

2020 年夏季 Java 小学期大作业：QUBYNews

实验报告

2018011365 张鹤潇

2018011364 项晓宁

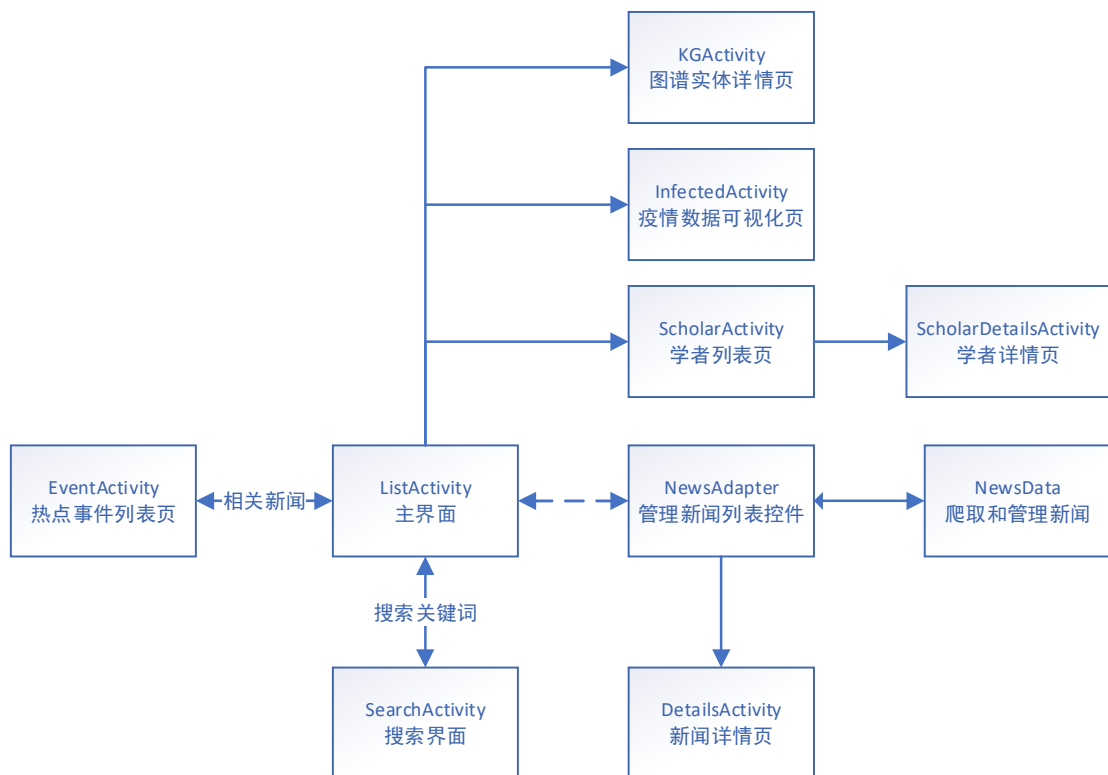
本项目实现了大作业要求中的所有**基础功能**和**扩展功能**，UI 设计简洁、美观、流畅，易用性强。

开发环境

Android Studio 4.0.1 Gradle 6.1.1 Java 1.8.0_242

Compiler SDK 30 (API 30 Android 10.0+)

1 代码结构



每个源代码文件的功能如下：

ListActivity：主界面，启动事件；

NewsAdapter：调用 **NewsData** 获得新闻数据，管理 **ListActivity** 的新闻列表控件；

NewsData: 爬取, 管理新闻数据;

DetailsActivity: 新闻详情页;

SelectActivity: 新闻搜索页;

EventActivity: 热点事件列表页;

InfectedActivity: 疫情数据可视化页;

ScholarActivity: 学者列表页;

ScholarDetailsActivity: 学者详情页;

KGActivity: 知识图谱实体详情页;

ParcelHelper: 辅助类, 解决与 parcel 相关的一些问题;

Util: 提供一些帮助函数。

分工如下:

功能	子功能	分工
系统支持	在安卓机上正常运行, 不崩溃	两人
页面布局	布局合理, 点击处理正确	张鹤潇
分类列表	删除和添加分类	张鹤潇
新闻列表	正确显示新闻列表的消息, 布局和展示, 点击进入新闻详情页面。	两人
	实现新闻的本地存储, 看过的新闻列表在离线的环境下也可以浏览。新闻是否看过的页面灰色标记。	
	上拉获取更多新闻, 下拉刷新最新新闻。	
	显示新闻的来源和时间	
	根据关键词在线搜索新闻, 搜索历史记录	
分享收藏	使用 QQ、微博 SDK 分享, 新闻详情页面点击分享可以分享到常用的 app, 分享内容带有新闻摘要	张鹤潇
疫情数据	全球各国疫情数据可视化, 以折线图形式呈现	张鹤潇
疫情图谱	对新冠疫情图谱内容进行查询和展示, 展示实体词条的描述, 关系和属性	张鹤潇
疫情新闻聚类	对疫情相关的新闻事件进行聚类、展示关键词和聚类新闻	两人
知疫学者	显示在疫情领域的高关注学者和追忆学者, 点开可以查看学者详细介绍	张鹤潇

报告：张鹤潇

展示视频：项晓宁

2 具体实现

本项目以 Java 和 Kotlin 两种语言混合编程实现。Kotlin 作为改良的 Java，是 Google 大力推广的安卓开发语言，它在 Java 原有 API 的基础上提供了更好的空类型管理、函数式编程、异步编程和数据类等功能，大大简化了 Java 繁琐的语法。本项目开发者自学了 Kotlin 语言，并将其应用到开发中。

对部分功能，特别介绍其实现细节。

功能	子功能	细节
分类列表	删除和添加分类	以内嵌 Spinner 的 AlertDialog 实现，并对 AlertDialog 加入淡入淡出动画。
新闻列表	正确显示新闻列表的消息，布局和展示，点击进入新闻详情页面。	新闻列表控件基于开源库 UltimateRecyclerView，以开源库 fuel 实现 http get，在运行过程中异步地加载新闻数据。
	实现新闻的本地存储，看过的新闻列表在离线的情况下也可以浏览。新闻是否看过的页面灰色标记。	已读新闻以 Parcel 格式存储。
	上拉获取更多新闻，下拉刷新最新新闻。	以控件自带 Refresh 动画作为上拉、下拉时的特效。
	根据关键词在线搜索新闻，搜索历史记录	搜索历史通过 SharedPreferences 存储。
分享收藏	使用 QQ、微博 SDK 分享，新闻详情页面点击分享可以分享到常用的 app，分享内容带有新闻摘要	以 QQ 分享替代微信分享功能。
疫情数据	全球各国疫情数据可视化，以折线图形式呈现	以开源库 MPAndroidChart 绘制折线图。
疫情新闻聚类	对疫情相关的新闻事件进行聚类、展示关键词和聚类新闻	以事件文本的 TF-IDF 为特征，用 K-MEANS 方法进行聚类，人为设定聚类数量为 25。

3 总结与心得

本次大作业充分锻炼了我们解决问题的能力。让我印象最深的一个 bug，是刷新新闻时

Listener 函数的返回和与新闻的更新不同步。我们检查了很久才意识到，http Get 是异步执行的，这两个事件不在同一个线程中。这加深了我对异步编程的理解。

本次大作业也充分锻炼了我们的工程能力：我们用 git 进行版本控制和协同开发，这对未来的课程大有帮助；我们自学了 Kotlin 语言，学会了几个第三方库的使用方式，这拓展了我们的技能树。

聚类本身是一件主观性非常强的事：没有什么方法能客观判定新闻聚类结果的好坏。基于此考虑，本项目采用了 TF-IDF+K-MEANS 的聚类算法，而没有选用其它更复杂的算法。我认为将聚类算法的复杂程度与大作业得分挂钩是非常不合理的，这不仅难以带来实用价值，还与本门课的主题相悖。望助教慎重考虑！