

Abstract of the Thesis

Social media is ubiquitous. With its huge active user base, it has opened up new ways of acquiring information, creating and sharing ideas, connecting with each other. It has notably changed the way people interact and carry on with their day-to-day activities engaging in the social media platforms. With the evolution of the role of social media in people's lives, the social media entities – hashtags in Twitter, topics in Quora, tags in Stack Overflow etc. which started as a way for people to organize similar contents together, propagate ideas, promote specific people or topics has now grown into a “language” in its own right.

Hashtags are the most prevalent and prominent social media entity with its presence across several social media platforms – Twitter, Facebook, Google+, Instagram, Pinterest etc. We consider two special kinds of hashtags – i) hashtag compound (two or more hashtags coalescing together to form a new one) ii) idiom (conversational themed hashtags) to study their linguistic formation, structure and put forward various socio-linguistic reasons for the adoption (popularity) of these hashtags.

Similar to hashtags, **topics** are also important entities in platforms like Quora. We study the dynamics of topic growth in Quora over time to understand how the Quora knowledge base is changing with the influx of new topics, growth or decay of older topics. We further study the popularity dynamics of topics – their migration across various popularity zone and then develop a framework to early predict the popularity (followership) of the topics. We also study the answerability of questions in Quora specifically investigating characteristics of open questions (questions with no answer) and answered question. We analyze various linguistic activities - the way users use their language and observe how such activities affect answerability of questions in Quora.

Further, we perform in-platform study of popularity of **books** (determined by the votes received in Readers' Choice Award) as well as a cross-platform study of Goodreads entities and attempt to establish the connection between such entities and the Amazon best sellers. We analyze the collective reading behavior on Goodreads platform and quantify various characteristic features of the Goodreads entities to identify differences between the Amazon best sellers and the other non-best selling books.

This thesis attempts to understand the behavior of these varied linguistic entities as they grow popular across these platforms. This investigation allows us to uncover many common linguistic features that seem to propel the growth of popularity of these entities which can be thought of as the central principle laid out in this thesis.