

哈哈哈哈哈

Web 数据挖掘过程中的数据预处理

数据的预处理是对 Web 上的数据检索后进行的数

据预处理，为数据挖掘模块提供挖掘所需要的数据。确定挖掘主题后，可使用 Google 的 Web API，利用 JBuilder 实现对 Google 的巨

大 Web 索引的搜索。但是，用户的一个查询请求往往会检索出庞大的结果集，而用户所需要的只是其中很小一部分，面对如此多的结果，用户仍然不知所措。所以必须用 HITS 算法来确定权威 Web 页面，这样就可以有效地去除无效网页。

主特征向量 e_1' 进行归一化；（7）将经过归一化后的特征向量 e_1' 中具有较大绝对值的元素作为 hubs 返回 [3]。