Abstract

The number of users of the internet is increasing exponentially every year; most of these users are using social networks, blogs and forums. When users in the Arab world need communicate with each other, they often use their colloquial dialect Arabic instead of the Modern Standard Arabic (MSA). To retrieve information about a specific topic, Modern Standard Arabic (MSA) Informational Retrieval will retrieve only MSA relevant data, but maybe there is helpful information published in Dialect Arabic, also if we need to directly translate from Dialect to English, this may be done by translating the Dialect to standard Arabic then to English.

This thesis aims to build an annotated corpus for Palestinian Dialect with Relevant Meta Data. The proposed methodology includes studying linguistic facts about Palestinian dialect and comparing it with Modern Standard Arabic in terms of morphology, orthography and lexical. As well as collecting Palestinian written text from different resources, then analyzing and annotating the corpus by using resources designed for

Egyptian dialect, after that annotating manually a list of words that can't be analyzed by existing resources, finally start using the existing annotation tool to double check over annotated corpus.