

# 知识搜索引擎实验报告

2018050027 卢颐临 2018050026 韩子沛

## 实验环境

python 3.7

Django

npm

Jdk1.8

Lucene 8.8.2

## 问题描述

- 查询词相关知识实体的搜索与排序
- 知识实体关系推理和搜索

## 实现模块

我们的实现分为三个模块，前端，Django端和Java server端。

### 前端

这就是一个普通的前端，用vue和js写的，即网站前台部分，运行在PC端浏览器上展现给用户浏览的网页。将用户提供的搜索关键词和条件信息传给后端。同时处理后端返回的数据使得其能以美观的方式展现给用户。

### Django端

Django端是作为前端和Java server端的数据传递中枢。将前端提供的数据，进行一定处理后通过Py4j调用Java端的相应接口，获得搜索结果。然后将其传回前端。同时Django端还负责对RAW数据的处理，并调用Java的端口写入库中。

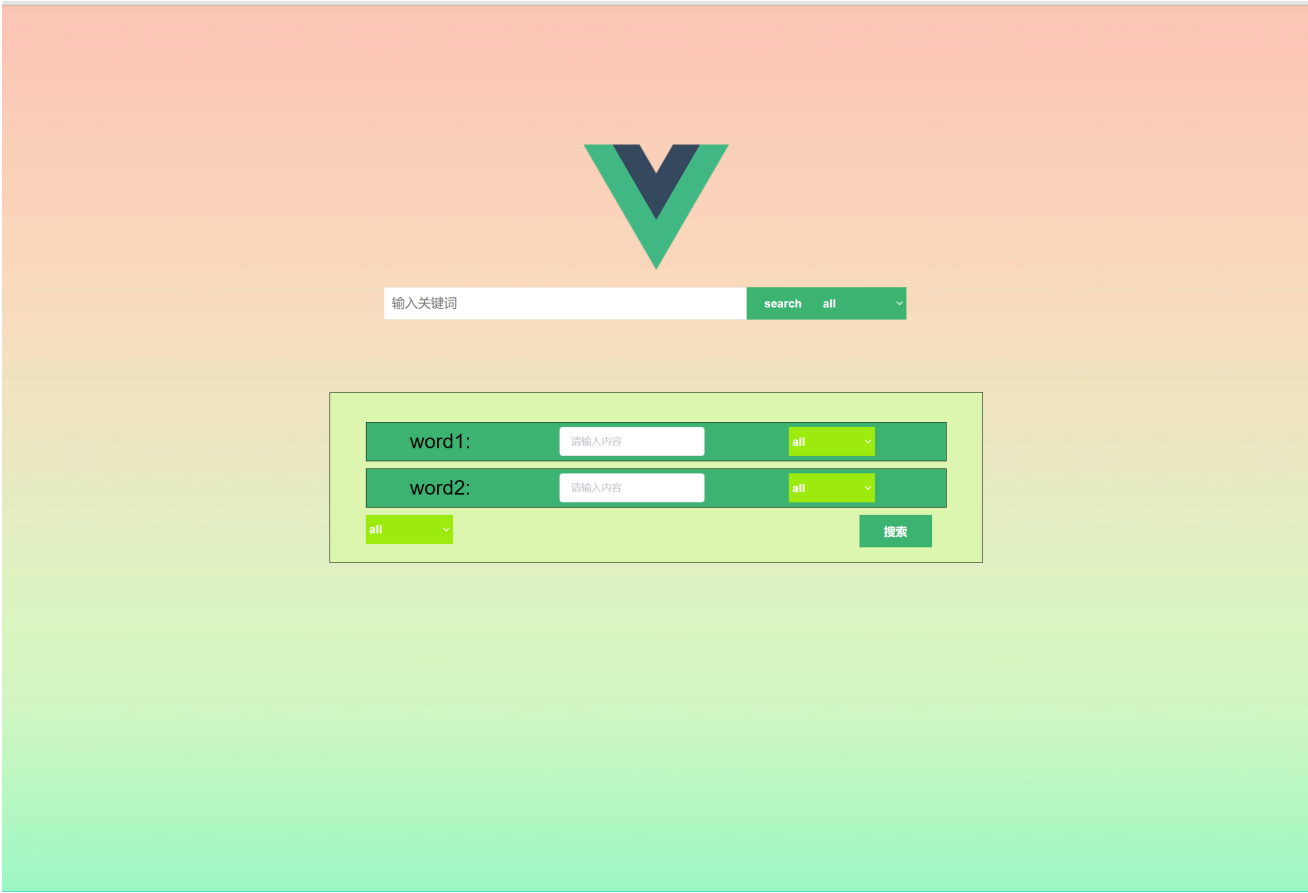
# Java server端

Java端是真正的后端，负责对数据基于lucene进行检索和排序。提供不同的接口以满足不同的搜索内容和方式。同时也提供了数据写入接口，以接收新的数据。

## 关键功能

第一个输入框用于搜索关键词，同时可以使用后面的Button限制搜索结果的知识实体的类型。

第二个输入框是用于搜索知识实体之间的关系。允许用户输入word1和word2然后同时可以使用Button限制搜索的知识实体的类型和只是实体之间的关系类型。



首页图片

这个页面展示了搜索关键词（北京）的展示结果，上方的搜索栏功能同首页的第一个搜索栏，可以继续搜索。下方列表展示的是搜索到的所有实体，信息包括实体的名称、类型、网址和评论。点击单个实体可以进入实体页，展示详细的实体信息。

返回首页	输入关键词	searchall
北京北京 (instance) https://baike.baidu.com/item/%E5%8C%97%E4%BA%AC%E5%8C%97%E4%BA%AC%2F1551988 《北京 北京》北京民谣歌手 郝云在2008年发布的专辑《郝云 北京》的主打歌曲，曾经流传于北京的大街小巷。		
北京北京 (instance) https://baike.baidu.com/item/%E5%8C%97%E4%BA%AC%E5%8C%97%E4%BA%AC%2F19210090 《北京北京》是G.E.M. 邓紫棋在 湖南卫视2016跨年晚会 开唱并于新年之际发布录音室单曲版本， 汪峰原唱、作词、作曲。 邓紫棋演唱版本由 Lupo Groinig重新编曲。		
北京，北京 (instance) https://baike.baidu.com/item/%E5%8C%97%E4%BA%AC%EF%BC%8C%E5%8C%97%E4%BA%AC%2F63146 《北京，北京》是 汪峰演唱的一首歌曲，歌曲由汪峰作词作曲，是2012年热播 电视剧《北京爱情故事》的插曲，歌曲收录在汪峰2007年的专辑《勇敢的心》中，于2007年6月23日发行。		
北京北京 (instance) https://baike.baidu.com/item/%E5%8C%97%E4%BA%AC%E5%8C%97%E4%BA%AC%2F66249 《北京北京》是由歌手 汪峰作词作曲并演唱的歌曲，收录于2007-06-01发行的专辑《勇敢的心》中，同时也是 电视剧《北京爱情故事》的片头曲。		
北京北京 (instance) https://baike.baidu.com/item/%E5%8C%97%E4%BA%AC%E5%8C%97%E4%BA%AC%2F1551998 《北京北京》，是 马唐“万物生长三部曲”的第二部，情节与风格都一脉相承。由青春步入成熟，从懵懂收获迷惘，浓重的 男性荷尔蒙是马唐作品一贯最大的特色。		
北京北京我爱北京 (instance) https://baike.baidu.com/item/%E5%8C%97%E4%BA%AC%E5%8C%97%E4%BA%AC%E6%88%91%E7%88%B1%E5%8C%97%E4%BA%AC 《北京北京我爱北京》是 关喆演唱的一首歌曲。::2008年北京奥运会闭幕式上，由 王力宏 (美国)、RAIN (韩国)、谭晶、陈慧琳 (中国香港)、韩雪、师鹏等6名中外歌手演唱激情、活力、动感的歌曲《北京，北京，我爱北京》。		
北京 (concept) http://baike.baidu.com/fenlei/%E5%8C%97%E4%BA%AC		
北京 (instance) https://baike.baidu.com/item/%E5%8C%97%E4%BA%AC%2F128981 北京，简称“京”，是中华人民共和国的 首都、直辖市、国家中心城市、超大城市、国际大都市，全国 政治中心、文化中心、国际交往中心、科技创新中心，是 中国共产党中央委员会、中华人民共和国中央人民政府、全国人民代表大会、中国人民政治协商会议全国委员会、中华人民共和国中央军事委员会所在地，也是 中部战区 司令部驻地。::北京位于 华北平原北部，背靠 燕山，毗邻 天津市和 河北省。北京的气候为典型的 北温带半湿润 大陆性季风气候。::北京是首批 国家历史文化名城和世界上拥有 世界文化遗产数最多的城市，三千多年的历史孕育了 故宫、天坛、八达岭长城、颐和园等众多名胜古迹。早在七十万年前，北京 周口店地区就出现了原始人群部落“北京人”。公元前1045年，北京成为 蓟、燕等 诸侯国的都城。公元938年以来，北京先后成为 辽 陪都、金、元、明、清国都。 1949年10月1日成为 中华人民共和国首都。::北京被全球权威机构GaWC评为 世界一线城市。 联合国报告指出，北京 人类发展指数居中国城市第二位 。2017年，北京 人均可支配收入达到57230元， 社会消费品零售总额11575.4亿元。北京 住户存款总额和人均住户存款均居全国第一。 2017年，北京 高新技术企业达到20183家。 福布斯2017年“ 中国大陆最佳商业城市排行榜”排第3位。::2015年7月31日， 国际奥委会主席 巴赫宣布北京携手 张家口获得2022年 冬季奥林匹克运动会的举办权。北京由此成为全球首个既举办过 夏季奥运会又将举办 冬季奥运会的城市。		
北京 (instance) https://baike.baidu.com/item/%E5%8C%97%E4%BA%AC%2F16822085 北京，今 内蒙古赤峰市宁城，原为 辽中京，金初仍沿用“中京大定府”旧称，海陵王贞元元年(1153年)，改名北京，置北京留守司，为北京路治所，辖有大定府、兴中府及利、义、锦、宗、建、兴等州。		
北京 (instance) https://baike.baidu.com/item/%E4%BA%AC%2F15808773 北京，简称“京”，是中华人民共和国的 首都、直辖市、国家中心城市、超大城市、国际大都市，全国 政治中心、文化中心、国际交往中心、科技创新中心，是 中国共产党中央委员会、中华人民共和国中央人民政府、全国人民代表大会、中国人民政治协商会议全国委员会、中华人民共和国中央军事委员会所在地，也是 中部战区 司令部驻地。::北京位于 华北平原北部，背靠 燕山，毗邻 天津市和 河北省。北京的气候为典型的 北温带半湿润 大陆性季风气候。::北京是首批 国家历史文化名城和世界上拥有 世界文化遗产数最多的城市，三千多年的历史孕育了 故宫、天坛、八达岭长城、颐和园等众多名胜古迹。早在七十万年前，北京 周口店地区就出现了原始人群部落“北京人”。公元前1045年，北京成为 蓟、燕等 诸侯国的都城。公元938年以来，北京先后成为 辽 陪都、金、元、明、清国都。 1949年10月1日成为 中华人民共和国首都。::北京被全球权威机构GaWC评为 世界一线城市。 联合国报告指出，北京 人类发展指数居中国城市第二位 。2017年，北京市 人均可支配收入达到57230元， 住户存款总额和人均住户存款 均居全国第一。 2017年，北京 高新技术企业达到20183家，数量居全国第一。 北京在福布斯2017年“ 中国大陆最佳商业城市排行榜”排第3位。::2015年7月31日， 国际奥委会主席 巴赫宣布北京携手 张家口获得2022年 冬季奥林匹克运动会的举办权。北京由此成为全球首个既举办过 夏季奥运会又将举办 冬季奥运会的城市。		
北京 (concept) https://zh.wikipedia.org/wiki/Category:%E5%8C%97%E4%BA%AC		
在北京 (instance) https://baike.baidu.com/item/%E5%9C%A8%E5%8C%97%E4%BA%AC%2F18730030 《在北京》是电影《爱拼北京》的主题曲，由 南征北战作词、作曲， 杨菲洋演唱。		
北京 (instance) https://baike.baidu.com/item/%E5%8C%97%E4%BA%AC%2F18302108 巴林左旗林东镇南波罗古城 古代称呼，由 金国所设，原为辽上京，金熙宗天眷元年（1138年）改名北京。		
在北京 (instance) https://baike.baidu.com/item/%F5%9C%A8%E5%8C%97%E4%BA%AC%2F9070442		

## 结果列表

这一页展示了唯一实体的所有信息，其中点击url网址可以在新窗口中打开相应网页，点击下方属性中的其他实体可以进入新的实体页。

[返回首页](#)[返回搜索页](#)

北京北京(instance)

<https://baike.baidu.com/item/%E5%8C%97%E4%BA%AC%E5%8C%97%E4%BA%AC%2F1551988>

comment : 《北京 北京》北京民谣歌手 郝云在2008年发布的专辑《郝云 北京》的主打歌曲，曾经流传于北京的大街小巷。

related : 

发现目标行动

北京

王朔

与青春有关的日子

property : 

谱曲 : [[bdi7339683|郝云]]

 | 

填词 : [[bdi7339683|郝云]]

 | 

编曲 : [[bdi7339683|郝云]]

 | 

歌曲原唱 : [[bdi7339683|郝云]]

 |

supplement : [郝云演唱歌曲]

实体展示页

这一页展示所有符合条件的关系，左边的实体是首页中输入的word1，右边的实体是首页中输入的word2，中间的词是俩个实体之间的关系。其中点击实体也可以进入相应的实体展示页。

[返回首页](#)

#### 北京北京 (instance)

《北京北京》是G.E.M. 邓紫棋在 湖南卫视2016跨年晚会 开唱并于新年之际发布录音室单曲版本，汪峰原唱、作词、作曲。邓紫棋演唱版本由 Lupo Groinig重新编曲。

#### 北京北京 (instance)

《北京北京》是G.E.M. 邓紫棋在 湖南卫视2016跨年晚会 开唱并于新年之际发布录音室单曲版本，汪峰原唱、作词、作曲。邓紫棋演唱版本由 Lupo Groinig重新编曲。

#### 邓紫棋 (instance)

G.E.M.邓紫棋，原名邓诗颖，1991年8月16日生于中国上海，4岁移居 香港，中国香港创作型女歌手。：;2008年7月10日以16岁之龄出道。同年10月16日发行首张EP《G.E.M.》，取得香港各大乐坛颁奖礼新人金奖。2011年5月，以19岁之龄在 香港红馆举行5场个人演唱会。2012年7月，发行第四张个人专辑《Xposed》，凭借该专辑邓紫棋在2013年获得 IFPI香港唱片销量大奖全年最高销量女歌手奖、最高销量国语唱片奖，并入围第24届金曲奖最佳国语女歌手奖。2014年1月，参加湖南卫视《我是歌手第二季》节目，获得总决赛亚军。3月31日，获第27届KCA 美国儿童选择奖“最受欢迎亚洲艺人”。12月10日，邓紫棋的蜡像入驻 香港杜莎夫人蜡像馆。2015年2月18日，参加 2015年中央电视台春节联欢晚会，自弹自唱其自创的歌曲《多远都要在一起》。5月12日，再登 福布斯中国名人榜，排名升至第11位。2015年7月，未满24岁便完成80场个人演唱会。截至北京时间2015年11月22日伦敦温布利终站，《G.E.M.X.X.X.Live"世界巡回演唱会》已经完成73场。打破之前华语乐坛女歌手单轮巡演59场的记录。8月15日，确认加盟 中国好声音第四季担任梦想导师，同年，成为首位在 瑞士阿尔卑斯山脉 少女峰举办音乐会的中国音乐人。2016年1月5日，邓紫棋入选《福布斯》'全球30岁以下30位最具潜力杰出音乐人'。

#### 邓紫棋 (instance)

G.E.M.邓紫棋，原名邓诗颖，1991年8月16日生于中国上海，4岁移居 香港，中国香港创作型女歌手。：;2008年7月10日以16岁之龄出道。同年10月16日发行首张EP《G.E.M.》，取得香港各大乐坛颁奖礼新人金奖。2011年5月，以19岁之龄在 香港红馆举行5场个人演唱会。2012年7月，发行第四张个人专辑《Xposed》，凭借该专辑邓紫棋在2013年获得 IFPI香港唱片销量大奖全年最高销量女歌手奖、最高销量国语唱片奖，并入围第24届金曲奖最佳国语女歌手奖。2014年1月，参加湖南卫视《我是歌手第二季》节目，获得总决赛亚军。3月31日，获第27届KCA 美国儿童选择奖“最受欢迎亚洲艺人”。12月10日，邓紫棋的蜡像入驻 香港杜莎夫人蜡像馆。2015年2月18日，参加 2015年中央电视台春节联欢晚会，自弹自唱其自创的歌曲《多远都要在一起》。5月12日，再登 福布斯中国名人榜，排名升至第11位。2015年7月，未满24岁便完成80场个人演唱会。截至北京时间2015年11月22日伦敦温布利终站，《G.E.M.X.X.X.Live"世界巡回演唱会》已经完成73场。打破之前华语乐坛女歌手单轮巡演59场的记录。8月15日，确认加盟 中国好声音第四季担任梦想导师，同年，成为首位在 瑞士阿尔卑斯山脉 少女峰举办音乐会的中国音乐人。2016年1月5日，邓紫棋入选《福布斯》'全球30岁以下30位最具潜力杰出音乐人'。

#### 邓紫棋 (instance)

G.E.M.邓紫棋，原名邓诗颖，1991年8月16日生于中国上海，4岁移居 香港，中国香港创作型女歌手。：;2008年7月10日以16岁之龄出道。同年10月16日发行首张EP《G.E.M.》，取得香港各大乐坛颁奖礼新人金奖。2011年5月，以19岁之龄在 香港红馆举行5场个人演唱会。2012年7月，发行第四张个人专辑《Xposed》，凭借该专辑邓紫棋在2013年获得 IFPI香港唱片销量大奖全年最高销量女歌手奖、最高销量国语唱片奖，并入围第24届金曲奖最佳国语女歌手奖。2014年1月，参加湖南卫视《我是歌手第二季》节目，获得总决赛亚军。3月31日，获第27届KCA 美国儿童选择奖“最受欢迎亚洲艺人”。12月10日，邓紫棋的蜡像入驻 香港杜莎夫人蜡像馆。2015年2月18日，参加 2015年中央电视台春节联欢晚会，自弹自唱其自创的歌曲《多远都要在一起》。5月12日，再登 福布斯中国名人榜，排名升至第11位。2015年7月，未满24岁便完成80场个人演唱会。截至北京时间2015年11月22日伦敦温布利终站，《G.E.M.X.X.X.Live"世界巡回演唱会》已经完成73场。打破之前华语乐坛女歌手单轮巡演59场的记录。8月15日，确认加盟 中国好声音第四季担任梦想导师，同年，成为首位在 瑞士阿尔卑斯山脉 少女峰举办音乐会的中国音乐人。2016年1月5日，邓紫棋入选《福布斯》'全球30岁以下30位最具潜力杰出音乐人'。

supplement

[ "邓紫棋" ]

comment

[ "邓紫棋" ]

comment

[ "北京" ]

comment

[ "北京" ]

comment

[ "北京" ]

关系展示列表

## 测试结果

我们使用的是Lucene提供的基于向量空间模型的排序算法计算公式为：

$$score(q,d) = coord(q,d) \times queryNorm(q) \times \sum_{t\_in\_q} (tf(t\_in\_q) \times idf(t))^2 \times t.getBoost() \times norm(t,d))$$

关键词

1.清华

<div>返回首页</div> <div>清华</div> <div>searchall</div>
<div>清华 (concept)</div> <div>http://baike.baidu.com/fenlei/%E6%B8%85%E5%8D%8E</div>
<div>清华同方 (concept)</div> <div>http://baike.baidu.com/fenlei/%E6%B8%85%E5%8D%8E%E5%90%8C%E6%96%B9</div>
<div>刘清华 (instance)</div> <div>https://baike.baidu.com/item/%E5%88%98%E6%B8%85%E5%8D%8E%2F7166859</div> <div>刘清华，女，汉族，1958年5月生，四川遂宁人，绵阳市审计局书记。</div>
<div>清华传奇 (instance)</div> <div>https://baike.baidu.com/item/%E6%B8%85%E5%8D%8E%E4%BC%A0%E5%A5%87</div> <div>《清华传奇》是2011年新世界出版社出版的图书，作者是 吴清军。</div>
<div>梁清华 (instance)</div> <div>https://baike.baidu.com/item/%E6%A2%81%E6%B8%85%E5%8D%8E%2F10219984</div> <div>梁清华，性别女，是浙江大学思想政治理论学科教研讲师。</div>
<div>张清华 (instance)</div> <div>https://baike.baidu.com/item/%E5%BC%A0%E6%B8%85%E5%8D%8E%2F4143627</div> <div>张清华，男，汉族，1965年3月出生，广东丰顺人，中共党员，工学博士，教授，现任 广东石油化工学院校长、党委副书记，广东省石化装备故障诊断重点实验室主任，广东高校石油化工装备故障诊断与信息化控制工程技术开发中心主任，兼任中国自动化学会技术过程故障诊断与安全性专业委员会委员、广东省自动化学会常务理事、广东省青年科学家协会理事、茂名市科协副主席、《自动化与信息工程》编委会委员、茂名市人大代表、茂南区人大代表等职务，是广东省高等学校“千百十工程”第四批省级培养对象。</div>
<div>刘清华 (instance)</div> <div>https://baike.baidu.com/item/%E5%88%98%E6%B8%85%E5%8D%8E%2F7166821</div> <div>男，汉。笔名：天涯石、江郎、征夫；自号：三味书画斋主人。业余爱好广泛，尤其对书、画“情有独钟”，并立志有所建树。</div>
<div>清华漫话 (instance)</div> <div>https://baike.baidu.com/item/%E6%B8%85%E5%8D%8E%E6%BC%AB%E8%AF%9D</div> <div>《清华漫话》是2009年清华大学出版社出版的图书，作者是清华大学校史研究室。</div>
<div>吴清华 (instance)</div> <div>https://baike.baidu.com/item/%E5%90%B4%E6%B8%85%E5%8D%8E%2F19387</div>
<div>清华逸事 (instance)</div> <div>https://baike.baidu.com/item/%E6%B8%85%E5%8D%8E%E9%80%B8%E4%BA%8B</div> <div>《清华逸事》是辽海出版社出版的图书，作者是黄延复。</div>
<div>梁清华 (instance)</div> <div>https://baike.baidu.com/item/%E6%A2%81%E6%B8%85%E5%8D%8E%2F14771</div> <div>梁清华，女，1949年3月出生，湖南耒阳人。大学毕业，湖南医科大学湘雅医院中西医结合研究所教授、主任医师、博士生导师。湖南省重点学科----中西医结合学科带头人。中国中西医结合风湿类病专业委员会委员，中国中西医结合实验医学专业委员会委员，中国中西医结合基础研究专业委员会委员，中南六省中西医结合防治风湿病协作委员会委员，湖南省中西医结合风湿免疫病学会副主任委员，湖南省中西医结合神经病学学会副主任委员，湖南省中西医结合基础研究专业委员会副主任委员。</div>
<div>刘清华 (instance)</div> <div>https://baike.baidu.com/item/%E5%88%98%E6%B8%85%E5%8D%8E%2F7166808</div> <div>刘清华是我国著名的武术 运动员，正是她在 九运会 武术套路决赛女子长拳比赛中，不负众望、技压群芳勇夺该项比赛冠军，为 北京代表团赢得了首枚沉甸甸的九运会金牌。</div>
<div>清华斋 (instance)</div> <div>https://baike.baidu.com/item/%E6%B8%85%E5%8D%8E%E6%96%B8</div> <div>清华斋始建于清 光绪二十六年（公元1900年），由 北京人、回民 马子清创建。马子清用卖面攒下的钱买了彰仪门附近路北的王振亭面铺，开办了 饽饽铺，起名“清顺斋”，以取己名中的“清”和寓意顺顺当当做生意。清华斋制作的 糕点有 酥皮的 龙凤饼、浆皮的 状元饼、大 薄脆及各式 月饼、麻元，蒸货有 蜂糕、豆糕、喇嘛糕（黄糕）、糕干条，炸货有 芙蓉糕、姜汁排叉、蜜麻花、萨其马，还有 槽子糕（蜜糕）、莲子缸炉”等近百个品种。</div>
<div>曹清华 (instance)</div> <div>https://baike.baidu.com/item/%E6%9B%B9%E6%B8%85%E5%8D%8E%2F42895</div> <div>曹清华，女，1946年11月生，毕业于陕西中医学院（现陕西中医药大学），曾在西安雁塔区中医院任职中医妇科副主任医师，曹清华于2007年7月25日获得“一种治疗关节炎的药物组合物及制备方法”的专利。专利号:ZL 2005 1 0002294.2。于中国卫生部网站查询其主要执业地点为西安莲湖曹清华诊所，执业证书编码：141610104000087，发证(批准)机关为西安市莲湖区卫生和计划生育局。</div>
<div>清华青田 (instance)</div>

2.清华大学



返回首页	输入关键词	search	all	▼
清华大学 (concept)				
http://baike.baidu.com/fenlei/%E6%B8%85%E5%8D%8E%E5%A4%A7%E5%AD%A6				
清华大学 (instance)				
https://baike.baidu.com/item/%E6%B8%85%E5%8D%8E%2F1832793				
清华大学（Tsinghua University），简称“清华”，由 中华人民共和国教育部直属，中央直管 副部级建制，位列“211工程”、“985工程”、“世界一流大学和一流学科”，入选“基础学科拔尖学生培养试验计划”、“高等学校创新能力提升计划”、“高等学校学科创新引智计划”，为 九校联盟、中国大学校长联谊会、东亚研究型大学协会、亚洲大学联盟、环太平洋大学联盟、清华—剑桥—MIT低碳大学联盟成员，被誉为“红色工程师的摇篮”。::清华大学的前身 清华学堂始建于1911年，因 水木清华而得名，是 清政府设立的留美预备学校，其建校的资 <span></span> 金源于1908年美国退还的部分 庚子赔款。1912年更名为清华学校。1928年更名为 国立清华大学。1937年 抗日战争全面爆发后南迁长沙，与 北京大学、南开大学组建 国立长沙临时大学，1938年迁至昆明改名为 国立西南联合大学。1946年迁回 清华园。1949年 中华人民共和国成立，清华大学进入了新的发展阶段。1952年全国高等学校院系调整后成为多科性工业大学。1978年以来逐步恢复和发展为综合性的研究型大学。::水木清华，钟灵毓秀，清华大学秉持“自强不息、厚德载物”的校训和“行胜于言”的校风，坚持“中西融汇、古今贯通、文理渗透”的办学风格和“又红又专、全面发展”的培养特色，弘扬“爱国奉献、追求卓越”传统和“人文日新”精神。恰如清华园 工字厅内对联所书——“槛外山光，历春夏秋冬、万千变幻，都非凡境；窗中云影，任东西南北、去来滄海，洵是仙居”。				
清华大学 (concept)				
https://zh.wikipedia.org/wiki/Category:%E6%B8%85%E5%8D%8E%E5%A4%A7%E5%AD%A6				
清华大学 (instance)				
https://baike.baidu.com/item/%E6%B8%85%E5%8D%8E%E5%A4%A7%E5%AD%A6				
清华大学（Tsinghua University），简称“清华”，由 中华人民共和国教育部直属，中央直管 副部级建制，是中国著名高等学府，位列“211工程”、“985工程”、“世界一流大学和一流学科”建设高校，入选“基础学科拔尖学生培养试验计划”、“高等学校创新能力提升计划”、“高等学校学科创新引智计划”，为 九校联盟、中国大学校长联谊会、东亚研究型大学协会、亚洲大学联盟、环太平洋大学联盟、清华—剑桥—MIT低碳大学联盟成员，被誉为“红色工程师的摇篮”。::清华大学的前身 清华学堂始建于1911年，因 水木清华而得名，是 清政府设立的留美预备学校，其建校的资 <span></span> 金源于1908年美国退还的部分 庚子赔款。1912年更名为清华学校。1928年更名为 国立清华大学。1937年 抗日战争全面爆发后南迁长沙，与 北京大学、南开大学组建 国立长沙临时大学，1938年迁至昆明改名为 国立西南联合大学。1946年迁回 清华园。1949年 中华人民共和国成立，清华大学进入了新的发展阶段。1952年全国高等学校 院系调整后成为多科性工业大学。1978年以来逐步恢复和发展为综合性的研究型大学。::清华大学秉持“自强不息、厚德载物”的校训和“行胜于言”的校风，坚持“中西融汇、古今贯通、文理渗透”的办学风格和“又红又专、全面发展”的培养特色，弘扬“爱国奉献、追求卓越”传统和“人文日新”精神。恰如清华园 工字厅内对联所书——“槛外山光，历春夏秋冬、万千变幻，都非凡境；窗中云影，任东西南北、去来滄海，洵是仙居”。				
清华大学 (instance)				
https://zh.wikipedia.org/wiki/清华大学				
“清华大学”（，缩写作），简称“清华”，旧称“清华学堂”、“清华学校”、“国立清华大学”，是一所位于中华人民共和国zh128zh11702清华园的公立大学。始建于1911年，因北京西北郊清华园得名。初为清政府利用zh133退还的部分zh13962所建留美预备学校“游美学务处”及附设“肄业馆”，于1925年始设大学部。抗日战争爆发后，清华与北大、南开南迁长沙，组建zh41148。1938年再迁昆明，易名zh8052。1946年迁回清华园复校，拥有文、法、理、工、农等5个学院。1949年中华人民共和国成立后，国立清华大学归属zh225675，更名“清华大学”；而原国立清华大学校长zh15659于1955年在台湾新竹复校，仍沿用原名。::1952年，中国高校进行院系调整，清华大学文、法、理、农、航天等院系外迁，吸纳外校工科，转为多科性工业大学，在土木、水利、计算机、核能等领域贡献卓越，被誉为“zh50655的摇篮”。1978年后，逐步恢复理科、人文社科、经济管理类学科。1999年，原中央工艺美术学院并入，成立美术学院。2006年，与zh73467合作办学，培养临床医学专业学生。2012年，原中国人民银行研究生部并入，成立五道口金融学院。2013年，黑石集团捐助成立清华大学苏世民书院及奖学金，与著名的牛津大学罗德奖学金及剑桥大学盖兹奖学金在捐助规模及名声上皆相当。::截至2016年12月，清华大学拥有zh162338、博物馆、图书馆、20个学院，及近200个科研机构、5家校办产业以及一个科技园区，分别为zh274372及其旗下的紫光集团、同方集团、诚志集团、清华科技园等。学校拥有固定资产超过191亿元人民币，控股资产超过3500亿元人民币，是zh10465、zh8004、双一流高等院校。2016年《QS世界大学排名》、《zh241667》均将清华大学排在中国大学首位。				
台湾清华大学 (instance)				
https://baike.baidu.com/item/%E6%96%B0%E7%AB%B9%E5%A4%A7%E5%AD%A6%2F4997487				
台湾清华大学（National Tsing Hua University, NTHU），简称“清大”，或清华大学（台湾），为中国顶尖级研究型大学之一，环太平洋大学联盟、东亚研究型大学协会成员。::台湾清华大学前身为1911年在 北京设立的 清华学堂。1925年设大学部。对日抗战期间，西迁至云南 昆明，与 国立北京大学、私立南开大学合组 国立西南联合大学。国共内战后，清华大学被分成两个，北京的 清华大学由 中华人民共和国政府接管，而台湾当局则于1955年于台湾省 新竹市让 清华大学复校，复校之初首设原子科学研究所，1964年恢复大学部。::截至2016年11月，台湾清华大学共设有理、工、原子科学、人文社会、生命科学、电机资讯及科技管理等10个学院。在台已造就英才超过五万人，校友包括诺贝尔奖得主 李远哲、“中央研究院”院士12人。				
清华大学教师 (concept)				
https://zh.wikipedia.org/wiki/Category:%E6%B8%85%E8%8F%AF%E5%A4%A7%E5%AD%B8%E6%95%99%E5%B8%88				
国立清华大学 (concept)				
https://zh.wikipedia.org/wiki/Category:%E5%9C%8B%E7%AB%8B%E6%B8%85%E8%8F%AF%E5%A4%A7%E5%AD%B8				
台湾清华大学 (instance)				
https://baike.baidu.com/item/%E5%9B%BD%E7%AB%8B%E6%B8%85%E5%8D%8E%E5%A4%A7%E5%AD%A6%2F3075325				
台湾清华大学（National Tsing Hua University, NTHU），简称“清大”，或清华大学（台湾），为中国顶尖级研究型大学之一，环太平洋大学联盟、东亚研究型大学协会成员。::台湾清华大学前身为1911年在 北京设立的 清华学堂。1925年设大学部。对日抗战期间，西迁至云南 昆明，与 国立北京大学、私立南开大学合组 国立西南联合大学。国共内战后，清华大学被分成两个，北京的 清华大学由 中华人民共和国政府接管，而台湾当局则于1955年于台湾省 新竹市让 清华大学复校，复校之初首设原子科学研究所，1964年恢复大学部。::截至2016年11月，台湾清华大学共设有理、工、原子科学、人文社会、生命科学、电机资讯及科技管理等10个学院。在台已造就英才超过五万人，校友包括诺贝尔奖得主 李远哲、“中央研究院”院士12人。				
台湾清华大学 (instance)				
https://baike.baidu.com/item/%E6%B8%85%E5%8D%8E%2F17917889				
台湾清华大学（National Tsing Hua University, NTHU），简称“清大”，或清华大学（台湾），为中国顶尖级研究型大学之一，环太平洋大学联盟、东亚研究型大学协会成员。::台湾清华大学前身为1911年在 北京设立的 清华学堂。1925年设大学部。对日抗战期间，西迁至云南 昆明，与 国立北京大学、私立南开大学合组 国立西南联合大学。国共内战后，清华大学被分成两个，北京的 清华大学由 中华人民共和国政府接管，而台湾当局则于1955年于台湾省 新竹市让 清华大学复校，复校之初首设原子科学研究所，1964年恢复大学部。::截至2016年11月，台湾清华大学共设有理、工、原子科学、人文社会、生命科学、电机资讯及科技管理等10个学院。在台已造就英才超过五万人，校友包括诺贝尔奖得主 李远哲、“中央研究院”院士12人。				

关系:

1. word1：清华大学 word2：北京大学





下面我们分别使用MAP, P@10, MRR对搜索引擎的性能进行定量分析:

	MAP	P@10	MRR
清华	1	0.1	1
清华大学	1	0.5	1
百度	1	1	1
可口可乐	0.87083	0.6	1
暹罗	0.65333	0.5	1
tsinghua	0.77585	0.7	1
pleural	1	1	1

测试中5个词的MRR都为1, 说明搜索引擎可以较好的搜索到我们先要搜寻的内容。但是对于关键词语义的理解较差。比如清华和清华大学, 清华的p@10仅有0.1, 因为在搜索结果中清华的搜索结果除了清华大学都是人物的名字, 并且绝大多数关键词的p@10都很低, 可以看出搜索引擎不能感知用户搜索的意图。尽管如此, 非冷门关键词对应的MAP都比较高, 说明搜索引擎在搜索结果排序上比较稳定。同时发现搜索引擎对于英文搜索支持的也比较好, 甚至冷门的pleural也能搜到。但我觉得也和我们的搜索引擎是只是搜索引擎的情况有关。

## 实验分工

卢颐临: Java server Django的通信部分

韩子沛: 前端和Django的数据预处理部分

## 使用的开源资料

- [Lucene](#)
- [element-ui](#)