained Language Representations with Message Passing on Hierarchical Relational Graphs. *CoRR* abs/2109.04223 (2021). arXiv:2109.04223 https://arxiv.org/abs/2109.04223

[15] Pooja Oza and Laura Dietz. 2021. Which entities are relevant for the story?. In *CEUR workshop proceedings*, Vol. 2860.

[16] Nina Poerner, Ulli Waltinger, and Hinrich Schütze. 2020. E-BERT: Efficient-Yet-Effective Entity Embeddings for BERT. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, Online, 803–818. https://doi.org/10.18653/v1/2020.findings-emnlp.71

[17] Jeffrey Pound, Peter Mika, and Hugo Zaragoza. 2010. Ad-Hoc Object Retrieval in the Web of Data. In *Proceedings of the 19th International Conference on World Wide Web* (Raleigh, North Carolina, USA) *(WWW '10)*. Association for Computing Machinery, New York, NY, USA, 771–780. https://doi.org/10.1145/1772690.1772769

[18] Hadas Raviv, David Carmel, and Oren Kurland. 2012. A Ranking Framework for Entity Oriented Search Using Markov Random Fields. In *Proceedings of the 1st Joint International Workshop on Entity-Oriented and Semantic Search* (Portland, Oregon, USA) *(JIWES '12)*. Association for Computing Machinery, New York, NY, USA, Article 1, 6 pages. https://doi.org/10.1145/2379307.2379308

[19] Stephen Robertson and Hugo Zaragoza. 2009. The Probabilistic Relevance Framework: BM25 and Beyond. *Found. Trends Inf. Retr.* 3, 4 (apr 2009), 333–389. https://doi.org/10.1561/1500000019

[20] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).

[21] Xiaozhi Wang, Tianyu Gao, Zhaocheng Zhu, Zhengyan Zhang, Zhiyuan Liu, Juanzi Li, and Jian Tang. 2021. KEPLER: A unified model for knowledge embedding and pre-trained language representation. *Transactions of the Association for Computational Linguistics* 9 (2021), 176–194.

[22] Jiacheng Xu, Kan Chen, Xipeng Qiu, and Xuanjing Huang. 2016. Knowledge graph representation with jointly structural and textual encoding. *arXiv preprint arXiv:1611.08661* (2016).

[23] Ikuya Yamada, Akari Asai, Jin Sakuma, Hiroyuki Shindo, Hideaki Takeda, Yoshiyasu Takefuji, and Yuji Matsumoto. 2020. Wikipedia2Vec: An Efficient Toolkit for Learning and Visualizing the Embeddings of Words and Entities from Wikipedia. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, Online, 23–30. https://doi.org/10.18653/v1/2020.emnlp-demos.4

[24] Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu. 2019. ERNIE: Enhanced Language Representation with Informative Entities. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, 1441–1451. https://doi.org/10.18653/v1/P19-1139

[25] Tao Zhou, Jie Ren, Matúš Medo, and Yi-Cheng Zhang. 2007. Bipartite network projection and personal recommendation. *Physical review E* 76, 4 (2007), 046115.

A DISCUSSION TOPICS

In general, I would like to discuss the following topics with my mentor and peers:

(1) Entity Retrieval
(2) Knowledge Graph embeddings
(3) Knowledge-enhanced BERT embeddings
(4) Future Research Directions
(5) Feedback

In particular, I would like to discuss the following:

(1) Discuss approaches to constructing query-aware entity descriptions for entity retrieval. In particular, how can we use pseudo-relevance feedback to construct those descriptions? Are there any suggestions other than using pseudo-relevance feedback to generate entity descriptions?
(2) I would also like to discuss the Knowledge Graph embeddings and their role in entity retrieval. Specifically, I would like to discuss the analysis of my ongoing work and how different entity embeddings affect entity similarity.
(3) Discuss the integration of query-aware context to identify relevant entities. Are there better approaches to integrating relevant information of a query in the Knowledge Graph embeddings?
(4) Are there any better approaches to identifying query-relevant information?
(5) Get suggestions on entity graph construction from the pseudo-relevance feedback text.
(6) Discuss the attention weights. Whether the attention weights also play a role in identifying relevance? Based on my ongoing work of Knowledge Graph embeddings.
(7) Discuss the current state-of-the-art in entity retrieval and future research directions.
(8) Feedback on my ongoing and proposed work.
(9) Understand and discuss my peers' research.

B ADVISOR'S STATEMENT ON SUPPORTING THE STUDENT'S ATTENDANCE

I am Laura Dietz, Associate Professor of Computer Science at the University of New Hampshire. I am Pooja Himanshu Oza's Ph.D. advisor since she joined our graduate program in 2018. I highly support Pooja for your PhD symposium as she is preparing for her Dissertation Proposal in the Academic Year 2022/2023. She is expected to graduate with a Ph.D. in Computer Science in Fall 2023.

Her research topic focuses on leveraging information from text and knowledge graphs for better query-specific entity representation. Her work was presented at ECIR's Text2Story workshop where she received wonderful feedback.

The pandemic hit Pooja very hard, with personal loss and motivational issues. She was very unlucky with her early research on using relation extraction technology to identify relevant relations, which while sounding promising in theory, did not lead to publishable results due to issues in the relation extraction pipeline. Ultimately, Pooja turned this unfortunate research experience around, which allowed her to clearly identify a research gap that is widely overlooked: most work on knowledge graph extraction obtains improvement on popular topics, including people and organizations, however the same technology fails when applied to general information needs, such as queries about environmental, natural science, and societal topics. Developing models that can represent entities for this important thematic complex is very important and hence will lead to high-impact research.

Pooja is currently preparing a full-paper reproducibility study and a full research paper as laid out under ongoing activities above. She has a clear vision for her final contribution before she graduates. She is at a critical point, where she would benefit most from the input of the research community. Attending the Ph.D. symposium at CIKM would also allow her to attend her first in-person research conference and finally knit the network that will allow her to establish a successful career.