

2020-7-14

# 《挑战极限报告》

——NotOnlyFilm 电影垂直搜索引擎

《项目实训》2 班 G23

组长：张佳瑶

组员：贺婷婷 应承峻 戴陈威 杨建伟

## 目录

1 挑战极限简介.....	2
2 深度搜索.....	3
2.1 查询分发与结果整合.....	3
2.2 多维排序.....	4
2.3 规则搜索.....	9
2.4 智能补全.....	13
2.5 搜索指数可视化.....	14
3 深度分析.....	16
3.1 影评关键字词云.....	16
3.2 同类型电影对比.....	17
3.3 猜你喜欢推荐.....	18
3.4 基于知识图谱的自动问答.....	19
3.5 关系图谱可视化.....	22

# 1 挑战极限简介

在十五天的项目实训课程实践过程中，团队开发完成了一款名为 Not Only Film 的电影领域垂直搜索引擎。除了完成搜索关键字得到相关结果并予以显示的基本功能外，团队对于极限挑战进行了一定的设计与实现，扩展了平台的功能，提升了用户的使用体验。

在深度搜索部分，团队开发完成了查询分发与结果整合、多维排序、规则搜索、智能补全、搜索指数可视化的功能；在深度分析方面，团队实现了影评关键词词云、同类型电影对比、猜你喜欢推荐、基于知识图谱的自动问答、关系图谱可视化的功能。

## 2 深度搜索

### 2.1 查询分发与结果整合

#### 2.1.1 简介

在搜索时能够分发查询关键字，在电影、资讯、种子三种不同类型的数据资源中进行搜索并返回符合关键字的结果。

在搜索结果页默认“全部”展示，聚合所有类型的数据资源，依据内容相关度以及数据资源类型排序展示。

#### 2.1.2 原理及实现

使用 Elastic Search 中的多索引查询结合 MultiQuery 多字段匹配实现，为不同索引的不同字段分配不同的权重进行查询，将返回的结果使用泛型数组 List<QueryResBo> 存储，所有的返回结果类 MovieResBo, RecommendResBo, MagnetResBo, MagnetResBo 全部继承 QueryResBo 基类。

#### 2.1.3 效果展示

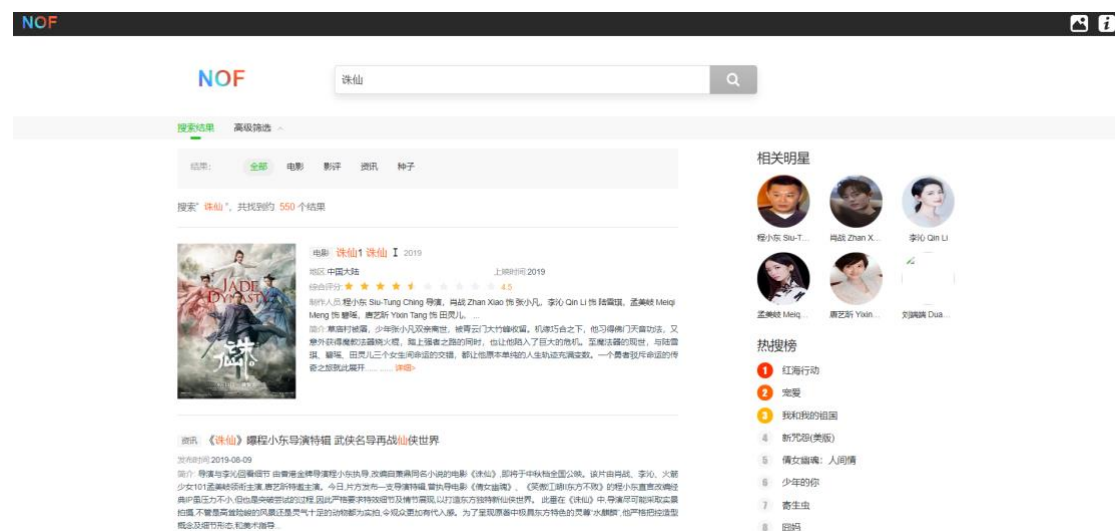


图 1 多类型资源整合后的搜索结果-1

资讯 《诛仙I》释片尾曲MV 肖战孟美岐演绎年少情真

发布时间:2019-09-06

简介: 由香港金牌导演程小东执导,改编自萧鼎同名小说的电影《诛仙》将于中秋档9月13日全国公映。该片由肖战、李沁、火箭少女101孟美岐领衔主演,唐艺昕特邀出演。片方今日公布电影片尾曲《情意结》及同名MV,由周深倾情献唱,空灵细腻的嗓音与画面中少年的情真意切直击心灵,缥缈的青云仙境渐渐显露阵容。张小凡怀抱碧瑶,周深作为华语乐坛特色与实力兼备的歌手,嗓音尤其纯净空灵,对于情感与意境的把控十分到位。此次在电影《诛仙》的片尾曲《情意结》中,周深以柔和飘逸的嗓音将张小凡少年时的纯真和命运的无奈娓娓道来。陆雪琪眼中含泪《情意结》整首歌曲旋律婉转悠扬,一句“如下辈子...”

来源:全球电影资讯网

资讯 《诛仙》唐艺昕展现“花式”吊威亚技能 高空倒吊&快速降落

发布时间:2019-08-16

简介: 由香港导演程小东执导,改编自萧鼎同名小说的电影《诛仙》,即将于中秋档全国公映。该片由肖战、李沁、火箭少女101孟美岐领衔主演,唐艺昕特邀出演。田灵儿拔剑 今日,片方发布了一支田灵儿“花式威亚”特辑,唐艺昕饰演的田灵儿是小说中重要的女性角色之一,天资出众,灵动可爱是她独特的标签。在特辑中,唐艺昕各种“花式”吊威亚、高空、倒吊、快速降落,这让她有轻微恐高症的唐唐在采访时回忆起来仍是有些害怕,但敬业的她每一个动作都反复琢磨,精益求精,争取给观众呈现出最好的“田灵儿”。电影《诛仙》改编自萧鼎同名小说,小说中的角色人物,在读者心中留下了深刻印象,尤其是灵动娇俏...

来源:全球电影资讯网

种子 诛仙 I /诛仙/诛仙电影版 Jade Dynasty

文件: 2019年奇幻《诛仙 I /诛仙电影版》HD国语中字

size:1CD

种子 诛仙 I /诛仙/诛仙电影版 Jade Dynasty

文件: 2019年奇幻《诛仙 I /诛仙电影版》BD国语中字

size:1CD

图 2 多类型资源整合后的搜索结果-2

## 2.2 多维排序

### 2.2.1 简介

能够针对某一特定类别的资源进行多种类型的排序,默认采用相关性从高到低的排序方式,可选评分从高到低、时间从近到远的排序方式。

支持对于电影类型数据按类型、地区、年份、时长进行筛选,允许设置多个筛选条件;支持对于资讯类型数据按发布时间进行筛选。

### 2.2.2 原理及实现

筛选时,通过构建 Elastic Search 中 BoolQuery 的 Filter,对于单值筛选可以使用 term 设置筛选条件,而对于多值筛选可以使用 terms 设置筛选条件。

对于排序,直接指定 Elastic Search 中 Sort 进行字段配置。此外通过排序重打分的的方式来牺牲排序的绝对性换取准确度的提升,以按时间从近到远的排序为例,

在使用排序的功能时,为了达到严格的排序效果,往往会严重降低查询的准确性,而这往往很难满足用户需要,因此我们需要对排序效果做出一定的牺牲,

来换取查询结果的准确性。即我们在对时间进行排序时，不是严格地按照时间进行排序，而是根据其匹配的分值得分×时间衰减率，进行重打分计算

时间打分模型：通过在 ScriptSort 引入指数衰减曲线和 Gauss 衰减函数曲线分别对评分和时间进行衰减处理。在 Gauss 衰减函数曲线中，我们根据年份的分布情况对下图中  $f(d)$  做了微调，经过多次调参我们确定了模型函数。

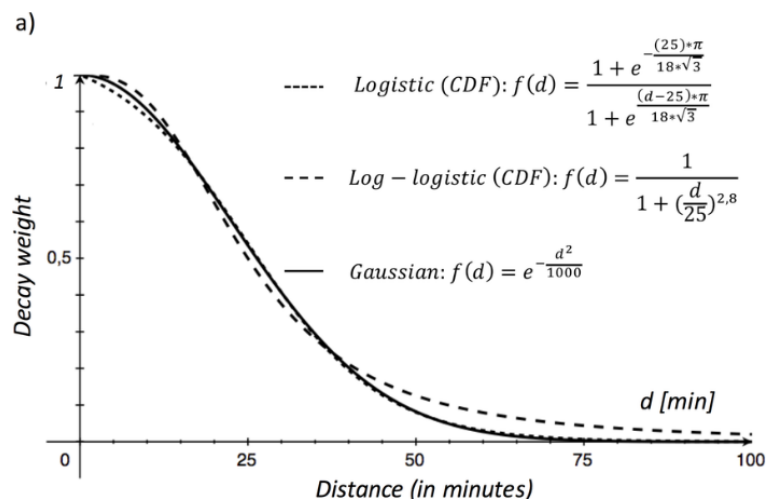




图 3 衰减曲线

## 2.2.3 效果展示

### 2.2.3.1 多维排序





搜索结果

高级筛选

搜索“子弹”，共找到约 1033 个结果



电影 消失的子弹 2012

地区 中国大陆 / 中国香港

上映时间 2012

综合评分  6.5

制作人员 罗志良 Chi-Leung Law 导演, 谢霆锋 Nicholas Tse 演员, 刘青云 Sean Lau 演员, 杨冪 Mi Yang 演员, 井柏然 Boran Jing 演员, ...

简介: 民国时期, 国内最大的兵工厂发生离奇命案, 死者被“幽灵子弹”所杀——犯罪现场找不到弹头。一时之间, 疑云密布, 人心惶惶, 此案堪称民国第一奇案。此案交由警局内负责管理档案的政务警察松东路调查 (刘青云 饰), 担任文职的松东路既不会拳脚功夫也不懂如何开枪, 却思想缜密, 对破案疑难杂案有独特方法。一心为民除害的热血警探郭追 (谢霆锋 饰) 不容于警局贪污腐败氛围, 联手松东路调查此离奇命案。 ..... [详细>](#)



电影 让子弹飞 2010

地区 中国大陆 / 中国香港


上映时间 2010


综合评分  8.8

制作人员 姜文 Wen Jiang 导演, 姜文 Wen Jiang 饰 张牧之, 葛优 You Ge 饰 马邦德, 周润发 Yun-Fat Chow 饰 黄四郎, 刘嘉玲 Carina Lau 饰 县长夫人, ...

简介: 民国年间, 花轿抬得县长的马邦德 (葛优 饰) 携妻 (刘嘉玲 饰) 及随从走马上任。途经南国某地, 遭劫匪张麻子 (姜文 饰) 一伙伏击, 随从尽死, 只夫妻二人侥幸活命。马为保命, 谎称自己是县长的汤 师爷。为汤师爷许下的财富所动, 张麻子摇身一变化身县长, 带着手下赶赴鹅城上任。有道是天高皇帝远, 鹅城地处偏僻, 一方霸主黄四郎 (周润发 饰) 只手遮天, 全然不将这个新来的县长放在眼里。张麻子痛打了黄的武教头 (姜武 饰), 黄 ..... [详细>](#)

图 4 默认按照相关性排序





搜索结果

高级筛选 ^

结果:

全部

电影

影评

资讯

种子

类型:

全部

剧情

喜剧

爱情

动作

惊悚

犯罪

悬疑

恐怖

历史

动画

音乐

科幻

战争

歌舞

奇幻

家庭

武侠

情色

灾难

儿童

地区:

全部

美国

中国大陆

日本

中国香港

英国

法国

韩国

中国台湾

印度

德国

加拿大

意大利

年份:

全部

2011

2012

2013

2014

2015

2016

2017

2018

2019

2020

时长:

全部

一小时内

一到三小时

三小时以上

排序:

相关性最高

最新

评分最高

搜索“子弹”，共找到约 5954 个结果



电影 分裂: 黑色子弹 Torn: Dark Bullets 2020

地区: 加拿大

上映时间: 2020

综合评分:  5.4

制作人员: 山姆·文森特 Sam Vincent 演员, 凯伦·霍尔尼斯 Karen Holness 演员, 多隆·贝尔 Doron Bell 演员, 莎伦·泰勒 Sharon Taylor 演员, 杰西·哈奇 Jesse Hutch 演员, ...

简介: 对警察枪击、种族主义和他们共同的联系进行原始的、毫无歉意的审视。没有说明我们知道答案，但不害怕直接提问。..... [详细>](#)

图 5 点击最新采用从远到近排序



## 2.2.3.2 多维筛选

搜索“爱”，共找到约 5954 个结果



电影 爱，爱你 อย่างนอกให้รู้วัก 2015

地区 泰国

上映时间 2015

综合评分: ★★★★★★☆☆☆☆ 6.3

制作人员 纳帕特·猜提昂通 Naphat Chaithiangthum 导演, 巴特潘·忽栓 Athaphan Poonsawas 饰 Gump, 诺拉帕·沙功颂 Pak Norrapat Sakulsong 饰 Nai, 查依·散顶腾古 Chanon Santinatornkul 饰 Pid, 吴俊 Khunatipapisiri Korn 饰 Zee ...

简介 Gump被妈妈推去安慰失恋的Sunz, Sunz是他妈妈朋友的一个侄子, 然而, 由于Sunz的魅力和Gump自己爱玩的习惯, Gump在遇到Sunz的第一眼开始, 感觉到了自己的动摇; 另一边, Gump的男友Nai遇到了他的高中时期的朋友Ball, Ball回来找寻他某些记忆中的东西。真是复杂的爱, 他们需要好好梳理一下了。..... [详细>](#)



电影 爱砸了 Screwed 2017

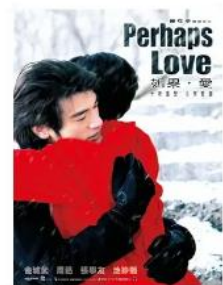
地区 芬兰

上映时间 2017

综合评分: ★★★★★★☆☆☆☆ 6.9

制作人员 尼尔斯·埃里克·艾克布洛姆 Nils-Erik Ekblom 导演, 米寇·考毗拉 Mikko Kauppila 饰 米库 Miku, 瓦特瑞·雷特宁 Valtteri Lehtinen 饰 伊莱亚斯 Elias, 桑娜·玛朱利 Sanna Majuri 饰 明娜 Minna, 萨米·胡塔拉 Sami Huhtala 饰 贾科 Jaakko, ...

简介 一次派对狂欢使得17岁的Miku被父母赶去乡间小屋暂住, 而这却为他和Elias的初次相逢提供了机会。两个家庭失和的男孩惺惺相惜、相依相偎, 两人的关系在这短暂的夏日时光中, 激起点点涟漪。..... [详细>](#)



电影 如果 爱 2005

地区 中国大陆 / 中国香港 / 马来西亚

上映时间 2005

综合评分: ★★★★★★☆☆☆☆ 7.7

制作人员 陈可辛 Peter Chan 导演, 赵良骏 Samson Chiu 导演, 金城武 Takeshi Kaneshiro 饰 林见东, 周迅 Xun Zhou 饰 孙纳, 张学友 Jacky Cheung 饰 聂文, ...

简介 想演戏的孙纳(周迅 饰)与想导演的林见东(金城武 饰)初遇见时, 两人都正处于生活与事业的低谷, 在相互鼓励和扶持中, 他们渐渐有了共同的可以看见希望的未来, 然而孙纳更想把现在的可能抓到手, 狠心与林见东分了手, 跟了可以给她戏的导演聂文(张学友 饰), 成为人人艳羡的大明星。..... [详细>](#)

图 6 默认相关性排序电影类型全部结果



图 7 多维筛选后得到符合要求的结果

## 2.3 规则搜索

### 2.3.1 简介

用户可以根据搜索符号规则组织搜索输入语句，进行搜索范围限定，获得更加精准的搜索结果。

Not Only Film 平台设计实现了以下四种高级规则搜索：

- “+<Key>” 表示 Key 必须在查询结果中出现。
- “-<Key>” 表示 Key 不应该在查询结果中出现。
- “<Key1> <Key2>” 表示表示两个关键字匹配任意一个均可。
- “:intitle <Key>” 表示匹配的查询结果的标题中必须要出现 Key。

### 2.3.2 原理及实现

根据通用的查询语法，“+<Key>”表示这个关键字必须在查询结果中出现，“-<Key>”表示这个字段不应该在查询结果中出现，“<Key1> <Key2>”表示表示两个关键字匹配任意一个均可，“:intitle <Key>”表示匹配的查询结果的标题中必须要出现这个关键字。

我们设计了一种通用的查询结构：使用 ES 的布尔查询加上排序选项以及高亮选项。分析可得，布尔查询主要由与（Must）、或（Should）、非（Must Not）以及一个过滤器组成。

因此，Must 的作用主要是匹配查询字符串中的“+”符号，Should 字符串主要匹配查询字符串中的空格，Must Not 主要匹配查询字符串中的“-”符号。

```
{
  "query": {
    "bool": {
      "must": [],
      "should": [],
      "filter": [],
      "must_not": []
    }
  },
  "sort": [],
  "highlight": {
    "pre_tags": "<em>",
    "post_tags": "</em>",
    "fields": {}
  },
  "from": 0,
  "size": 10
}
```

## 2.3.3 效果展示

搜索“罪恶”，共找到约 1362 个结果



电影 罪恶之城 Sin City 2005

地区 美国

上映时间 2005

综合评分: ★★★★★★☆☆ 8.4

制作人员 弗兰克·米勒 Frank Miller 导演, 罗伯特·罗德里格兹 Robert Rodriguez 导演, 杰西卡·阿尔芭 Jessica Alba 饰 Nancy, 布鲁斯·威利斯 Bruce Willis 饰 Hartigan, 克萊夫·歐文 Clive Owen 饰 Dwight, ...

简介 影片改编自著名漫画大师弗兰克·米勒的经典同名漫画, 由三个故事组成, 都发生在虚拟的罪恶之城, 这里, 警察和暴徒已无分别, 暴力摧毁一切, 也保护一切。 故事一: 罪城 .....[详细>](#)



电影 罪恶之家 An Inspector Calls 2015

地区 英国

上映时间 2015

综合评分: ★★★★★★☆☆ 8.3

制作人员 艾斯林·沃什 Aisling Walsh 导演, 大卫·休里斯 David Thewlis 饰 The Inspector, 米兰达·理查森 Miranda Richardson 饰 Sybil Birling, 苏菲·兰朵 Sophie Rundle 饰 Eva, 肯·斯托特 Ken Stott 饰 Arthur Birling, ...

简介 亚瑟 (肯·斯托特 Ken Stott 饰) 是一位成功的资本家, 在不久之后即将加官进爵。希拉 (克洛伊·皮里 Chloe Pirrie 饰) 是亚瑟的女儿, 这位单纯善良的千金小姐在家宴中接受了男友杰拉德 (凯尔·索列尔 Kyle Soller 饰) 的求婚, 两人喜结连理。亚瑟和妻子对这门亲事显然十分满意, 而希拉的弟弟艾瑞克 (芬恩·科尔 Finn Cole 饰) 却显然并不怎么高兴。 .....[详细>](#)



电影 阳光下的罪恶 Evil Under the Sun

地区 英国

上映时间:

综合评分: ★★★★★★☆☆ 8.3


制作人员 盖伊·汉弥尔顿 Guy Hamilton 导演, 彼得·乌斯捷诺夫 Peter Ustinov 演员, 玛吉·史密斯 Maggie Smith 演员, 科林·布莱克利 Colin Blakely 演员, 黛安娜·里格 Diana Rigg 演员, ...

简介 阿伦娜·马歇尔 (黛安娜·里格 Diana Rigg 饰) 是一个美艳不可方物的女明星, 天性放浪的她从不掩饰对其他男性的喜爱, 也十分享受被男性追求的感觉, 从而彻底忽略了丈夫和养女的感受。帕特里克 (尼古拉斯·克莱 Nicholas Clay 饰) 是个血气方刚的英俊小伙, 虽然已经有了体弱多病的妻子克里斯汀 (简·伯金 Jane Birkin 饰), 但他依然难以抵挡向他暗送秋波的阿伦娜的诱惑。今年, 位于莱瑟库 .....[详细>](#)

图 8 “罪恶”搜索结果



搜索“罪恶+之城”，共找到约 498 个结果




**电影 罪恶之城 Sin City 2005**

地区: 美国 上映时间: 2005

综合评分: ★★★★★★☆☆☆☆ 8.4

制作人员: 弗兰克·米勒 Frank Miller 导演, 罗伯特·罗德里格兹 Robert Rodriguez 导演, 杰西卡·阿尔芭 Jessica Alba 饰 Nancy, 布鲁斯·威利斯 Bruce Willis 饰 Hartigan, 克萊夫·歐文 Clive Owen 饰 Dwight, ...

简介: 影片改编自著名漫画大师弗兰克·米勒的经典同名漫画, 由三个故事组成, 都发生在虚拟的罪恶之城, 这里, 警察和暴徒已无分别, 暴力摧毁一切, 也保护一切。 故事一: 罪恶城 ..... [详细>](#)




**电影 低压槽：欲望之城 低压槽 2018**

地区: 中国大陆 / 中国香港 上映时间: 2018

综合评分: ★★★★★☆☆☆☆☆ 4.4

制作人员: 张家辉 Nick Cheung 导演, 张家辉 Nick Cheung 饰 于秋, 徐静蕾 Jinglei Xu 饰 惠美子, 何炅 Jiong He 饰 阿占, 余男 Nan Yu 饰 Jackie, ...

简介: 孤城, 一座浮华却堕落的罪恶之城, 这里充斥着金钱与利益的争夺。然而, 在这个亡命之徒的乐园里, 仍有少数人坚持着自己心中的正义..... 警方卧底于秋在上司阿占的运筹帷幄下, 多次成功打入当地帮派, 斡旋于多方黑色势力之中。于秋在黑帮生活中却好像渐渐迷失自我 ..... [详细>](#)



**电影 上帝之城 Cidade de Deus 2002**

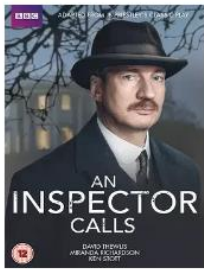
地区: 巴西 / 法国 上映时间: 2002

综合评分: ★★★★★★★★★★ 8.9

制作人员: 费尔南多·梅里尔斯 Fernando Meirelles 导演, 卡迪亚·兰德 Kátia Lund Co-director, 亚历桑德路·罗德里格斯 Alexandre Rodrigues 饰 Rucanã, Rorkat, 基卡·博罗索 Kika Boreto 饰 Leandro

图9 “罪恶+之城”搜索结果

搜索“罪恶-之城”，共找到约 1320 个结果




**电影 罪恶之家 An Inspector Calls 2015**

地区: 英国 上映时间: 2015

综合评分: ★★★★★★★★★★ 8.3

制作人员: 艾斯林·沃什 Aisling Walsh 导演, 大卫·休里斯 David Thewlis 饰 The Inspector, 米兰达·理查森 Miranda Richardson 饰 Sybil Birling, 苏菲·兰朵 Sophie Rundle 饰 Eva, 肯·斯托特 Ken Stott 饰 Arthur Birling, ...

简介: 亚瑟 (肯·斯托特 Ken Stott 饰) 是一位成功的资本家, 在不久之后即将加官进爵。希拉 (克洛伊·皮里 Chloe Pirrie 饰) 是亚瑟的女儿, 这位单纯善良的千金小姐在家宴中接受了男友杰拉德 (凯尔·索列尔 Kyle Soller 饰) 的求婚, 两人喜结连理。亚瑟和妻子对这门亲事显然十分满意, 而希拉的弟弟艾瑞克 (芬恩·科尔 Finn Cole 饰) 却显然并不怎么高兴。 ..... [详细>](#)



**电影 阳光下的罪恶 Evil Under the Sun 2018**

地区: 英国 上映时间: 2018

综合评分: ★★★★★★★★★★ 8.3

制作人员: 盖伊·汉弥尔顿 Guy Hamilton 导演, 彼得·乌斯捷诺夫 Peter Ustinov 演员, 玛吉·史密斯 Maggie Smith 演员, 科林·布莱克利 Colin Blakely 演员, 黛安娜·里格 Diana Rigg 演员, ...


简介: 阿伦娜·马歇尔 (黛安娜·里格 Diana Rigg 饰) 是一个美艳不可方物的女明星, 天性放浪的她从不掩饰对其他男性的喜爱, 也十分享受被男性追求的感觉, 从而彻底忽略了丈夫和养女的感受。帕特里克 (尼古拉斯·克莱 Nicholas Clay 饰) 是个血气方刚的英俊小伙, 虽然已经有了体弱多病的妻子克里斯汀 (简·伯金 Jane Birkin 饰), 但他依然难以抵挡向他暗送秋波的阿伦娜的诱惑。今年, 位于莱瑟库 ..... [详细>](#)

图10 “罪恶-之城”搜索结果

搜索结果

高级筛选

搜索"intitle:之城", 共找到约 3809 个结果



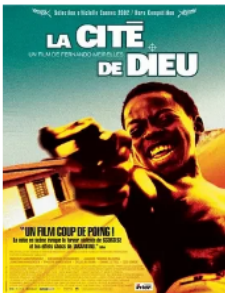
电影 罪恶之城 Sin City 2005

地区 美国 上映时间:2005

综合评分: 8.4

制作人员 弗兰克·米勒 Frank Miller 导演, 罗伯特·罗德里格兹 Robert Rodriguez 导演, 杰西卡·阿尔芭 Jessica Alba 饰 Nancy, 布鲁斯·威利斯 Bruce Willis 饰 Hartigan, 克莱夫·欧文 Clive Owen 饰 Dwight, ...

简介 影片改编自著名漫画大师弗兰克·米勒的经典同名漫画, 由三个故事组成, 都发生在虚拟的罪恶之城, 这里, 警察和暴徒已无分别, 暴力摧毁一切, 也保护一切。 故事一: 罪城 ..... [详细>](#)




电影 上帝之城 Cidade de Deus 2002

地区 巴西 / 法国 上映时间:2002

综合评分: 8.9

制作人员 费尔南多·梅里尔斯 Fernando Meirelles 导演, 卡迪亚·兰德 Kátia Lund Co-director, 亚历桑德罗·罗德里格斯 Alexandre Rodrigues 饰 Buscapé - Rocket, 莱安德鲁·菲尔米诺 Leandro Firmino 饰 Zé Pequeno - Li'l Zé, 菲利普·哈根森 Philippe Haagensen 饰 Bené - Benny, ...

简介 巴西里约热内卢的贫民窟, 这里是“上帝之城”, 更是魔鬼也会叹息着转身的地方。 阿炮 (Alexandre Rodrigues 饰) 带着我们到了这里, 他见证了这里二十多年来被残暴、贪婪、复仇、野心、背叛、掠夺所裹挟的混乱生活以及最终导致的一场灾难 ..... [详细>](#)



电影 暮光之城 Twilight 2009

地区 美国 上映时间:2009

综合评分: 7.5

制作人员 凯瑟琳·哈德威克 Catherine Hardwicke 导演, 克里斯汀·斯图尔特 Kristen Stewart 饰 Bella Swan, 罗伯特·帕丁森 Robert Pattinson 饰 Edward Cullen, 比利·伯克 Billy Burke 饰 Charlie Swan, 阿什丽·格林尼 Ashley Greene 饰 Alice Cullen, ...

简介 因为父母离异, 贝拉 (克里斯汀·斯图尔特 Kristen Stewart 饰) 随父亲来到小镇生活。在新学校里, 贝拉留意到一群举止诡异的怪人, 他们特立独行, 很是神秘。在实验课上, 她遇到了怪人中的一个——金发帅哥爱德华 (罗伯特·帕丁森 Robert Pattinson 饰), 他俊朗的外形与出众的谈吐令她着迷。在交往中, 贝拉渐渐对爱德华产生了感情, 但其他怪人对她心存戒备。与此同时, 小镇里接连发生死亡事件, ..... [详细>](#)

图 11 “intitle:之城”搜索结果

## 2.4 智能补全

### 2.4.1 简介

在搜索框中根据用户已输入的文字, 提示智能补全后的结果。

### 2.4.2 原理及实现

为了实现在搜索框中输入前缀自动补全的功能, 有很多可选的实现方案。其中最常见的使用正则匹配或是 PrefixQuery, 即使用 key\* 的形式来匹配关键词, 但是这样往往要对索引进行全文搜索, 其搜索效率在数据量大的时候非常低。

为了提高搜索效率，我们使用了 ES 中的 Completion Suggester 进行前缀匹配，它通过改造索引的结构来提升性能。在索引结构的 title 字段中，与前面的拼音一样，我们新加一个 suggest 域，并设置类型为 completion（经过测试研究，我们发现不能对 title 直接进行 completion 类型的设置，它的分词会导致该类型消耗约 20 倍的存储空间）。Completion Suggester 与 ES 中提供的其他 Suggester 相比，采用了不同的数据结构，索引并非通过倒排来完成，而是将分析过的数据编码成 FST 和索引一起存放，FST 有着非常高的数据压缩率和查询效率。对于一个开放状态的索引，FST 会被 ES 整个装载到内存里的，能够进行极快速度的前缀查找。

### 2.4.3 效果展示

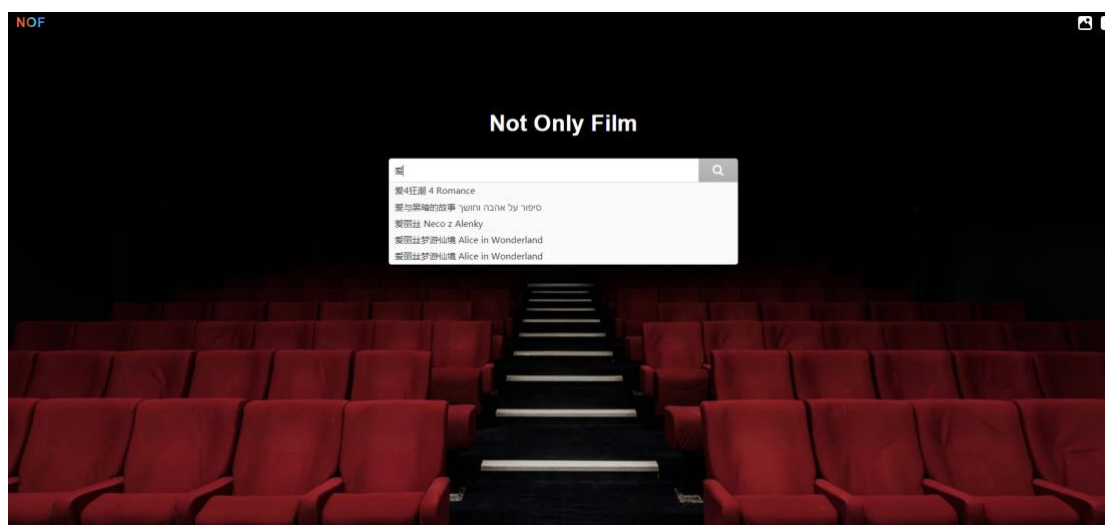


图 12 搜索框智能补全效果

## 2.5 搜索指数可视化

### 2.5.1 简介

通过折线图的形式展示最近 7 天的搜索指数，表现电影的热度趋势。

### 2.5.2 原理及实现

由于网站用户较少，目前采用爬取百度指数的方式获得搜索指数的数据。在第一次爬取数据时爬取近 14 天的数据，然后在之后的日子里每天启动一次脚本爬取前一天的数据，从而保证数据的时效性。

## 2.5.3 效果展示

近7天搜索指数



图 13 近 7 天搜索指数折线图





## 3.2 同类型电影对比

### 3.2.1 简介

对于一部电影，将其与同类型电影进行对比，得到不同维度上的数据，表现该类影片在同类型影片中的地位。

### 3.2.2 原理及实现

对于电影 A，其类型为 B，选择所有类型为 B 的电影，在维度 C 上进行升序排序，取得电影 A 的排名后计算其排名在同类电影中的分布位置，得到电影 A 好于“电影 A 排名/B 类型电影总数”B 类型片的数据分析结论。

按照如上逻辑，从电影评分、评分人数、近期热度三个维度对一个电影所属的所有类型进行同类推理得到分析结论。

- 电影评分：由于系统同时爬取了 IMDB 评分和豆瓣评分，所以在计算电影评分时采用评分人数进行加权后得到电影评分，计算公式为：

$$\text{电影评分} = \frac{\text{评分}_{\text{IMDB}} * \text{评分人数}_{\text{IMDB}} + \text{评分}_{\text{豆瓣}} * \text{评分人数}_{\text{豆瓣}}}{\text{评分人数}_{\text{豆瓣}} + \text{评分人数}_{\text{IMDB}}}$$

- 评分人数：由于系统同时爬取了 IMDB 评分和豆瓣评分，所以在计算电影评分人数时采用评分人数之和，计算公式为：

$$\text{评分人数} = \text{评分人数}_{\text{豆瓣}} + \text{评分人数}_{\text{IMDB}}$$

- 近期热度：取该电影近 14 日搜索指数的平均，计算公式为：

$$\text{近期热度} = \text{average}(\text{近 14 日搜索指数})$$

### 3.2.3 效果展示

#### 同类对比



图 15 同类型电影对比

## 3.3 猜你喜欢推荐

### 3.3.1 简介

系统提供推荐功能，根据用户的访问记录、电影的详细信息向用户推荐用户可能喜欢的电影。

### 3.3.2 原理及实现

推荐的参考依据：(1)近 10 次的电影访问记录；(2)同类型的电影；(3)相同电影搜索记录的用户搜索过的其他电影。综合三个方面的信息，最终为用户推荐他可能喜欢的电影。推荐电影展现在电影详情页，根据用户实时的访问记录，动态地推荐相似类型的电影。

### 3.3.3 效果展示



图 16 猜你喜欢

## 3.4 基于知识图谱的自动问答

### 3.4.1 简介

本系统构建基于知识图谱的自动问答，面向电影领域回答相关问题。

例如，演员的生日、出生地、英文名、出演的电影、合作的演员等信息，电影的上映日期、评分、类型等信息。

### 3.4.2 原理及实现

#### 3.4.2.1 知识图谱本体建模

本系统为面向电影领域的垂直搜索引擎，涉及的概念和范围是固定可控的，数据精度较高，因此构建领域知识图谱。本图谱是一个电影领域的知识图谱，采用自顶向下的方法来构建本体。

在 Protégé 中创建电影知识图谱的类。电影知识图谱中的类有电影、电影类型、电影人物等。其中电影、电影类型、电影人物是互斥的关系。电影人物下根据不同的派别，有子类如喜剧演员等。

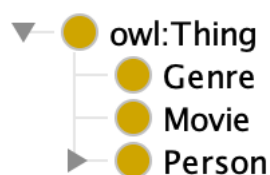


图 17 类

在 Protégé 中创建类的对象属性。“hasActedIn”表示某电影人物参演了某电影，“hasActor”表示某电影有哪些电影人物，“hasGenre”表示某电影有哪些电影类型。“hasActedIn”属于电影人物类，取值范围是 Movie，它的逆属性是“hasActor”。

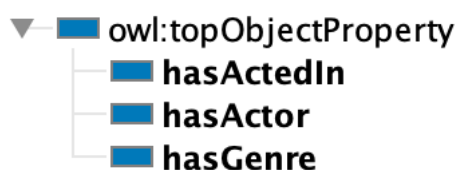


图 18 对象属性

在 Protégé 中创建类的属性。主要的属性有电影类型名字、电影简介、电影评分、电影上映日期、电影人物名字、电影人物生平等等。

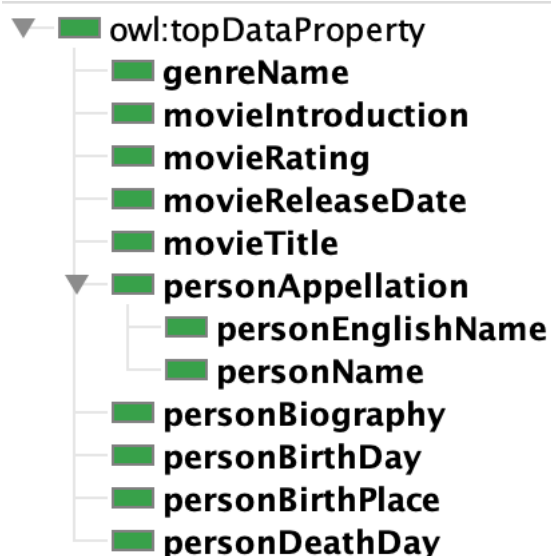


图 19 数据属性

最后建立完成的本体结构为：

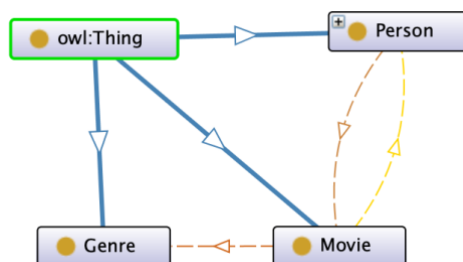


图 20 本体结构

### 3.4.2.2 数据准备

将从豆瓣和 TMDb 网站爬取获得的信息存入到关系数据库 MySQL 中。演员的基本信息包括姓名、出生日期、生平等。电影的基本信息包括名称、简介、评分、上映日期、类型等。采用数据库到三元组的直接映射的方式，使用 D2RQ、Jena 将 MySQL 数据库中的数据转换为三元组，存入 Fuseki 数据库中。

在 Fuseki 中存储准备好的三元组数据，并且配置配置文件，将我们设计好的本体文件放入 Fuseki 数据库的配置文件中，启动 OWL 推理机。从推理机中可以推理得到原本数据中未知的信息，如一部电影有哪些演员出演，某个演员一共演出了多少部电影。

### 3.4.2.3 KBQA

利用构建的电影知识图谱实现问答。由于面向电影行业，我们构建电影人物名字和电影名字的外部词典。使用 jieba 并且加载外部词典配合正则匹配完成分词和实体识别。使用 REfo 完成语义匹配，匹配完成后使用 SPARQL 向 Fuseki 服务器发送查询，打印最终结果。最终实现电影领域的简易的 KBQA。

### 3.4.3 效果展示



图 21 KBQA

## 3.5 关系图谱可视化

### 3.5.1 简介

为了提供用户更好、更直观的搜索体验，我们将电影详情、电影类型、电影人关系可视化，本系统构建了全部电影和单独电影的关系图谱。

### 3.5.2 原理及实现

从数据爬取中获得全部电影的信息，解析出合适的关系范围。对于全部电影，构建全部电影之间的关系。考虑到计算机渲染，本系统为每个电影解析适当大小的关系图谱，围绕一个电影构建该电影的电影类型、电影人与电影本身的关系，同时再添加电影人出演的其他电影的关系，一共是两层的关系。

关系范围用 JSON 格式存储，借助开源工具 InteractiveGraph，使用 JSON 绘制相应的关系图谱。

### 3.5.3 效果展示

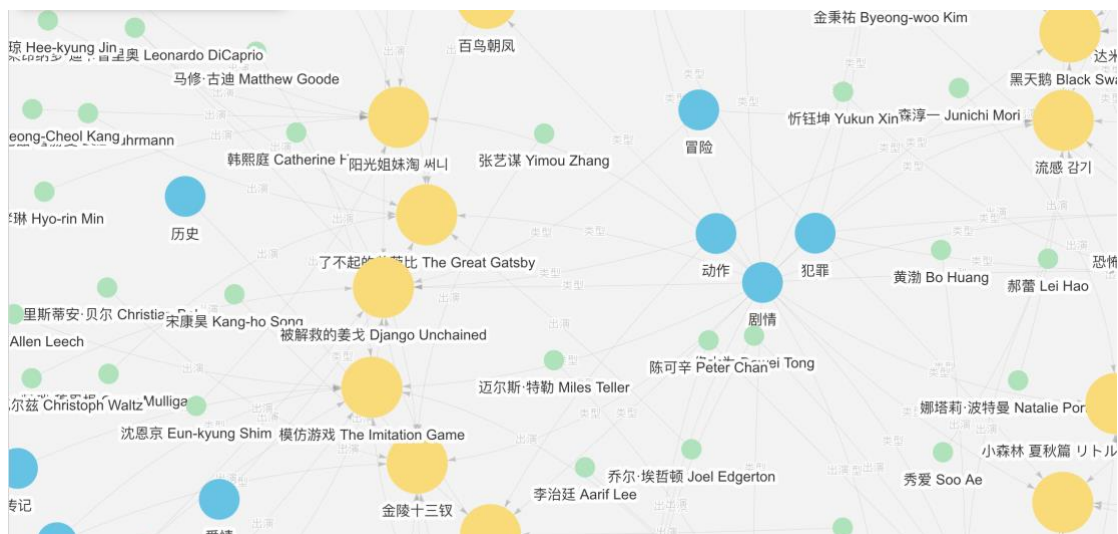


图 22 全部电影关系图谱

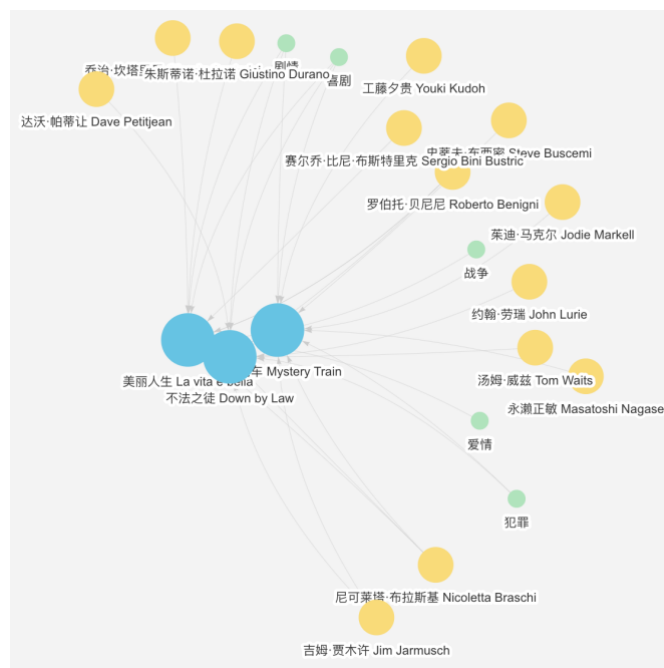


图 23 美丽人生的关系图谱