Minutes of Meeting

2017/7/18-电话会议

Recorder:陈栋

Participant：黄凯波（老师）、陈栋、闵志强、杜雨霏、李宜瑾

I Meeting Content

1. 总结小组上周的工作进展，具体如下：
2. 将上周爬取的公司节点和关系边插入到了OrientDb数据库中。
3. 搜寻细化公司信息（Email,Address,Other Information）

问题：没有自动化，纯手工的操作。这样是非常浪费时间和精力的，以后需要脚本化、自动化。

1. 黄老师对下一阶段的任务进行说明：
2. 期刊信息的搜集：

首先，搜集农业研究领域的分类；

其次，分别对每一个分类查找期刊3-4个；

最后，构建农业期刊图库的数据库结构并形成文档（加上之前的公司内容）提交至git；并对期刊数据进行爬取。并将摘要中不是农业的词汇去掉。

1. 学习Python Flask框架，尽快应用。
2. 了解熟悉单元测试内容，需要了解为什么这样做？其不合理的地方在哪里？
3. 黄老师提出了依托数据派通过北京林业大学的社团、研究生会及学生会搞大数据创新实战活动的构想。作为远期目标。
4. 黄老师讲解了OrientDB和Neo4j的区别和利弊。选择使用OrientDB的原因是由于OrientDB不单纯是图数据库，而且还有管理文档的功能，另外其天生对分布式支持，并且更新很快。Neo4j是一个纯图数据库。
5. 明确了小组讨论及汇报时间，暂定为一周一次，可以通过会面、视频、音频等方式进行。

II Plan

1. 期刊信息的搜集，搜集农业研究领域分类、分别对每一个分类查找期刊3-4个，并构建数据库结构。

**执行人：**全体小组成员

**时间：**1周（7月18日-7月25日）

1. 爬取期刊数据，并入库。

**执行人：**全体小组成员

**时间：**1周（7月25日-8月1日）

1. 学习Python Flask框架，要求看懂框架，并自己完成一个Demo（连接图库，实现一个动态网页），组织汇报并在组内共享学习。

**执行人：**闵志强、陈栋

**时间：**1周（7月18日—7月25日）

1. 学习单元测试相关概念及内容，并会使用一个单元测试工具或代码。

**执行人：**杜雨霏、李宜瑾

**时间：**1周（7月18日—7月25日）

1. google API文档的提供+公司信息获取渠道的搜集+微软文档共享平台的替换+数据库脚本的提交。

**执行人：**黄老师

**时间：**尽快