



## 知识工程课程期末大作业

学 院 名 称	软件学院
专 业 名 称	人工智能
姓 名 学 号	马 健 ZF1721248 殷晓婷 ZF1721406 刘子晗 ZF1721245 雷一明 ZF1721213
指 导 教 师	罗杰
得 分 系 数	马 健：26% 殷晓婷：26% 刘子晗：24% 雷一明：24%
日 期	2018 年 6 月

## 目 录

1. 引言.....	1
2. 系统设计概述.....	1
2.1 系统总体设计.....	1
2.2 DBpedia 介绍.....	1
2.3 echarts 介绍.....	1
3 系统详细设计.....	2
3.1 系统框架搭建.....	2
3.2 系统数据处理.....	3
3.3 结果可视化.....	4

## 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放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 javascript 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，系统的结果可视化通过 echarts 插件完成。

### 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。

DBpedia 2014 版的资料集拥有超过 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 种物种和 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 等搜寻引擎检索的对象。

### 2.3 echarts 介绍

ECharts，缩写来自 Enterprise Charts，商业级数据图表，一个纯 Javascript 的图表库，可以流畅的运行在 PC 和移动设备上，兼容当前绝大部分浏览器（IE6/7/8/9/10/11，chrome，firefox，Safari 等），底层依赖轻量级的 Canvas

类库 ZRender，提供直观，生动，可交互，可高度个性化定制的数据可视化图表。创新的拖拽重计算、数据视图、值域漫游等特性大大增强了用户体验，赋予了用户对数据进行挖掘、整合的能力。

支持折线图（区域图）、柱状图（条状图）、散点图（气泡图）、K线图、饼图（环形图）、雷达图（填充雷达图）、和弦图、力导向布局图、地图、仪表盘、漏斗图、事件河流图等 12 类图表，同时提供标题，详情气泡、图例、值域、数据区域、时间轴、工具箱等 7 个可交互组件，支持多图表、组件的联动和混搭展现。

## 3 系统详细设计

### 3.1 系统框架搭建

系统是一个基于 javascript 的网页，网页主体包括检索、搜索结果及可视化展示、推荐结果及可视化展示几个部分。检索由界面 logo、检索框、检索按钮、检索类别几个组件构成，如图 1 所示。



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，列表含有由 10 条检索结果信息，统计可视化部分会显示 dbpedia 数据库中检索到的所有对应的实体类别和数量。

推荐部分根据检索内容的类别，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。

### 3.2 系统数据处理

系统接收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，最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。

```
prefix foaf: <http://xmlns.com/foaf/0.1/>

Select distinct ?type ?value

where {

?s foaf:name ?sname .

?s name bif:contains name .

?s ?type ?value .

}
```

其中 **name** 代表要查询的出版物的名称，如：**Harry Potter**。

(2) 推荐的相关语句如下所示。

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX owl: <http://www.w3.org/2002/07/owl#>

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT COUNT(?movie) SAMPLE(?movie)

where {

dbr: name dct:subject ?o .

?movie dct:subject ?o .

FILTER (?movie != dbr: name) .

}

GROUP BY ?movie ORDER BY DESC(COUNT(?movie)) limit 7
```

我们取前 7 个推荐的电影并在页面上显示出来并可视化。

### 3.3 结果可视化

结果可视化有三个部分。

(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，有相关图片、语言、国家、作者、简介和相关链接，此部分效果如图 2 所示。



图 2 出版物相关信息

(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种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**，如图 3 所示。其检索结果可视化效果分别以三个进度条的形式显示，选取 echarts 中的条形图进行展示，可视化效果图左边名称代表所搜索的出版物种类，右边数字代表该种类出版物的数量，效果如图 4 所示。

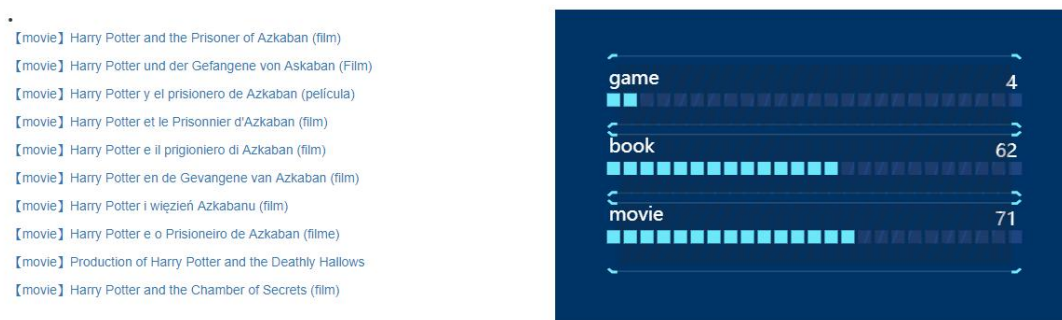


图 3 搜索出版物信息结果图

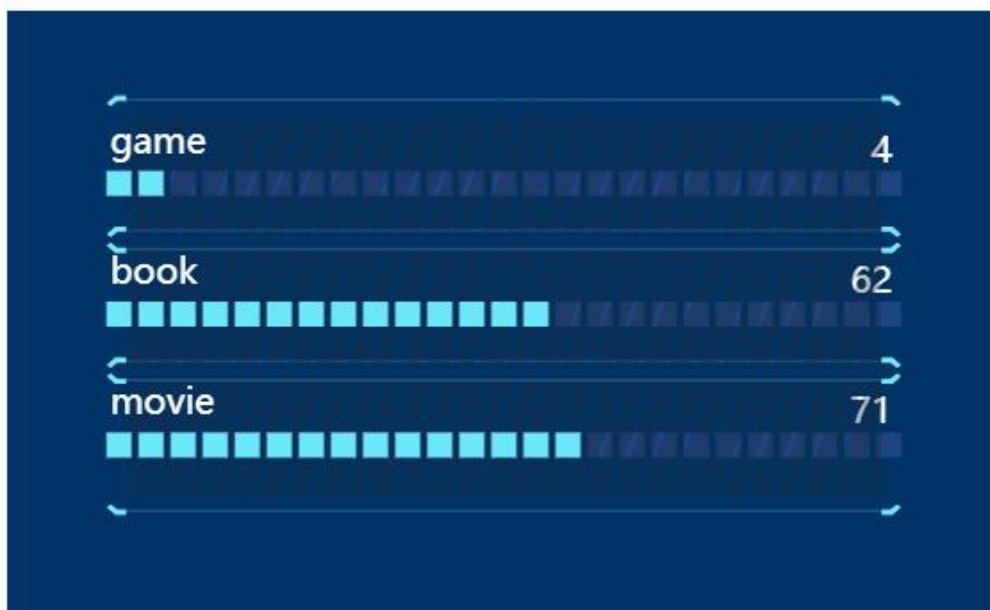


图 4 出版物检索结果统计图

(3) 推荐结果列表中，每条结果对应着不同的共享次数，共享次数越高推荐指数也高，每部电影的共享次数分别通过推荐查询结果获取，然后构成条形图的参数，将鼠标悬停在图上可以看到每种推荐的共享次数，可视化效果如图 5 所示。

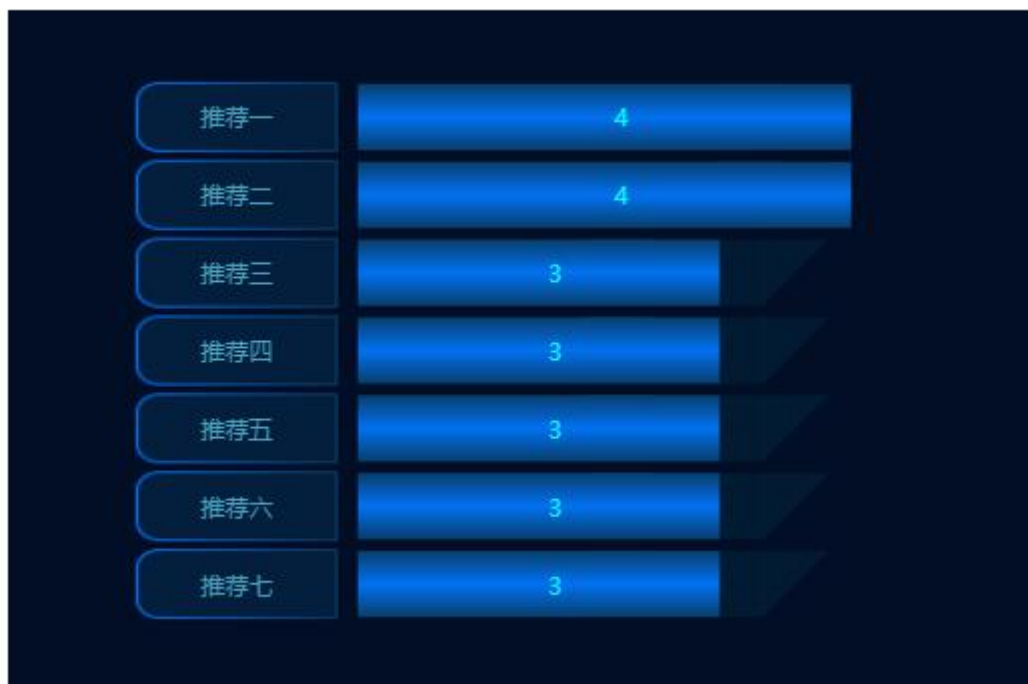


图 5 出版物推荐结果统计图