ABSTRACT

*Topic models* are hierarchical Bayesian models used to discover latent semantic structure within collections of documents, allowing them to be reduced from millions of words to a few dozen interpretable topics. This paper presents three closely related methods: latent Dirichlet allocation, correlated topic models, and structural topic models. I discuss the estimation challenges associated with topic modeling and compare the three methods by analyzing a collection of 182,308 posts contributed by the general public to the statistics and machine learning community website [stats.stackexchange.com](http://stats.stackexchange.com).