

Towards segmentation of hashtags for Instagram post analysis

Dobroslav Egorov¹, Alexander Komarov¹, Oleg Kravchenko¹

¹Bauman Moscow State Technical University, Moscow, Russia

Nowadays there are many solutions in digital market that provide automated massive interactions with the audience in Instagram. However the ways of picking up the audience are very general and the accuracy of targeting right people might be very low. Therefore it is important to clusterize the audience in Instagram and understand what the users you target write about and in which language. One of the ways to find it out is analyzing the hashtags for users' posts. However there are some pitfalls.

In general, there are two related problems. Disambiguation problem is the first one, and the second is about unknown words handling. The common way to solve the second problem is a hidden Markov model implementation. In present paper to solve the first problem a parallel corpora approach is considered.