

一. “人物关系知识图谱”

主要功能：用户在搜索人时常常希望了解人物在社会或者在影视作品/小说中的社交网络等的关系，“人物关系图”通过收集数据、检索分析、展示等步骤为用户展示人物的关系，帮助用户更深入得了解搜索的人物。

1. 相关人物推荐

ItemCF 算法，以百度的用户近期历史搜索记录为数据（数据形式为“用户-搜索人物”对），衡量用户所搜索的人物之间的相关性，并且加入两个人物作为关键词同时出现在同一页面的频率作为衡量相关性的另一部分因素。设 C1 为 ItemCF 算法确定的人物间相关分数，C2 为同时出现频率确定的人物间的相关分数，则总的相关性分数 C：

C = α *C1 + β *C2

可以推测，α 值大的话用户的粘度和停留时间会增加，β 值大的话用户获取信息定位会更准确

2. 数据来源

用户查询人物关键词时，自动从百度百科的数据库中检索该人物的关系数据，如果百度百科中包含其与其他人物的关系数据，则从百度百科中抽取；如果百度百科不包含数据，则将该人物以及相关推荐人物推送至贴吧，从贴吧抽取关系

a. 百度百科：

在人物的百度百科中，寻找具有相同值的属性，通过简单的语义分析确定其关系。或者能够得到后台的信息更好。



b. 人物/小说/电影的贴吧：

在百科中编辑的人只在少数，而在特定人物或小说的贴吧中，大家的参与热情更高，了解程度也较好，所以更容易获得高质量的关系数据。在贴吧中，可以使用

人物一，人物二，关系

的格式收集人物具体关系，如“牛顿、爱因斯坦、同为物理学家”。通过 ItemCF 或者同时出现的概率之类的方法来找出高相关性的人物供给贴吧（当然贴吧也可以自己提出高相关性的人物），由贴吧确定人物的具体关系，如同学、夫妻。

难点:

- 如何设计贴吧收集数据的功能, 包括这个功能放置的位置、形式
- 如何对收集的数据进行审核。目前考虑可以由吧主审核, 或者通过投票、点赞的方式
- 数据库的结构
 - 如两个人可能有多种关系, 如何存储和展示?
 - 电视剧、书籍等可能需要知道此作品所有人的关系, 所以需要记录某本书包含那些人, 这些人同属一大类。所以可能需要为数据库增加一些字段。

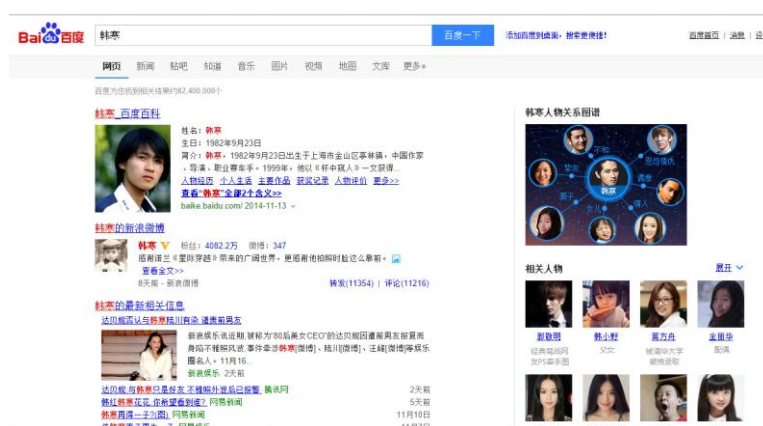
3. 收到用户搜索请求后的处理

如果用户的搜索关键词中存在数据库中的人名, 那么就触发这个人物关系网络的功能, 由后台检索数据库, 找到此人的关系, 按照固定的格式发送到前端。

4. 结果的前端展示

a. 展现方式

人物关系网络图和原有的“知心推荐”的列表的人物并不一定是一致的, 关系网络图更着重体现与用户所检索的人物有实际联系、关系的人物。因此要分开显示于原右侧推荐栏上方, 以缩略图形式展现数据。如现有的韩寒知识图谱:



b. 推荐理由

推荐理由直接在知识图谱上显示为二者的关系

5. 根据贴吧的关系反馈构建学习系统

以用户人工添加的关系标签作为反馈, 学习人物关系分类的模式 (待定)

6. 其他:

对于重名问题可以参考现在的搜索算法, 默认显示排名最高者 (百度百科如何实现排序), 对于人名+其他信息的搜索可以根据相关性进行排名。

二. “美食推荐”

主要功能：通过持续地分析用户的搜索习惯、饮食习惯，根据评价等信息为用户推荐可能喜欢的餐厅

1. 什么是用户喜欢的餐厅：

用户通常在哪些餐厅用餐，这些餐厅有一些评价。将这一餐厅与用户所在位置周围的餐厅的评价进行对比，找到最相似的餐厅作为最靠前的推荐，之后才是按照评价的星级、距离等进行排序。

2. 如何确定用户通常在哪些餐厅用餐：

一种可能的方法是在用户搜索餐厅后分析用户的位置信息，如果位置信息定位在某个餐厅，则认为用户在此餐厅用餐，如果用户在用餐后未给出负面评价，则认为用户喜欢。另外也可以请用户选择就餐地点和评价。

3. 如何对比几家餐厅的评价获得相似性。