期末大作业

Compute the PageRank scores on the Wikipedia dataset

Dataset: WikiData.txt

The format of the lines in the file is as follow:

FromNodeID ToNodeID

数据集：WikiData.txt

文件中各行的格式如下：FromNodeID ToNodeID

在此项目中，您需要报告前100个NodeID及其PageRank得分。您可以选择不同的参数（例如，传送参数）来比较不同的结果。您必须报告的一项结果是，将teleport参数设置为0.85时。除了基本的PageRank算法外，您还需要实现Block-Stripe Update算法。

In this project, you need to report the Top 100 NodeID with their PageRank scores. You can choose different parameters, such as the teleport parameter, to compare different results. One result you must report is that when setting the teleport parameter to 0.85.

In addition to the basic PageRank algorithm, you need to implement the Block-Stripe Update algorithm.

期末大作业

作业要求：

1. 语言：C/C++/JAVA/Python
2. 考虑dead ends 和spider trap 节点
3. 优化稀疏矩阵
4. 实现分块计算
5. 程序需要迭代至收敛
6. 不可直接调接口，例如实现pagerank时，调用Python的networkx包
7. 结果格式(.txt文件)： [NodeID] [Score]

截止日期：

2020年5月1日24点

期末大作业

作业提交：

1. 作业发送至邮箱：webbigdata@163.com
2. 实验报告内容（包括但不限于）：数据集说明、关键代码细节、云主机运行截图、实验结果、结果分析等
3. 程序源码。C，C++，JAVA，Python，四者之一。无需提交数据集，但是需要在实验报告中说明数据调用位置。
4. 程序执行结果文件。必须与实验报告中讨论的内容一致，严格按照作业要求生成结果文件。格式不一致将无法进行正确率的计算，影响最终成绩。
5. 可执行文件。请说明具体运行方式，提交前请确认在其他电脑上也可以运行。C/C++ 编译选择release 方式进行静态编译（debug方式生成的exe文件在其他电脑上运行可能会有缺少dll文件等问题）。JAVA 和Python 请借助第三方软件生成exe可执行文件，并集成相关依赖包。
6. 上传文件命名:

   学号1+姓名1+学号2+姓名2（同一组的同学上传相同的文件）例: 00000_李明_11111_王芳
7. 各小组独立完成，请勿抄袭

期末大作业

评分标准：

1. 晚交一天，从本次作业成绩（score）中扣2分。5天以上，score=（score-10）*0.8。晚交10天以上， score=（score-10）*0.6。作业以学院网站上提交或发送到邮箱(webbigdata@163.com)时间为准。
2. 得分点：

    实验报告+源码+可执行文件+结果文件 50分

    结果是否正确 20分

    是否考虑dead ends 和spider trap 节点 10分

    是否优化稀疏矩阵 10分

    是否实现分块矩阵 10分

如有疑问，请发邮件至 webbigdata@163.com