Atitit 大数据体系图

大数据 技术

* 数据采集
  + gui自动化
  + 爬虫
  + Nui自动化
  + Ocr技术
  + Tts语音处理
  + 文档处理（office zip等）
  + html文档处理解析
* 转换与处理识别 挖掘等
  + 图像处理
  + 图像内容识别（人脸识别等
  + 数据分类
  + 图像分类等
  + tag标注
  + nlp自然语言处理（包括分词等）
  + 多线程与分布式mapreduce模型等
* 存储技术
  + 分布式文件存储（ftp hdfs nfs webdav等）
  + 数据压缩 压缩文件等
  + 索引（全文索引，btree，hash索引等）
  + Nosql mongodb redis zookeeper memcache等
  + hadoop系列（hadoop hdfs hive spark等）
* 搜索与索引
  + solr Elasticsearch es
  + elk lucence
* 算法系列
  + map reduce filter foreach系列
  + jion算法系列
  + groupby 分组聚合系列
  + dsl系列（sql linq）
* 可视化