2018-10-09

项目名称： 人工智能在内容处理领域的应用

项目简介：

应用自然语言处理相关技术实现文本和图片的分类/聚类、信息抽取等具体任务；研究深度学习技术在NLP中的应用；利用模式识别、机器学习、自然语言处理等相关技术将研究对象转化为原型demo；探索自然语言处理技术在内容领域中的创新应用。

待开发的文本处理工具集合：

（第一阶段 / 优先级排序）

* PDF文件图片的批量提取 （PNG / JPG）
* PDF批量解除密码，并删除 PDF水印对象，形成干净的 PDF文件 （提供示例）
* PDF读取页码、机构名称、日期等信息后自动更名
* PDF文件目录索引的批量提取 ( TXT )
* PDF全文件文本的批量提取 （TXT）
* PDF 批量加密，加水印，防止第三方篡改