Mode:众数

Cross tabulations:交叉表

Confusion Matrix：混淆矩阵，也称为是误差矩阵

Multi-disciplinary field：多学科领域

Linear algebra：线性规划

Reinforced learning： 加强学习，监督学习和非监督学习的总和

Logarithmic：对数的

Tune：调整，调谐

摩尔定律：摩尔定律是由英特尔（Intel）创始人之一戈登·摩尔（Gordon Moore）提出来的。其内容为：当价格不变时，集成电路上可容纳的元器件的数目，约每隔18-24个月便会增加一倍，性能也将提升一倍。换言之，每一美元所能买到的电脑性能，将每隔18-24个月翻一倍以上。这一定律揭示了信息技术进步的速度。

尽管这种趋势已经持续了超过半个世纪，摩尔定律仍应该被认为是观测或推测，而不是一个物理或自然法。预计定律将持续到至少2015年或2020年 [1] 。然而，2010年国际半导体技术发展路线图的更新增长已经放缓在2013年年底，之后的时间里晶体管数量密度预计只会每三年翻一番。 [1]

Hadoop分布式文件系统（HDFS），Hadoop Distributed File System。

计算密集型：computationally intensive

数据密集型：data intensive

结构化数据：结构化数据也称作行数据，是由二维表结构来逻辑表达和实现的数据，严格地遵循数据格式与长度规范，主要通过关系型数据库进行存储和管理。

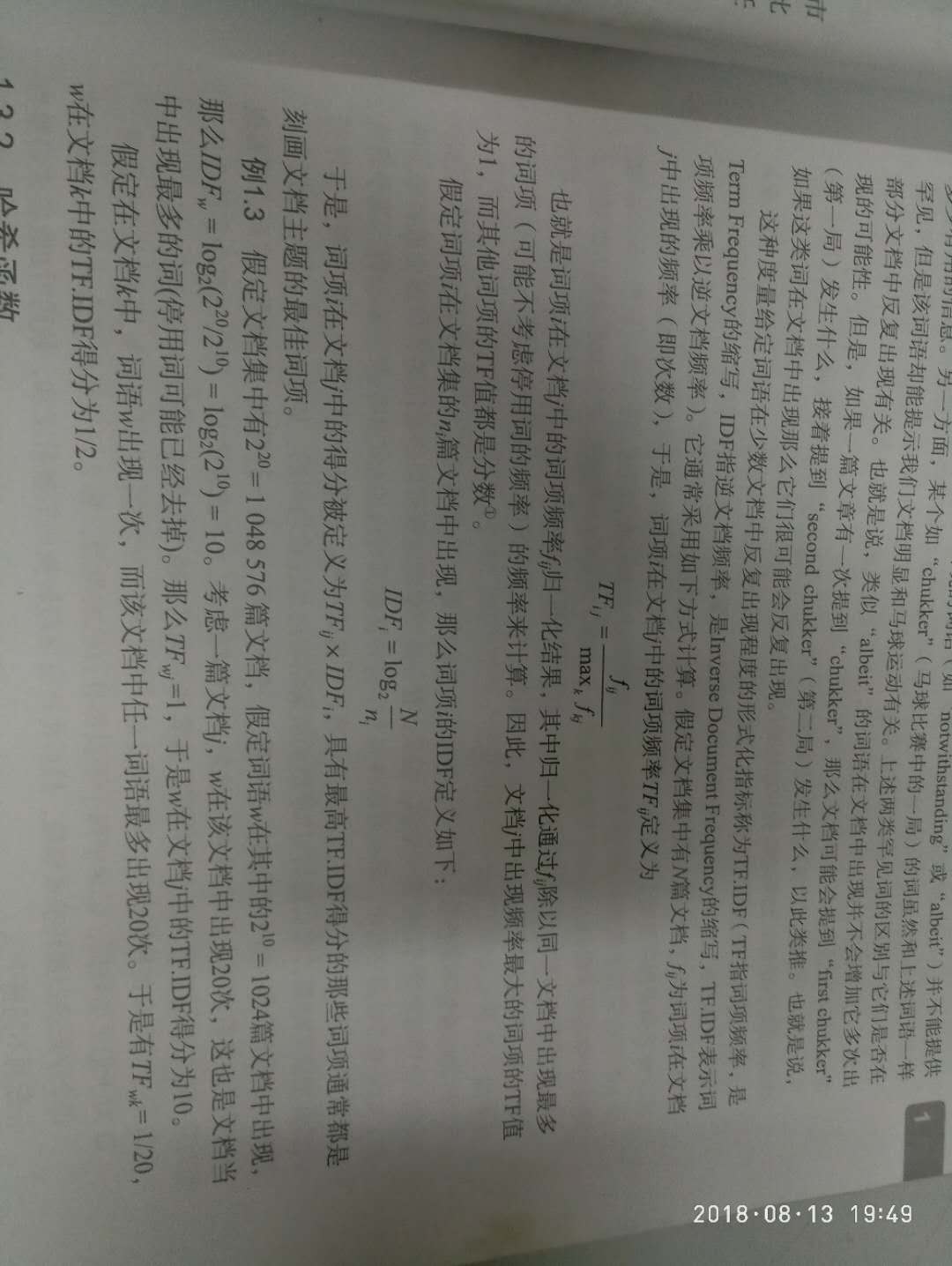
驼峰式大小写：小驼峰式命名法（lower camel case），第一个单词以小写字母开始，第二个单词的首字母大写；大驼峰式命名法（upper camel case），每一个单词的首字母都采用大写字母

Bigram：连续的两组词

Trigram：连续的三组词

数据挖掘：data mining

数据汇总形式：PageRank，聚类

度量给定词语在少数文档中反复出现程度的形式化指标成为TF.IDF（TF指词项频率，Term Frequency，IDF是逆文档频率，是Inverse Document Frequency的缩写，TF.IDF表示词项频率乘以逆文档频率）

JPEG（Joint Photographic Experts Group），联合图像专家小组

QPS（Query Per Second），每秒查询量

JSON(JavaScript Object Notation, JS 对象简谱) 是一种轻量级的数据交换格式。它基于 ECMAScript (欧洲计算机协会制定的js规范)的一个子集，采用完全独立于编程语言的文本格式来存储和表示数据。简洁和清晰的层次结构使得 JSON 成为理想的数据交换语言。 易于人阅读和编写，同时也易于机器解析和生成，并有效地提升网络传输效率。中文名 JavaScript 对象简谱

Django是一个开放源代码的Web应用框架，由Python写成。采用了MVC的框架模式，即模型M，视图V和控制器C。它最初是被开发来用于管理劳伦斯出版集团旗下的一些以新闻内容为主的网站的，即是CMS（内容管理系统）软件。并于2005年7月在BSD许可证下发布。这套框架是以比利时的吉普赛爵士吉他手Django Reinhardt来命名的。

Gephi是一款开源免费跨平台基于JVM的复杂网络分析软件,，其主要用于各种网络和复杂系统，动态和分层图的交互可视化与探测开源工具。