



知识工程大作业

学院名称	软件学院
专业名称	人工智能
姓名学号	胡强强 ZF1721201 龚凌云 ZF1721137 焦超 ZF1721209 杨帆 ZF1721349
指导教师	罗杰
得分系数	胡强强：25% 龚凌云：25% 焦超：25% 杨帆：25%
日期	2018年6月
GITHUB地址	https://github.com/Soulemp ty/-

目 录

1. 引言.....	1
2. 系统设计概述.....	1
2.1 系统总体设计.....	2
2.2 DBpedia 介绍.....	2
2.3 可视化数据爬取.....	3
3. 系统详细设计.....	3
3.1 系统框架搭建.....	3
3.2 系统数据处理.....	4
(1) 检索的主要查询语句如下所示.....	4
(2) 推荐的相关语句如下所示.....	5
3.3 结果可视化.....	5

1. 引言

随着信息技术的发展，每年都有很多的出版物发行，包括电影、游戏以及图书等多个种类，如何找到用户感兴趣的出版物这一问题尚未得到十分有效的解决。基于知识图谱的出版物检索和推荐系统旨在基于链接开发数据库和知识图谱的相关内容，为用户提供出版物的检索、推荐以及检索结果可视化统计等功能。

2. 系统设计概述

2.1 系统总体设计

系统总体包括系统检索、推荐以及结果统计可视化几个功能，项目整体是一个基于 Django 的网页，系统检索和推荐需要基于 dbpedia 数据库，通过 sparql 的查询语句完成相应功能，通过查询的信息，通过 scrapy 爬取查询照片，将照片信息返回到网页前端，进行可视化。

2.2 DBpedia 介绍

DBpedia 是一个特殊的语义网应用范例，它从维基百科(Wikipedia)的词条里撷取出结构化的资料，以强化维基百科的搜寻功能，并将其他资料集连结至维基百

科。透过这样的语义化技术的介入，让维基百科的庞杂资讯有了许多创新而有趣的应用，DBpedia 同时也是世界上最大的多领域知识本体之一，也是 Linked Data 的一部分，美国科技媒体 ReadWriteWeb 也将 DBpedia 选为 2009 年最佳的语义网应用服务。

DBpedia2014 版的资料集拥有找过 458 万的物件，包括 144 万 5000 人、73 万 5000 个地点、12 万 3000 张唱片、8 万 7 千部电影、1 万 9000 种电脑游戏、24 万 1000 个组织、25 万 1000 个物种和 6000 个疾病。其资料不仅被 BBC、路透社、纽约时报所采用，也是 Google、Yahoo 等搜索引擎检索的对象。

2.3 可视化数据爬取

系统使用 python 库种的 selenium 自动化测试工具，配和 ChromeDriver 来根据搜索的信息和推荐的信息来爬取出目标的图片返回到系统前端。

3. 系统详细设计

3.1 系统框架搭建

系统是一个基于 Django 的页面，网页主题包括检索、搜索结果及可视化展示、推荐结果及可视化展示几个部分。检索由检索框、搜索提示、检索按钮和检索类别几个组件构成，如图 1 所示。

请输入搜索内容

搜索

☐ 电影 ☐ 游戏 ☐ 图书

图 1 系统检索部分设计图

检索结果包括检索结果主要内容展示，一个实体关键字检索之后可能会返回多个不同类别的出版物信息，这里选取了第一条检索结果进行展示，展示内容包括出版物的图片、语言、国家、作者以及内容简介。然后返回其他的检索结果列表，列

表含有 5 条检索结果信息，并去重去掉查询的目标。统计可视化部分会先是 dbpedia 数据库种检索到的所有对应的实体类别和数量。

推荐部分根据检索的类别，在 dbpedia 中检索与待检索出版物类别相似的出版物加入到推荐列表并进行展示，并对相关的推荐指数进行显示。在获取到搜索结果后，根据数据库返回的电影信息，选择特征明显的特征，按照这些特征分别到 Depedia 中搜索数据。

- 电影方面会推荐相同导演的作品给用户
- 游戏方面会将相同的游戏开发团队的游戏推荐给用户
- 书籍推荐会根据同一作者的作品推荐给用户

3.2 系统数据处理

系统接受搜索框的输入内容后将其保存下来，然后通过查询语句进行检索，通过 SPARQLWrapper 工具完成 sparql 语句的查询，然后将查询结果保存至一个列表集合中，在页面上进行显示。

(1) 检索的主要查询语句如下所示

在获取用户输入的电影名称后，从数据库中获取相应的电影信息。选择特征明显的属性：

```

rdf:type dbo:Film;
rdfs:comment ?comment; （评论）
rdfs:label ?name; （名称）
dbp:language ?language; （语言）
dbp:country ?country; （国家）
dbo:director ?director; （导演）
dbo:starring ?actor; （演员）
dbo:wikiPageExternalLink ?wikiLink （维基百科信息）

```

当查询游戏的信息时候，就输入游戏的名称，从数据库中获取相应的游戏信息。

```

rdf:type dbo:VideoGame;
rdfs:label ?name; （名称）
rdfs:comment ?comment; （评论）
dbo:developer ?developer; （开发团队）
dbo:genre ?genre （类型）

```

当查询书的信息，就输入书的名称，从数据库中获取相应的书的信息。

```

rdf:type dbo:Book;
rdfs:label ?name; （名称）
rdfs:comment ?comment; （评论）

```

dbo:author ?author; （作者）
dbo:thumbnail ?thumbnail; （短文）
dbp:country ?country; （国家）
dbp:genre ?genre; （类别）
dbp:title ?title （主题）

(2) 推荐的相关语句如下所示

对结果进行推荐，电影用同一个导演，图书用同一个作者，游戏用同一个公司

```
movie_re = "" select distinct ?movie where{  ?movie dbo:director
dbr:%s  }LIMIT 5""""%(str(director[28:]))

game_re = "" select distinct ?game where{  ?game dbo:developer
dbr:%s  }LIMIT 5""""%(str(developer[28:]))

book_re = "" select ?book where{  ?book dbo:author dbr:%s  }LIMIT
5""""%(str(author[28:]))
```

3.3 结果可视化

结果可视化有三部分。

(1)输入要检索出版物的名称，会检索出该部分直接的简介信息，有关图片、语言、国家、作者、简介和相关链接。

(2)检索结果有三种类别，电影、图书以及游戏，可以通过复选框选择三个类别出版物的具体类别，页面显示出该出版物的类别，分别是 movie, game 和 book。其检索结果可视化分别查询的图片，作者，语言，国家等类别，和对应的简介和搜索相关推荐作品。

以下图 2，3，4 分别是电影、游戏、书籍的搜索和推荐结果。

电影名称: 阿凡达



导演: James_Cameron

国家: United_States

语言: English

演员: Stephen_Lang Zoe_Saldana Sigourney_Weave...

简介:

《阿凡达》（英语：Avatar）是2009年上映的一部美国科幻史诗式电影，由詹姆斯·卡梅隆撰写剧本并执导，主要演员有山姆·沃辛顿、西耶·沙达纳、雪歌妮·薇佛、黛雪儿·罗德里奎兹和史蒂芬·朗。电影设定于2154年，当时人类正在南门二恒星系生态茂盛的潘朵拉卫星上开采珍稀矿产难得素。采矿殖民地的扩张威胁到了当地部落纳美人的生存；纳美人是土生土长于潘多拉星球上的有梦和智慧能力的类人种族。电影的标题「阿凡达」，指的是经过基因改造而能为部分人类所控制的纳美人体。人类使用他们来同潘朵拉星上的原住民展开交流。阿凡达的开发始于1994年，卡梅隆当时为这部电影撰写了80页的剧本论述。柯麦隆原计划在1997年的《快活号》摄制完成后就开始制作这部电影，并打算在1999年上映，但最终未能如愿。柯麦隆解释说，当时的技术手段还不足以表现出他想像的电影场景。为电影中的外星人创造语言的工作开始于2005年夏天，柯麦隆也在2006年初开始发展剧本及其虚构世界。

推荐列表

泰坦尼克号



真实的谎言



深渊




图 2 电影检索推荐结果图

请输入搜索内容

搜索

游戏名称: 星际争霸II：虫群之心



开发者: Blizzard_Entertainment

游戏类型: Real-time_strategy


发布者: Real-time_strategy

简介:


《星际争霸II：虫群之心》（StarCraft II: Heart of the Swarm）是暴雪娱乐出品的科幻即时战略游戏《星际争霸II：自由之翼》的续篇，同时也是《星际争霸II三部曲》中的第二部，其第三部为《星际争霸II：虚空之遗》。本资料片主以异虫（Zerg）为主题，包含了约20个新战役任务，以及若干新兵种单位、科技和地图。虫族之心于2013年3月12日在美洲、欧洲、东南亚、台湾等地同步发售。（中国大陆为2013年7月17日）

推荐列表

魔兽世界：巫妖王之怒



星际争霸II：虚空之遗



暗黑破坏神III




图 3 游戏检索推荐结果图

1

书名: 悲惨世界



作者: Victor_Hugo

国家: France

语言: English

类型: Epic novel, historical fiction

维基链接: <http://www.theatlantic.com/unbound/essays/lesmisfa.htm>

简介:

《悲惨世界》（法语：Les Misérables，英语发音：/leɪˈmɪzəˈrɑːbəl/，另有中文译名《孤星泪》，原意为“悲惨的人们”、“可怜的人们”），是法国作家维克多·雨果（Victor-Marie Hugo）于1862年所发表的一部长篇小说，是19世纪最著名的小说之一。小说描绘19世纪初20年间几个法国人物的生活背景，涵盖拿破仑战争和1832年巴黎共和党人起义等政治现象叙述。故事的主线围绕主人公获释罪犯尚万强试图赎罪的历程。小说试图检视他的赎罪行为在当时的社会环境下的所造成的影响。这部宏大的小说，融进法国的历史，以及巴黎的建筑、政治、道德哲学、法律、正义、宗教信仰，检视善、恶和法律的本质，同样还有爱情与爱情的种类和本质。雨果的创作灵感来自一位真实的罪犯和警察，他把这个真实人物的性格分成故事中的两个人物。悲惨世界通过它不计其数的舞台和银幕的改编作品被世人所了解。最著名的改编作品是同名音乐剧。

推荐列表



图 4 书籍检索推荐结果