# Abstract

The Linked Open Data (LOD) cloud is currently a primary source of background knowledge for tasks in a wide variety of domains and across many scientific fields. The structured nature and the usage of well-defined open standards make it convenient to contribute to and build upon. However, since the major part of the LOD is ultimately crowdsourced and mostly populated and updated manually, some of the content in the LOD can become stale, inconsistent and lack coverage. Social media, on the other hand, uniquely allow the real world events to be accurately reflected with little or no delay in the form of posts and profile updates. A major downside of this vibrant source of knowledge that is contained in the social media is its lack of structure, significant noisiness and restrictive APIs that make it hard to extract, analyze and use it in the downstream tasks.

In this thesis, I present the task of linking entities in a knowledge base (KB) to the corresponding social media profiles as an attempt to bridge the structured LOD cloud and the vibrant social media. As will be shown, such linking allows knowledge transfer between the two worlds: on the one hand, enabling the Semantic Web practitioners to harvest this vast amount of valuable, up-to-date data from the social media; on the other hand, the social media researchers can use the structured LOD knowledge much more efficiently, simplifying the pipelines and improving performance for tasks such as Type Prediction, Entity Linking, and User Profiling. I implement such knowledge transfer using DBpedia as a KB, since it is a cornerstone dataset in the LOD, and Twitter as a social media, due to its popularity and relative accessibility. However, approaches developed here are designed to be general and could be applied to other social media and KBs.

To this end, firstly, I introduce SocialLink — a project designed to link KBs to social media profiles. SocialLink consists of (i) a linking approach that is able to produce high-quality entity-profile pairs, (ii) a LOD-compliant dataset of alignments between DBpedia and Twitter, (iii) the Social Media Toolkit (SMT) system providing additional functionality on top of SocialLink. SocialLink employs a custom deep neural network-based architecture designed to efficiently exploit many modalities of data representing entities and profiles within DBpedia and Twitter.

In second, I demonstrate how SocialLink can facilitate tasks in both Semantic Web and Social Media Analysis. In particular, I employ the abovementioned knowledge transfer to achieve state-of-the-art performance in Type Prediction task on DBpedia. Additionally, SocialLink is used to infer user interests on Twitter and to implement a novel approach that I proposed to prevent such inference. Finally, the Entity Linking capabilities of SocialLink are exploited to augment the social media management application called Pokedem and to provide an additional performance boost to a conventional Entity Linking pipeline achieving the second-best performance in EVALITA 2016 competition.

SocialLink[1] and its applications are open source projects with all the code, datasets, tutorials and experimental results available online. The approaches presented in this thesis have been validated with extensive evaluation and covered in a number of publications.

**Keywords**:
Machine Learning, Social Media, Semantic Web, User Profiling, Entity Matching, Deep Learning

[1] http://sociallink.futuro.media/