写论文总结

万维网(WWW)缺乏-个全面的网站目录。因此,使用机器学习技术进行网站分类是一种新兴的可能性。在本论文中使用朴素贝叶斯机器学习算法，根据网站主页的内容对网站进行分类。

为了从数以百万计的网站上找到信息，互联网用户使用各种搜索工具，大致分为: 1.基于爬虫的搜索引擎(SE)，如谷歌，必应，雅虎等2.主题目录，如DMOZ (目 录Mozilla)，图书管理员互联网索引(LI) ete。 和3.元搜索引擎，如Metacrawler，Clusty等。

NB方法是文本分类中最有效、最直接的方法之一,在以往的数据挖掘研究中取得了很好的效果。