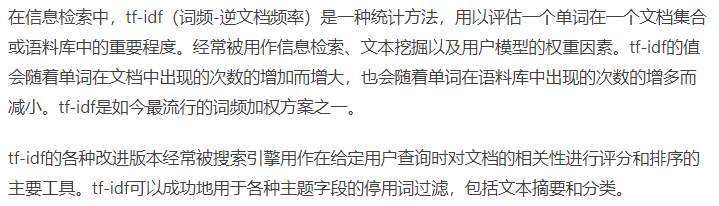
# 原理阶段

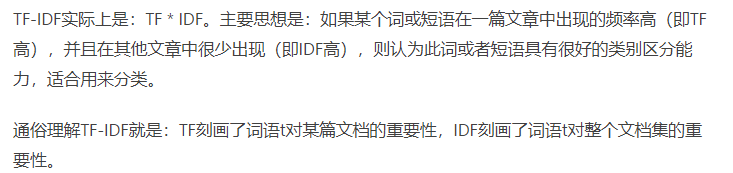
1. **一个问题？**

有一篇很长的文章，用计算机提取它的关键词（Automatic Keyphrase extraction），完全不加以人工干预，请问怎样才能正确做到？

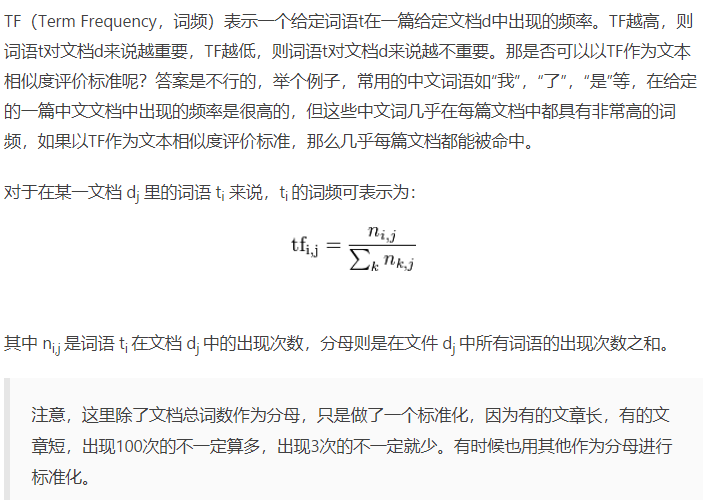
1. **Tfidf定义**



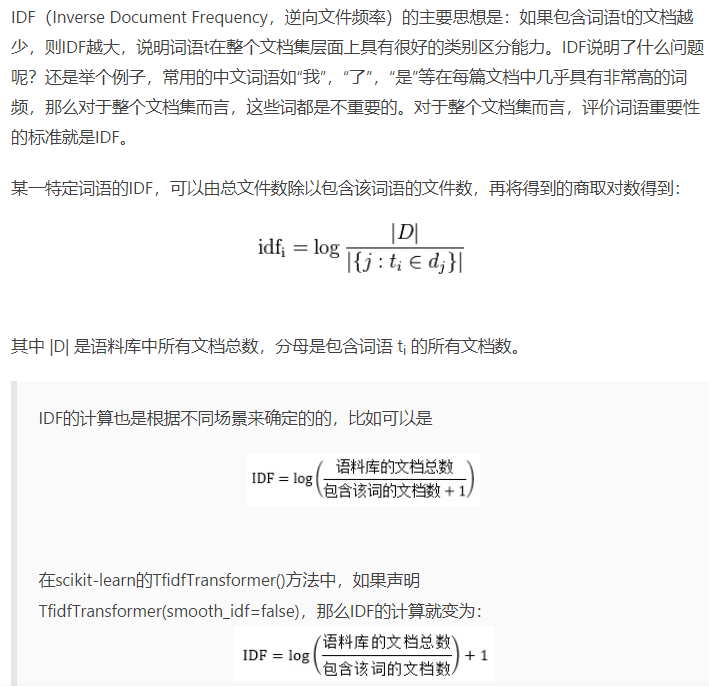
1. **Tfidf原理简述**

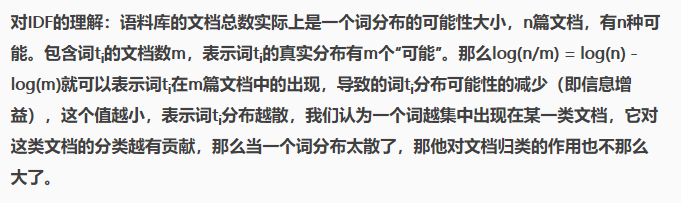


1. **TF（Term Frequency，词频）**

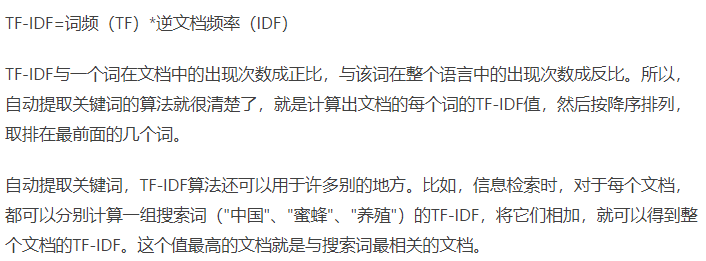


1. **IDF（Inverse Document Frequency，逆向文件频率）**

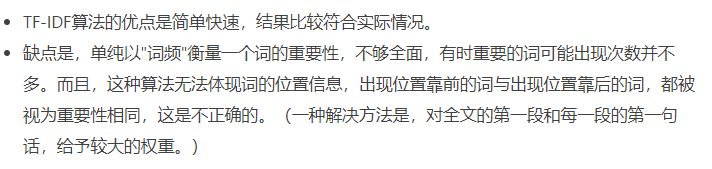




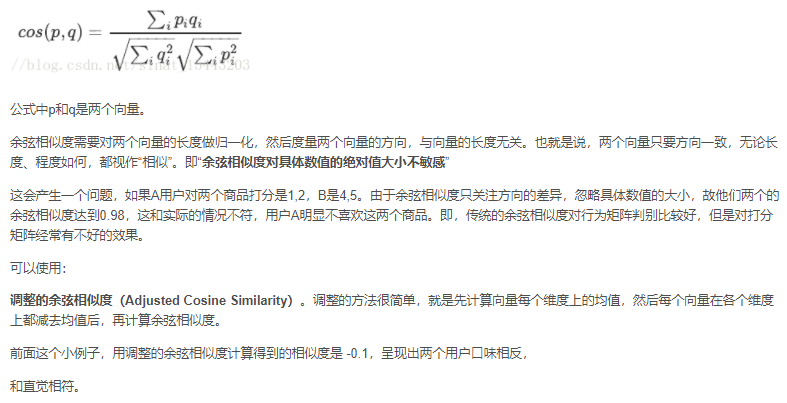
1. **计算TF-IDF**



1. **TF-IDF 优缺点**



1. **余弦相似度**



1. 一个小案例



