Introduction to elgooG Search Engine

Zhipeng Chen

July 20, 2020

Initial interface

Result interface

How to make it

(1) Crawler

(2) Normalized and divided

(3) Build the inverted index

(4) Solve queries

Optimization

(1) The value of word is related to the times that the word appears in website, and it also related to the lenght of the word.

$$Value_{i,id} = (1 + \lg(t_{i,id})) * (\lg(\frac{N}{n_{id}}))^{1.75} * (\lg(len_{id} + 1))$$

(2) The value of word is related to whether it is a number.

$$Value_{i,id} = (Value_{i,id})^{'} * (\lg(len_{id} + 1) + 1.6)$$

Optimization

(3) The sorce of website is related to the lenght.

$$Lenght_i = \sqrt{\sum (\lg t_{i,id})^2}$$

$$Sorce_i = Sorce_i^{'} / lenght_i^{0.495}$$

Optimization

(4) Different websites have different weight.

If the page is about prefessor:

$$Sorce_i = Sorce_i^{'} * 1.6$$

If the page is about laboratory or the overview of School of Information:

$$Sorce_i = Sorce_i^{'} * 1.2$$

If the page is about news or notice:

$$Sorce_i = Sorce_i^{'} / 1.2$$

Example-1

slpooG

数据工程与知识工程教育部重点实验室   Search

数据工程与知识工程教育部重点实验室 - 机构设置 - 中国人民大学信息学院

... 数据工程与知识工程教育部重点实验室成立于2006年。是"985工程"二期"数据工程与知识工程"科技创新平台承担单位，中国人民大学建设的第一个省部级重点实验室，现任主任为教育部科技进步一等奖获得者杜小勇教授。实验室以中国人民大学的学科优势为依托，以计算机科学与技术、图书情报与档案管理取得一级学科博士学...

2016年金融信息学研讨会通知 - 师资科研 - 中国人民大学信息学院

发布时间: 2016-11-29 02:27:28 点击: 中国科学院数字与系统科学研究院

... 中国科学院管理、决策与信息系统重点实验室(MADIS)上海市智能信息处理重点实验室(IIPL)教育部数据工程与知识工程重点实验室(DEKE) 日程：报告摘要详见附件。 附件：2016年金融信息学研讨会议安排请大家携自己笔记 光 ...

杜小勇 - 师资科研 - 中国人民大学信息学院

... 大学)，理工处处长、教育部数据工程与知识工程重点实验室主任。中国人民大学信息学院计算机应用技术专业退职教授，博士生导师。教育经历 1983 年毕业于原杭州大学数学系并留校任教，1985-1988 年在中国人民大学信息学院计算机应用专业攻读硕士学位，毕业后晋级攻读数据与知识工程研究所副所长。1993 年获日本政府奖 ...

张缘 - 师资科研 - 中国人民大学信息学院

... 同年加入中国人民大学数据工程与知识工程教育部重点实验室，研究方向包括大数据系统、高性能计算、异构数据库。担任2019级图灵试验班主任。教育经历 2012年-2017年，清华大学，计算机科学与技术系，博士； 2008年-2012年，西安电子科技大学，计算机学院，本科工作经历2017年9月-2019年6月，中国人民大学 ...

张新勇 - 师资科研 - 中国人民大学信息学院

... 院副教授，"感知计算与用户研究"实验室负责人，信息系研究生工委。教育经历 1993.09—1997.07：南京林业大学，林业管理，农学学士； 2002.09—2005.03：北京理工大学，系统工程，工学硕士； 2005.04—2008.03：清华工科大学，基础工学（人机交互），工学博士。工作经历 5、1997 ...

数据仓库与商务智能实验室 - 中国人民大学信息学院

... 数据仓库与商务智能实验室成立于2002年。实验室负责人为陈红教授。该实验室附于数据工程与知识工程教育部重点实验室和中国人民大学信息学院计算机系。实验室现有教师4人（其中教授2人，讲师2人），在读博士生12人、硕士生18人。实验室分为三个研究小组：(1) 数据挖掘；(2) 面向现代硬件平台的数据仓库 ...

云计算与大数据实验室 - 中国人民大学信息学院

..."云计算与大数据实验室"是由信息系授权带领的，致力于云计算、非结构化数据、海量数据、数据库等方向研究的团队，隶属于数据工程与知识工程教育部重点实验室（DEKE）和信息学院计算机系。在随着规模数的领域下，实验室既存在数据库和云计算领域都取得了主要的成果。除教育2011-2015连续三年在本国际数据库顶级会议AC ...

Example-2

Example-3

elpooG
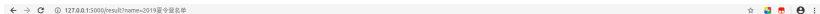
2019夏令营名单 　　Search

2019年中国人民大学信息学院优秀大学生夏令营营员名单 - 学院公告 - 中国人民大学信息学院

发布时间：2019-08-21 17:55 ;来源：教务科

... 以下是2019年信息学院优秀大学生夏令营营员名单，请确定来参加夏令营的同学于2019年6月26日前发邮件到2014103596@ruc.edu.cn确认参加本次夏令营，若26号前没有收到确认邮件，则视同放弃本次营员资格。同学们发邮件时，请务必按此署名（邮件主题：姓名+所申请的专业+参加信息学院夏令营+直博 ...

李锡荣 - 师资科研 - 中国人民大学信息学院

... 术报告； + 2019.11 Learn to Represent Queries and Videos for Ad-hoc Video Search, TRECVID 2019 workshop, Gaithersburg + 2019.10 Deep Learning for Video Retrieval ...

2018-2019学年国家励志奖学金增补公示 - 学院公告 - 中国人民大学信息学院

发布时间：2019-09-29 16:47 ;来源：党团学办公室

... 出，现增补一名同学，名单如下： 对上述名单如有意见，请于10月8日前致扣762513007或发送邮件至dszc@ruc.edu.cn。 ...

2019年中国人民大学信息学院优秀大学生夏令营获奖名单 - 学院公告 - 中国人民大学信息学院

发布时间：2019-07-12 19:25 ;来源：教务科

... 2019年7月5日至7日，来自全国各高校的优秀大学生与中国人民大学信息学院的硕士导师和博士导师们进行了面对面交流，同学们充分了解了学院各个院系团队的研究方向和科研情况。学院通过对参营学生的全面考核与选拔，评选出以下夏令营奖学金。 我院拟向学校优先推荐以上获得一等奖的同学，其次再推荐获得二等 ...

**2018-2019学年文明宿舍推荐评审结果公示 - 学院公告 - 中国人民大学信息学院**

发布时间：2019-11-17 23:21 ;来源：党团学办公室

... 根据学校2019年度，优秀宿舍长评审工作安排及2018-2019学年学校、学院宿舍营造成绩，产生我院推荐文明宿舍名单。现公示如下： 对上述名单如有意见建议，请于11月20日前发送邮件至 xiaotongwang@ruc.edu.cn。 ...

2019年中国人民大学信息学院 全国优秀大学生夏令营入营须知 - 学院公告 - 中国人民大学信息学院

发布时间：2019-06-24 11:25 ;来源：教务科

... 的同学： 欢迎你参加2019年中国人民大学信息学院全国优秀大学生夏令营！以下是有关夏令营的一些事项说明： 1、夏令营举办时间：2019年7月5日—7月7日 2、报到时间：2019年7月5日下午2：30—5：00；3、报到地点：中国人民大学信息楼一层大厅；报到时请携带本人学生证、身份证；本校学生也必须办理报到，签到等 ...

Example-3
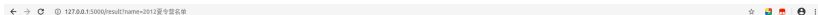
elpooG | 2012夏令营名单 | Search

2012信息学院优秀大学生夏令营录取我院学生名单 - 学院公告 - 中国人民大学信息学院
发布时间：2012-06-29 11:26
姓名 冯雪

2012信息学院优秀大学生夏令营录取学生名单（第二批，外校生）- 学院公告 - 中国人民大学信息学院
发布时间：2012-07-02 16:32
...2012信息学院优秀大学生夏令营录取学生名单（第二批，外校生）以下是录取参加2012信息学院优秀大学生夏令营的外校生名单，请以下同学于7月5日前发邮件到panxuyan@ruc.edu.cn确认是否参加本次夏令营。发邮件时，务必署著名（姓名，及所申请的专业）。赵珂 崔欣茹 王鑫 郭雅楠 郑 ...

范举 - 师资科研科 - 中国人民大学信息学院
一年生，博士，副教授。2012年7月于清华大学计算机系获得博士学位，2012年至2015年在新加坡国立大学从事博士后研究工作，现为中国人民大学数据工程与知识工程教育部重点实验室副教授。更多信息，请参见个人主页：http://iir.ruc.edu.cn/~fanj/ 教育经历 2007-2012 清华大学计算机系 博 ...

2012年Google优秀奖学金初选名单 - 人才培养 - 中国人民大学信息学院
发布时间：2012-04-05 14:12
本科生：1.谢怡然 2.曹博硕 3.李梅英 4.侯日佳 硕士生：1.林春彬 2.游伟 3.文活

2011年优秀大学生夏令营开始招生 - 学院公告 - 中国人民大学信息学院
发布时间：2011-05-16 17:40
...1.高校计算机相关专业(本科生)。有志于参加应届计算机科学与技术科学研究工作的其他专业本科生；2、学习成绩优秀。本科总评成绩排名在年级前20%（在其他方面有突出表现的学生除外）。满足推荐免试硕(博)士研究生标准的学生优先。3、对计算机科学与技术研究有浓厚的兴趣，愿意从事科学研究工作。4、英语 ...

2012中国人民大学信息学院全国优秀大学生夏令营入营须知 - 学院公告 - 中国人民大学信息学院
发布时间：2012-07-08 10:31
...公正的选拔，被录取为2012中国人民大学信息学院全国优秀大学生夏令营的一员，以下是有关夏令营的一些事项说明：1、夏令营举办的时间：2012年7月16日~7月21日。3、乘车路线：北京站：乘坐地铁2号线（内环）在宣武门站换乘地铁4号线（安河桥北方向），在人民大学站下，步行至中国人民大学东门 北京西站：...

Thank you for listening!