In this thesis, I try to find a definition of a news topic to make topic detection implementable and its quality measurable. I describe various methods — a “simple” words counting, optionally with stopwords. I also describe TF-IDF and the text categorization problem. I touch the subject of text clustering. Then I briefly describe approaches called latent semantic indexing and latent Dirichlet allocation. The thesis includes my experiments with “simple” words counting, TF-IDF and text categorization on database of articles from several online news websites; I also describe the creation of this database. Precision and recall are used as a metric to text categorization approach.