# Implicit Entity Recognition, Classification and Linking in Tweets

Hawre Hosseini
hawre.hosseini@ryerson.ca
Laboratory for Systems, Software and Semantics (LS$^3$), Ryerson University
Toronto, Canada

## ABSTRACT

Linking phrases to knowledge base entities is a process known as *entity linking* and has already been widely explored for various content types such as tweets [1][4][6]. A major step in entity linking is to recognize and/or classify phrases that can be disambiguated and linked to knowledge base entities, i.e., *Named Entity Recognition and Classification*. Unlike common entity recognition and linking systems, however, we aim to recognize, classify, and link entities which are *implicitly* mentioned, and hence lack a surface form, to appropriate knowledge base entries. In other words, the objective of our work is to recognize and identify core entities of a tweet when those entities are not explicitly mentioned; this process is referred to as *Implicit Named Entity Recognition and Linking* [5] [2].

## 1 OVERVIEW

Prior work [3][5] has shown that a noticeable amount of tweets contain implicit mentions to entities; therefore, performing implicit entity linking is expected to offer a considerable amount of value to systems that leverage entity linking as a component. Most notably, implicit entity linking is expected to result in a more precisely targeted information extraction in such applications as, sentiment analysis, trend detection, event monitoring, and news recommendation. As another example, identifying user needs in dialogue systems might require detection of implicit mentions. In this paper, we present our research goals (RGs) and progress so far.

**RG1.** *Implicit Named Entity Recognition and Classification (INERC).* Traditional named entity linking is comprised of two steps: first, named entity recognition and classification (NERC), which involves recognizing the phrase that is probably referring to a named entity and classifying the type of the entity it should be linked to; and second, named entity linking (NEL) which is the act of linking the recognized phrase to a unique entity in a knowledge base, which is usually performed in two steps, i.e., candidate selection and candidate disambiguation. One of the limitations of prior work is that they overlook the first step, i.e., NERC, by assuming that a given tweet contains an implicit entity of a known type. This assumption makes the implicit entity linking step of little practical use to real world problems. Therefore, a research goal is to perform implicit named entity recognition and classification (INERC). A major resource required is a gold standard dataset similar to those provided in traditional NERC; such datasets provide a taxonomy of most prevalent classes of entity mentions as well as instances of those mentions. We aim at providing such a dataset.

**RG2.** *Implicit Named Entity Linking (INEL).* Having recognized implicit mention of an entity, the next step is to link it to the corresponding knowledge base entry. Therefore, linking implicit mentions comprises a research line of this thesis. The authors in [5] have proposed an approach to perform implicit entity linking. Their method is heavily based on exploiting the linked data present in the knowledge graph. While this source of information is invaluable and has proven helpful in previous studies, we aim at exploiting different information resources in order to improve both efficiency as well as effectiveness of the linking process.

**RG3.** *Implicit Named Entity Linking Applications.* The other research direction is to investigate the value-add that INEL provides to systems using named entity recognition. For instance, how can linking implicit mentions of user generated content boost performance of news recommendation systems? This is particularly important as traditional NEL systems fall short in linking implicit mentions, mainly since they rely on recognizing and linking explicit mentions, which have a surface form in text. Another research question is the applicability of the notion of implicit mentions and its pertinent methods to different information extraction schemes and frameworks. For instance, how can we extract structured information from unstructured sources with a high degree of precision?

This proposed PhD research is accomplished in three main phases. In Phase 1, we have accomplished the second part of research goal 1, i.e. preparing a gold standard dataset [3]. Furthermore, we have developed and evaluated an approach in order to accomplish research goal 2 [2]. In phase 2, we aim at performing INERC using the resources provided through Phase 1. In the last phase, we aim at applying the developed models of the previous phases for different purposes as described in research goal 3.

## REFERENCES

[1] Leon Derczynski, Diana Maynard, Giuseppe Rizzo, Marieke van Erp, Genevieve Gorrell, Raphaël Troncy, Johann Petrak, and Kalina Bontcheva. 2015. Analysis of named entity recognition and linking for tweets. *Information Processing & Management* 51, 2 (2015), 32–49.
[2] Hawre Hosseini, Tam T. Nguyen, and Ebrahim Bagheri. 2018. Implicit Entity Linking Through Ad-Hoc Retrieval. In *IEEE/ACM 2018 International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2018, Barcelona, Spain, August 28-31, 2018*. 326–329. https://doi.org/10.1109/ASONAM.2018.8508612
[3] Hawre Hosseini, Tam T. Nguyen, Jimmy Wu, and Ebrahim Bagheri. 2019. Implicit Entity Linking in Tweets: an Ad-hoc Retrieval Approach. *Applied Ontology* (2019).
[4] Xiaohua Liu, Yitong Li, Haocheng Wu, Ming Zhou, Furu Wei, and Yi Lu. 2013. Entity Linking for Tweets. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1304–1311.
[5] Sujan Perera, Pablo N. Mendes, Adarsh Alex, Amit P. Sheth, and Krishnaprasad Thirunarayan. 2016. Implicit Entity Linking in Tweets. In *ESWC'16*. 118–132.
[6] Alan Ritter, Sam Clark, Oren Etzioni, et al. 2011. Named entity recognition in tweets: an experimental study. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 1524–1534.


*SIGIR '19, July 21–25, 2019, Paris, France*

ACM ISBN 978-1-4503-6172-9/19/07…$15.00
https://doi.org/10.1145/3331184.3331416