## 《信息与知识获取》课程作业要求与评分标准

2020-2021学年春季《信息与知识获取》课程总成绩包括3次作业成绩，其中作业1的成绩作为期中成绩，作业2+作业3的成绩作为期末成绩，总成绩就是期中成绩（0.2）与期末成绩（0.8）的加权分。考虑到每次作业工作量与时间要求，都是学生自主组建作业小组来完成，小组成员可以是1-3人。默认情况下，组内成员分数相同，如果组内同学共同认为分数需要有差异，需要在报告中具体说明。

#### 打分方法：

每次作业成绩根据各组最终提交的作业报告与程序代码的完成情况，由助教和老师共同打分。

#### 作业要求与评分标准如下：

作业1要求：

以组为单位，进行主题文献调研，主题是“智能手机传感器的国内外发展情况如何？”。目前，智能手机中内嵌了很多种类的传感器，包括图形、图像、声音、触摸、运动、文字感知等等，请首先全面了解各类传感器的基本功能，然后，自主选择其中至少一类传感器作为主要调研方向，对比分析其与人类感知的关系，了解其基本工作原理，并对于该类传感器的国内外研究与应用的发展情况进行较为深入的调研、分析和总结。调研报告内容应包括传感器对环境和社会可持续发展的影响的分析。鼓励发挥各位的创新型思维能力，对未来可能的发展方向给出建议或预测。撰写调研报告。报告格式自拟，请务必在显著位置写明组内成员的姓名、学号与分工完成情况。字数不低于1500字。

评分标准：

完成调研报告，内容覆盖上述全部要求，可得到80分。在此基础上，调研范围广，分析透彻，条理清晰，有自己对于问题的发现、思考与分析，有创新性的探讨，报告撰写认真，文笔流畅，表达清楚，文字准确，图文并茂，酌情加1-20分。

提交时间：2021年3月21号

提交文件命名要求：组长姓名-成员1姓名-成员2姓名

提交方式：所有提交材料压缩成一个文件提交，提交到295459877@qq.com

邮件主题：信息与知识获取作业1

作业2：

基本要求：自己动手设计实现一个信息检索系统，中、英文皆可，数据源可以自选，数据通过开源的网络爬虫获取，规模不低于100篇文档，进行本地存储。中文可以分词（可用开源代码），也可以不分词，直接使用字作为基本单元。英文可以直接通过空格分隔。构建基本的倒排索引文件。实现基本的向量空间检索模型的匹配算法。用户查询输入可以是自然语言字串，查询结果输出按相关度从大到小排序，列出相关度、题目、主要匹配内容、URL、日期等信息。最好能对检索结果的准确率进行人工评价。界面不做强制要求，可以是命令行，也可以是可操作的界面。提交作业报告和源代码。

扩展要求：鼓励有兴趣和有能力的同学积极尝试多媒体信息检索以及优化各模块算法，也可关注各类相关竞赛。自主开展相关文献调研与分析，完成算法评估、优化、论证创新点的过程。

评分标准如下（按照100分计算）：

1. 完成基本的信息检索功能且有对环境和社会可持续发展影响的考虑，系统能够正常运行，并提交源代码和实验报告：60分；
2. 完成要求的信息检索功能且有对环境和社会可持续发展影响的考虑，系统能够正常运行，并按时提交源代码和实验报告：61~70分；
3. 在2的基础上，且实验报告撰写认真、思路清晰、表达准确：71~80分;
4. 在3的基础上，支持检索结果准确率人工评价：81~90分；
5. 在4的基础上，融入了自己的创新性思考或对多媒体信息检索进行了尝试：91-100分。

提交时间：2021年6月15日之前

提交文件命名要求：组长姓名-成员1姓名-成员2姓名

提交方式：所有提交材料压缩成一个文件提交，提交到295459877@qq.com

邮件主题：信息与知识获取作业2

作业3：

基本要求：自己动手设计实现一个信息抽取实验系统，中、英文皆可，可以在作业2信息检索系统的基础上实现，也可以单独实现。特定领域语料根据自己的兴趣选定，规模不低于100篇文档，进行本地存储。对自己感兴趣的特定信息点进行抽取，并将结果展示出来。其中，特定信息点的个数不低于5个。可以调用开源的中英文自然语言处理基本模块，如分句、分词、命名实体识别、句法分析。信息抽取算法可以根据自己的兴趣选择，至少实现正则表达式匹配算法的特定信息点抽取。最好能对抽取结果的准确率进行人工评价。界面不作强制要求，可以是命令行，也可以是可操作的界面。提交作业报告和源代码。鼓励有兴趣和有能力的同学积极尝试优化各模块算法，也可关注各类相关竞赛。

扩展要求：鼓励有兴趣和有能力的同学积极尝试多媒体信息抽取以及优化各模块算法，也可关注各类相关竞赛。自主开展相关文献调研与分析，完成算法评估、优化、论证创新点的过程。

评分标准如下（按照100分计算）：

1. 完成基本的信息抽取功能且有对环境和社会可持续发展影响的考虑，系统能够正常运行，并提交源代码和实验报告：60分；
2. 完成要求的信息抽取功能且有对环境和社会可持续发展影响的考虑，系统能够正常运行，并按时提交源代码和实验报告：61~70分；
3. 在2的基础上，且实验报告撰写认真、思路清晰、表达准确：71~80分;
4. 在3的基础上，支持抽取结果准确率人工评价：81~90分；
5. 在4的基础上，融入了自己的创新性思考或对多媒体信息抽取进行了尝试：91-100分。

提交时间：2021年6月15日之前

提交文件命名要求：组长姓名-成员1姓名-成员2姓名

提交方式：所有提交材料压缩成一个文件提交，提交到295459877@qq.com

邮件主题：信息与知识获取作业3