ECIR 23 Tutorial: Neuro-Symbolic Approaches for Information Retrieval

Laura Dietz[1][0000-0003-1624-3907], Hannah Bast[2][0000-0003-1213-6776], Shubham Chatterjee[3,1][0000-0002-6729-1346], Jeff Dalton[3][0000-0003-2422-8651], Edgar Meij[4][0000-0003-0516-3688], and Arjen de Vries[5][0000-0002-2888-4202]

[1] University of New Hampshire, USA dietz@cs.unh.edu
[2] University of Freiburg, Germany bast@cs.uni-freiburg.de
[3] University of Glasgow, Scotland jeff.dalton@glasgow.ac.uk, shubham.chatterjee@glasgow.ac.uk
[4] Bloomberg, United Kingdom emeij@bloomberg.net
[5] Radboud University, The Netherlands arjen@acm.org

Abstract. This tutorial will provide an overview of recent advances on neuro-symbolic approaches for information retrieval. A decade ago, knowledge graphs and semantic annotations technology led to active research on how to best leverage symbolic knowledge. At the same time, neural methods have demonstrated to be versatile and highly effective. From a neural network perspective, the same representation approach can service document ranking or knowledge graph reasoning. End-to-end training allows to optimize complex methods for downstream tasks.
We are at the point where both the symbolic and the neural research advances are coalescing into neuro-symbolic approaches. The underlying research questions are how to best combine symbolic and neural approaches, what kind of symbolic/neural approaches are most suitable for which use case, and how to best integrate both ideas to advance the state of the art in information retrieval.

Keywords: Neural Networks · Semantics · IR

1 Motivation

Being able to reason on what is relevant for an information need is important for all kinds of information retrieval tasks: web search, question answering, dialogues, image search, task assistance, or e-commerce. As traditional keyword-matching approaches are successively being replaced with neural-representation approaches [17, 15], the question is whether symbolic approaches still have merit.

A decade ago, advances in knowledge graphs and semantic annotations, such as via entity linking, led to significant improvements in text ranking tasks [22, 6]. These, in turn, set new standards for entity-oriented downstream tasks like question answering [8, 1]. Now, neural representations for semantic annotations or other kinds of symbols have taken hold in the field of knowledge management. The information retrieval community is split into research that solely relies on neural representations (abandoning symbols altogether) and research that integrates neural and symbolic approaches.

Symbolic approaches have been studied in information retrieval over the years. The IR community has had a continued interest in entity retrieval tasks [9, 3, 2]. Sometimes, information needs are best answered using knowledge from external databases [24]—other times text can contextualize knowledge [20]. Furthermore, effective query expansion via pseudo-relevance feedback relies upon approaches that analyze retrieved documents—and reason about why these are relevant.

The goal of this tutorial is to consolidate findings and initiate a synergistic transfer across different IR-relevant use cases with respect to neuro-symbolic approaches.

2 Format and Target Audience

In this full-day tutorial, we will provide different perspectives on neural-symbolic methods, provide different perspectives on the topic, and discuss customizations for different use cases.

Our goal is to provide useful information to a *wide variety* of audiences. The first part of the tutorial will be *introductory*, designed to bring audience members up to speed who have only basic knowledge in neural representations and/or symbolic approaches, such as knowledge graphs and entity linking.

The second part will of interest to both a *beginners and intermediate audience*, where different speakers provide their own perspective on the topic and look at different use cases where we need to reason on what is relevant.

To conclude the tutorial, we will invite all speakers and some additional guests to a panel discussion. Our goal is to spur a discussion of what works where, when, and why.

3 Topics Covered

1. **Foundational Topics**
   Neural Text Representations. BERT and other large neural language models (LLMs) of text have led to tremendous increases in performance improvements. LLM-based document re-ranking models are either based on Siamese-models like the Duet Model [16], or transformers, such as mono-BERT or duo-BERT [15]. We also cover neural query expansion [21] and query rewriting [14].
   Symbols and Knowledge Graphs. Several repositories of symbolic knowledge are readily available: Entities derived from Wikipedia pages, or nodes in a knowledge graph such as DBpedia or Wikidata. Word-oriented knowledge bases for Common Sense Reasoning are COMET [12] as well as ConceptNet. Neural representations of such symbols are provided in E-BERT [19], Wikipedia2Vec [25], or BERT-ER [3]. Graph neural networks such as HOPE [18], allow to reason across the graph structure.

Text-Symbol Alignment and Semantic Annotations. The task of Entity Linking [11] (aka Wikification) is to annotate unstructured text with detected and disambiguated entity identifiers. Such entity links in the text can serve as a set of logical symbols to reason with. The entity links also provide a means to align the text with nodes in a knowledge graph to perform inferences in. For some tasks, such as for conversations, specialized entity linking methods obtain better performance [13]. Other fine-grained information can be extracted from text with relation extraction, semantic role labeling, type predictions, and entity-aspect linking. Neural alignment methods allow the utilization of information in text and symbols for better ranking quality, such as EM-BERT [9].

2. **Perspectives**

Reasoning about Relevance. Retrieval models aim to reason about what is relevant. Hence, we summarize related ideas from other areas, such as logic-based reasoning in knowledge graphs as well as Natural Language Inference (NLI). Some systems include retrieval into their neural inference models, such as REALM [10]. Following on research on probabilistic reasoning with neural approaches for logic-based reasoning in knowledge bases, such as FuzzQE [5]. Chain-of-Thought reasoners [23] are leveraging neural few-shot learners to generate well-reasoned arguments.

Ranking Wikipedia Entities/Aspects. Given a query and a knowledge graph, the entity ranking task is to retrieve and rank entities from the knowledge graph according to their relevance to the query. Entity ranking has also been shown to be useful for tasks that require an explicit semantic understanding of text [4]. Two broad directions for entity ranking are (1) Non-neural approaches that leverage symbols and semantic annotations in text, and (2) Neural approaches that leverage dense representations of entities learnt using neural networks. Finally, we discuss future directions for learning better entity representations for IR.

Explainability for Pseudo-relevance Feedback. Traditionally, pseudo-relevance feedback (PRF/RM3) is a technique to identify relevant terms for query expansion. This idea has been generalized to identify relevant entities for expanding queries with expansion entities [6], or augmenting neural representations [21].

In "Explainability" the focus shifts from making correct predictions to explaining why a prediction was made. One explainability approach is to analyze model gradients to approximate input importance [7]. In a PRF setting, such information can be used to glean information on why a document was deemed relevant, with the goal to augment and refine the search query.

3. **Different Use Cases**

Use Case: Question Answering on Knowledge Graphs. This task accepts a question in arbitrary natural language, which should then be translated to a corresponding structured query (for example, in SPARQL) on a given knowledge graph. The currently best approaches for solving this problem [1] are all inherently neuro-symbolic: the knowledge is given in strongly structured (symbolic) form, yet the learning is neural. Correspondingly, the challenges are twofold. The *symbolic* challenge is to understand the nature of the structured queries, which are often surprisingly complex and non-trivial—even for seemingly simple questions. The *neural* challenge is to learn a high-quality translation model that can handle also complex questions and requires little supervision.

Use Case: Task-based Assistance. Information agents support complex real-world tasks and must not only retrieve relevant information, but also perform complex tasks using external symbolic tools and computation. This requires grounded reasoning about information and world state that is multimodal across text, images, video, and structured knowledge. Further, they must support the user in explainable and controllable fashion that involves eliciting structured information and storing it in personal knowledge graphs to incorporate structured symbolic constraints ("make it vegan") as well as being adaptable to mood, situation, and skill level.

Use Case: Generating Relevant Articles. Some usage scenarios ask for long, multi-faceted answers without the need for a user to interact. The goal is to foresee obvious next questions, and be forthcoming with such information without being explicitly asked. To satisfy this use case requires to solve a range of inter-dependent tasks: (1) high-recall retrieval with broad coverage, (2) query-specific clustering for subtopic-detection, (3) organization of content into a sequential structure, and (4) summarization and natural language generation.

4. **Discussion Panel**: The goal is to identify synergistic opportunities across different use cases. We are discussing approaches that (according to the literature) are supposed to work, but do not yet yield satisfactory results, leaving ample room for improvement. We also debate some controversial questions, such as "Since we have neural text representations, do we really need symbolic approaches?" The tutorial presenters and panel speakers are selected because they represent a broad spectrum of expert opinions on the topic.

4 Presenters

Laura Dietz, Associate Professor, University of New Hampshire (main contact). Dr. Dietz focuses on integrating relevant-oriented tasks, using full-text search, Wikipedia knowledge, and fine-grained semantic annotations, along with subtopic extraction, content organization, and natural language generation. She organized the KG4IR Tutorial and Workshop Series (ICTIR 2016, WSDM 2017, SIGIR 2017, SIGIR 2018) and the TREC Complex Answer Retrieval track (2017–2019).

Hannah Bast, Full Professor, University of Freiburg. Dr. Bast is interested in all aspects of information retrieval, with a focus on efficiency, ease of use, and fully functional systems. Her search systems power DBLP, Google Maps, and maybe soon Wikidata. Her work combines indexing and search in full text and structured knowledge for downstream applications such as question answering.

Shubham Chatterjee, Research Associate, University of Glasgow. Dr. Chatterjee focuses on neural entity-oriented information retrieval and extraction, particularly text understanding using entities and entity ranking. His goal is to build an intelligent search system that can answer open-ended information needs.

Jeff Dalton, Senior Lecturer, University of Glasgow. Dr. Dalton focuses on methods for effectively leveraging knowledge for complex information-seeking tasks. His work on entity-based query feature expansion published at SIGIR in 2014 is one of the first to demonstrate the effectiveness of using general-purpose knowledge graphs for search. He is a Turing AI Acceleration Fellow at the Turing Institute with a prestigious UKRI fellowship, and the lead organizer of the TREC Conversational Assistance track.

Edgar Meij, Head of Search and Discovery, Bloomberg AI. Dr. Meij leads several teams of researchers and engineers that work on information retrieval, semantic parsing, question answering, and smart contextual suggestions under severe latency constraints. Together, these researchers and engineers build, maintain, and leverage the company's search, autocomplete, and question-answering systems. He has taught several tutorials and organized various workshops on knowledge graphs, entity linking, and semantic search at top-tier conferences.

Arjen de Vries, Full Professor, Radboud University. Dr. de Vries uses structured and unstructured information to improve information access. He works on entity linking as well as entity retrieval, demonstrating that having knowledge of the entities in the query can help improve retrieval performance for entity-oriented search tasks. Dr. de Vries organized the first information retrieval evaluation campaigns that looked beyond documents into entities—the Enterprise Search track at TREC and later, the Entity Ranking track at INEX.

References

1. Bast, H., Haussmann, E.: More accurate question answering on freebase. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. pp. 1431–1440 (2015)
2. Cao, N.D., Izacard, G., Riedel, S., Petroni, F.: Autoregressive entity retrieval. CoRR **abs/2010.00904** (2020), https://arxiv.org/abs/2010.00904
3. Chatterjee, S., Dietz, L.: BERT-ER: Query-Specific BERT Entity Representations for Entity Ranking. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 1466–1477. SIGIR '22, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3477495.3531944
4. Chatterjee, S., Dietz, L.: Predicting Guiding Entities for Entity Aspect Linking. In: Proceedings of the 31st ACM International Conference on Information and Knowledge Management. CIKM '22, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3511808.3557671

5. Chen, X., Hu, Z., Sun, Y.: Fuzzy logic based logical query answering on knowledge graphs. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 3939–3948 (2022)
6. Dalton, J., Dietz, L., Allan, J.: Entity query feature expansion using knowledge base links. In: Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 365–374. SIGIR '14, Association for Computing Machinery, New York, NY, USA (2014). https://doi.org/10.1145/2600428.2609628
7. Funke, T., Khosla, M., Rathee, M., Anand, A.: Zorro: Valid, sparse, and stable explanations in graph neural networks. IEEE Transactions on Knowledge and Data Engineering (2022)
8. Gerritse, E.J., Hasibi, F., de Vries, A.P.: Graph-embedding empowered entity retrieval. In: Advances in Information Retrieval, Proceedings of the 42nd European Conference on Information Retrieval (ECIR 2020). pp. 97–110. Lecture Notes in Computer Science, Springer, Cham (2020). https://doi.org/10.1007/978-3-030-45439-5_7
9. Gerritse, E.J., Hasibi, F., de Vries, A.P.: Entity-aware transformers for entity search. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 1455–1465. SIGIR '22, Association for Computing Machinery, New York, NY, USA (2022). https://doi.org/10.1145/3477495.3531971
10. Guu, K., Lee, K., Tung, Z., Pasupat, P., Chang, M.W.: Realm: retrieval-augmented language model pre-training. In: Proceedings of the 37th International Conference on Machine Learning. pp. 3929–3938 (2020)
11. van Hulst, J.M., Hasibi, F., Dercksen, K., Balog, K., de Vries, A.P.: Rel: An entity linker standing on the shoulders of giants. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 2197–2200 (2020)
12. Hwang, J.D., Bhagavatula, C., Le Bras, R., Da, J., Sakaguchi, K., Bosselut, A., Choi, Y.: (comet-) atomic 2020: On symbolic and neural commonsense knowledge graphs. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 6384–6392 (2021)
13. Joko, H., Hasibi, F., Balog, K., de Vries, A.P.: Conversational entity linking: Problem definition and datasets. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 2390–2397 (2021)
14. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L.: Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 7871–7880 (2020)
15. Lin, J., Nogueira, R., Yates, A.: Pretrained transformers for text ranking: BERT and beyond. CoRR **abs/2010.06467** (2020), https://arxiv.org/abs/2010.06467
16. Mitra, B., Craswell, N.: An updated duet model for passage re-ranking. arXiv preprint arXiv:1903.07666 (2019)
17. Mitra, B., Craswell, N., et al.: An introduction to neural information retrieval. Foundations and Trends® in Information Retrieval **13**(1), 1–126 (2018)
18. Ou, M., Cui, P., Pei, J., Zhang, Z., Zhu, W.: Asymmetric transitivity preserving graph embedding. In: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1105–1114 (2016)

19. Poerner, N., Waltinger, U., Schütze, H.: E-BERT: Efficient-Yet-Effective Entity Embeddings for BERT. In: Findings of the Association for Computational Linguistics: EMNLP 2020. pp. 803–818. Association for Computational Linguistics, Online (Nov 2020). https://doi.org/10.18653/v1/2020.findings-emnlp.71
20. Ponza, M., Ceccarelli, D., Ferragina, P., Meij, E., Kothari, S.: Contextualizing trending entities in news stories. In: Proceedings of the 14th ACM International Conference on Web Search and Data Mining. pp. 346–354 (2021)
21. Pradeep, R., Nogueira, R., Lin, J.: The expando-mono-duo design pattern for text ranking with pretrained sequence-to-sequence models. arXiv e-prints pp. arXiv–2101 (2021)
22. Reinanda, R., Meij, E., de Rijke, M., et al.: Knowledge graphs: An information retrieval perspective. Foundations and Trends® in Information Retrieval **14**(4), 289–444 (2020)
23. Saparov, A., He, H.: Language models are greedy reasoners: A systematic formal analysis of chain-of-thought. arXiv preprint arXiv:2210.01240 (2022)
24. Xiong, C., Liu, Z., Callan, J., Hovy, E.: Jointsem: Combining query entity linking and entity based document ranking. In: Proceedings of the 2017 ACM SIGIR Conference on Information and Knowledge Management. p. 2391–2394. CIKM '17, Association for Computing Machinery, New York, NY, USA (2017). https://doi.org/10.1145/3132847.3133048
25. Yamada, I., Asai, A., Sakuma, J., Shindo, H., Takeda, H., Takefuji, Y., Matsumoto, Y.: 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. pp. 23–30. Association for Computational Linguistics, Online (Oct 2020). https://doi.org/10.18653/v1/2020.emnlp-demos.4