周报 1（12.2~12.15）

本组课题为基于 Spark 的景点推荐系统，截至开题时，我们完成了对“去哪儿网”景点数据的爬取。老师对数据来源给出了建议，于是这两周，我们对“携程”的景点数据也进行了爬取，丰富了数据源。

遇到的问题：在大规模爬取数据时，我们发现“去哪儿网”的景点页数若超过某个页数，就会一直重复某一页的景点内容。