**专 利 申 请 技 术 交 底 书**

**专利卷号（法务部填写）：**

**如有疑问，请咨询集团法务部。**

**1688事业部/国际事业部/商业智能部1688DW/ 技术质量部1688技术质量员工请联系李海波，分机：39508，Email:** [ricky.lihb@alibaba-inc.com](mailto:ricky.lihb@alibaba-inc.com)**；**

**淘宝/天猫事业部/聚划算事业部/航旅事业部/淘点点事业部/技术质量部共享业务员工请联系季江胜，分机：39694，Email:** [jiangsheng.jijs@alibaba-inc.com](mailto:jiangsheng.jijs@alibaba-inc.com)**；**

**一淘及搜索事业部/无线事业部/来往事业部/旺旺与客户端事业部/阿里妈妈事业部/技术质量部阿里妈妈技术/数字娱乐事业部员工请联系宋家扬，分机：39832,Email:** [jiayang.songjy@alibaba-inc.com](mailto:jiayang.songjy@alibaba-inc.com)**;**

**小微金服员工请联系潘葚（周泰），分机：37079，Email:** [shen.pans@alibaba-inc.com](mailto:shen.pans@alibaba-inc.com)**;**

**阿里云事业部/云OS事业群/集团技术保障/数据平台事业部/共享业务事业部/菜鸟/技术质量部中间件及大数据员工请联系岳强，分机：39750，Email:** [qiang.yueq@alibaba-inc.com](mailto:qiang.yueq@alibaba-inc.com)**。**

**以下内容请发明人填写：**

**发明名称：\_\_\_\_一种针对知识性购物query的实体挖掘和推荐装置 \_\_\_**

**发明人姓名：\_\_\_\_\_谢朋峻、周鑫、郎君\_\_\_\_\_\_\_**

**发明人花名：\_\_\_\_\_程辰、旷谷、上乘\_\_\_ \_\_**

# 发明人工号：­ \_\_\_\_051276、078503、071372\_\_\_\_\_\_\_

**第一发明人身份证号码（**必填项，针对中国发明人**）：\_\_51132319871028211X \_\_\_\_\_\_\_\_\_**

**技术联系人：\_\_\_程辰\_\_\_**

**电话：\_\_\_\_ 12849 \_ \_**

**E-MAIL：\_\_\_\_\_\_\_\_ chengchen.xpj@taobao.com \_\_\_\_\_\_\_\_**

**所属公司/部门\_\_\_\_一淘及搜索事业部-搜索和推荐算法研发部-新方向探索-知识库**

**注：1、以上内容均需要填写。**

**2、“发明人花名”一栏中，淘宝发明人必填。**

**术语解释：**

知识性购物query：

我们把电商query分为两种，一种是普通购物性query，比如“连衣裙”、“电视机”；另外一种是知识性购物query，它的意图是希望获得一种知识性的导购答案。比如“送给小孩的满月礼物”、“含蛋白质高的食物”、“杭州特产”等。

1. **本发明是否已在或计划在产品中实施？如果是，请列出产品名。是否已经或计划对外书面公开？如果是，请具体说明。**

计划实施。

名称: 知识图谱combo。即将在搜索结果页导航区呈现。

1. **本发明是否有敏感的部分不适合作为专利申请公开？例如，会不会对黑客或竞争对手提供优势？**

没有敏感地方。

1. **本发明要解决的技术问题是什么？**

针对知识性购物query，从社区数据挖掘实体作为答案推荐给用户，解决目前知识性购物query导购信息较差的问题。

**4、详细介绍技术背景,并描述已有的与本发明最相近似的实现方案**（包括两部分：1、作为本发明基础的且帮助理解本发明公知技术内容；2、与本发明最接近的已有技术方案的说明（对于方法，应说明现有方法的步骤，对于装置，应当说明结构组成及其关系））

随着电子商务应用的日益普及，网络购物慢慢融入普通用户的衣食住行中。搜索作为很多用户习惯性的购物入口，大家会在搜索框输入各种感兴趣的query。在用户输入query后，购物网站会提供相关的导购信息帮助用户明确用户意图。常用的搜索结果页导购方式有两种形态：

1. 导航

导航区通过筛选的方式让用户一步一步明确需要购买的商品，是帮助用户确定购物意图的一种有效方式。

# CN 103218719 A “一种电子商务网站导航方法及系统”。

通过汲取类目点击导航和类目商品数量导航的精华，综合考虑query关键词对应的点击、购买等历史因素，以及查询词相关的商品数量信息等，提供与搜索意图最相关的类目或属性，最终以导航的形式帮助用户明确用户意图。

1. 相关搜索

相关搜索是在用户搜索某query后，提供跟当前query相似或者相关的query供用户跳转。

CN 103279486 A “一种提供相关搜索的方法和装置”。

通过将与当前query共现于同一会话的其他query构成当前query的候选推荐项；再根据相似度对候选推荐项进行聚类，就得到当前query的候选推荐簇。在线推荐的时候，结合输入query的语义得到推荐的query簇，再根据各个簇内候选项的搜索次数，最终将query推荐给用户。

**5、以因果关系推理的方式推导出现有技术的缺点是什么？针对这些缺点，说明本发明的目的及能够达到的技术效果。**（现有技术的缺点是针对于本发明的优点来说的，本发明不能解决的缺点不必写）

对普通的购物明确型query，导航和相关搜索都能提供良好的导购信息给用户。但是对于知识性query，他们都不能很好的满足用户的意图。

1. 商品导航的缺点

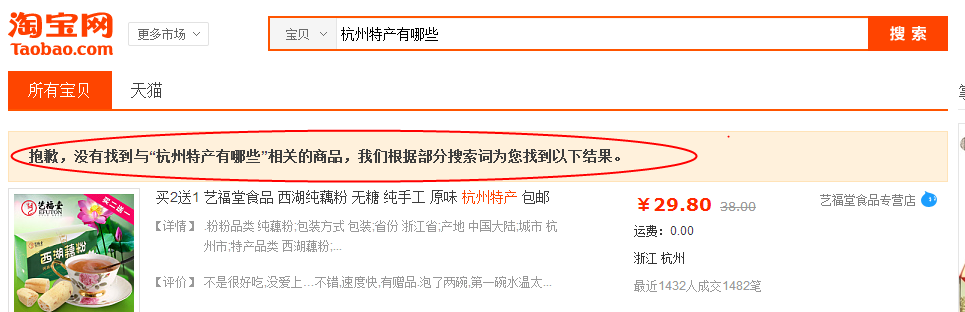
当前的商品导航本质是通过query关键字召回商品，再根据用户对召回商品集合的CPV（类目、属性、属性值）的点击反馈来计算不同的CPV的重要度，根据重要度推荐给用户。

这种方式的缺陷是它完全依赖召回商品的集合和商品自身的类目属性体系。当知识性query较长导致召回商品较少或者商品类目属性较宽泛时，导航区提供的信息导购性就很差。如图1，图2下：

图1



图2



1. 相关搜索的缺点

相关搜索的推荐候选项来自于用户输入的query，正因为此，它受制于用户认知。当搜索知识性query时，相关搜索呈现的都是类似的query，不能满足用户获取答案的需求。如下图3.



本发明采用信息抽取的方式，先识别知识性query，再对与知识性query相关的外网社区数据进行信息抽取，挖掘出想要的实体作为答案沉淀到知识库。用户线上搜索相关query时，基于知识库推荐答案给用户，线上作用形式类似下图4，图5

图4



图 5



当用户点击推荐的标签时，就会跳转到标签对应的新query，这也是跟导航很大的不同。通过这种方式，不仅能满足用户query的意图，还能更好的引导用户的购物行为。

**6、本发明技术方案的详细阐述，应该结合流程图、原理框图、电路图、系统结构图进行说明**（发明中每一功能的实现都要有相应的技术实现方案；所有附图都应该有详细的文字描述；方法专利都应该提供流程图，并提供相关的系统装置。5.1 本部分为专利申请最重要部分，需要详细提供；5.2 发明必须是一个技术方案，不能只有原理，也不能只做功能介绍；5.3 对于软件、业务方法，要提供流程图；5.4 必须结合流程图、原理框图、电路图、系统结构图等附图进行说明。）

系统框图：

****

基本流程：

1. **离线计算部分**

模块1 知识性query识别模块

我们通过对电商query进行分析总结，提炼出不同知识性query的类型和模板，提炼的query模板类似于：

1. 地名+“特产” ：表示希望获取某地的特产知识
2. “送”+ 称呼+“礼物”：表示希望获取送礼的导购知识
3. 品类词 +“品牌”：表示希望获取某品类的畅销品牌
4. 品类词 + “配件”：表示希望获取某品类的其它配件

基于大量类似的模板，我们可以扩展很多知识性query。

模块2 社区数据抓取模块

采用通用的爬虫技术，在社区网站（比如百度知道、淘宝问答）我们抓取知识性query的搜索结果数据，并对爬取的网页数据进行解析，我们不光解析社区数据的文本内容，也会把数据文本的编辑者、支持数、反对数等信息解析出来。抓取的结果数据示例如下：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Query | 来源网站 | 回答文本 | 作者 | 支持数 | 反对数 |
| 送给男朋友的礼物 | 百度知道 | 围巾,因为你送他的时候要对他这么说,我要用围巾把你围住因为你是我的一辈子. | chengchen | 2 | 0 |
| 送给男朋友的礼物 | 搜搜问答 | 送礼物给男人是一件很头痛的事,那些什么都没有的男人,你根本不会爱上他,什 么都有的男人,你又可以送什么给他呢? 衣服、领带、皮带、手表、公事包 | Kuanggu | 0 | 0 |
| 送给男朋友的礼物 | 淘宝问答 | 钱包、皮带、手表 | shangcheng | 1 | 0 |

模块3 答案生成模块

有了每个query结构化好后的文本答案数据，还需要从文本答案中抽取需要的品类实体以及对实体打分。

A)实体抽取

从自由文本中识别实体的方法很多，比如基于规则的方法、基于隐马尔科夫模型的方法、基于条件随机场的方法等。本文采取基于规则的方法。从文本答案抽取的实体会有很多类型，由于本文主要解决知识性购物query的导购问题，只需要将品类实体筛选出来。结果样式如下：

|  |  |  |
| --- | --- | --- |
| Query | 回答文本 | 实体 |
| 送给男朋友的礼物 | 围巾,因为你送他的时候要对他这么说,我要用围巾把你围住因为你是我的一辈子. | 围巾 |
| 送给男朋友的礼物 | 送礼物给男人是一件很头痛的事,那些什么都没有的男人,你根本不会爱上他,什 么都有的男人,你又可以送什么给他呢? 衣服、领带、皮带、手表、公事包 | 衣服、领带、皮带、手表、公事包 |
| 送给男朋友的礼物 | 我送男朋友送过钱包、皮带、手表，楼主可以参考 | 钱包、皮带、手表 |

B）实体分数计算

有个每条query对应的文本答案，有了每个文本答案对应的实体，就可以结合文本答案网站的质量、答案本身的支持度(支持数-反对数)来对实体进行投票算分。对于某条query A，实体entity1的分数公式如下：

其中

表示entity1是否在网站i的答案j中出现，出现则为1，不出现则为0。

表示网站i的权重，这个可以通过人工指定或者基于pagerank算法得到。

表示文本答案j的权重，这个可以通过答案的支持和反对数得到，默认为1。

通过以上2步，我们可以得到query的实体答案及分数，形式如下：

|  |  |
| --- | --- |
| Query | 实体及分数 |
| 送给男朋友的礼物 | 手表:55;钱包:46;打火机:32;腰带:22;围巾:22;剃须刀:20;手链:18;皮带:18;领带:18。。。 |
| 宝宝满月礼物 | 长命锁:28;奶粉:28;红包:27;手镯:16;银手镯:13;婴儿床:8;银锁:7;推车:7;银镯子:7;尿不湿:7;鞋子:6;帽子:6;尿布:6 |

模块4 人工review模块

抽取实体过程中难免会有badcase,为了确保推荐数据的准确性，对于知识性query的答案，还需要人工来check，确保没问题后融入知识库。

**二．在线应用部分**

如框图所示，用户通过终端向在线服务器提交搜索请求时，如果搜索的关键字命中知识库中已有的知识query，就直接将对应答案以标签的形式在导航区呈现给用户。用户可以点击标签继续购物行为。

**7、在实施本发明中，是否有某种计算机语言，软件或硬件能够达到最佳的实施效果？**

离线数据的统计牵涉到大规模的数据处理，所以需要大规模并行计算的需要，需要云计算平台。同时需要支持实时查询的keyvalue系统。

**8、本发明的关键点和欲保护点是什么？**（请提炼出本发明的技术创新点，以提醒专利代理人注意，便于其撰写权利要求书）*。*

针对购物知识性query，从社区数据挖掘实体作为答案推荐给用户，引导用户更好地购物。

**9、能实现本发明目的的其他替代方案**（尽量多列举可实现本发明的的其他替代方案，以便充分公开本发明）

注意：感谢您对阿里巴巴集团专利工作的支持和所付出的辛勤工作，我们将按照以下条件对您的提案进行审核，通过审核后我们将按照您所属BU的《专利奖励办法》中的规定，发放相应的奖励。

审核条件：

-是基础技术还是周边技术；

-是否是公司重点项目所涉及的技术；

-是否解决了业界重大的技术问题；

-是否业界有近似产品或技术；

-是否代表业界未来发展方向；

-是否易于发现竞争对手的侵权行为及证据；

-是否满足专利法的相关规定。