**Main points**

* Many subject based internet directories are present on the internet
* The websites are classified in various hierarchical categories. The proper classification of websites have made these web directories immensely popular and quite important to
* As per a recent survey, it had been estimated that presence of 584 million websites on the internet. But these internet directories have only a few percent of website listed with them. This is mainly due to the fact that manually or semi-automatic classification of website is a tedious and costly affair
* In this paper we present a fast, effective and much more accurate classification algorithm for classification of websites.
* A presence of an automatic website classifier based on machine learning and data mining techniques is therefore an excellent alternative for manual classification of data.
* Maintain directory services automatically
* Analysis of webpage is quite different from analysis of text document
* Information contained in various tags (meta ,title , ahref) have a significant impact on the website categorization
* In this paper we present a unique approach to classify websites based on the content of their webpages using a probabilistic approach.