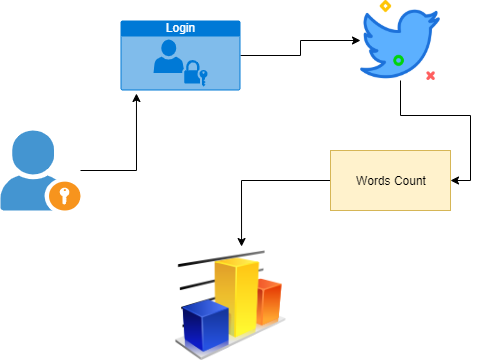
**INTRODUCTION**

Word representation attempts to represent aspects of word meanings. For example, the representation of “cellphone” may capture the facts that cellphones are electronic products, that they include battery and screen, that they can be used to chat with others, and so on. Word representation is a critical component of many natural language processing systems as word is usually the basic computational unit of texts. A straight forward way is to represent each word as aone-hot vector, whose length is vocabulary size and only one dimension is 1, with all others being 0. However, one hot word representation only encodes the indices of words in a vocabulary, but fails to capture rich relational structure of the lexicon. To solve this problem, many studies represent each word as a continuous, low-dimensional and real valued vector, also known as word embeddings. Existing embedding learning approaches are mostly on the basis of distributional hypothesis , which states that the representations of words are reflected by their contexts. As a result, words with similar grammatical usages and semantic meanings, such as “hotel” and “motel”, are mapped into neighboring vectors in the embedding space. Since word embeddings capture semantic similarities between words, they have been leveraged as inputs or extra word features for a variety of natural language processing tasks, including machine translation, syntactic parsing, question answering, discourse parsing, etc.



Architecture