**专 利 申 请 技 术 交 底 书**

**\*本专利申请技术交底书是阿里巴巴集团及其关联公司的发明人为申请专利所填写的发明内容资料，该专利申请技术交底书仅供阿里巴巴集团及其关联公司内部使用以及提供给与其合作的专利代理机构撰写专利申请文件所使用。请相关人员遵循商业行为准则、保密协议以及职业道德规范等，禁止向非相关人员传播、披露该技术交底书所涉及的内容。**

蚂蚁金服平台产品技术部/微贷事业部/网商银行员工请联系潘葚（周泰），分机：39975，Email：[shen.pans@alibaba-inc.com](mailto:shen.pans@alibaba-inc.com)；

蚂蚁金服国际事业部/国内事业群/无线事业部/金融&理财&保险事业部员工请联系谭磊（炼城），分机：42658，Email： [allen.tl@alibaba-inc.com](mailto:allen.tl@alibaba-inc.com)；

蚂蚁金服大安全员工请联系吴菊华（沈岸），分机：13888，Email： juhua.wjh@alibaba-inc.com；

淘宝/天猫/聚划算/航旅事业部/淘点点/菜鸟/城市生活事业部/商家业务事业部员工请联系季江胜，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/旺旺/阿里妈妈/数娱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)；

阿里云事业群产品技术部/网站运营及集团上云事业部/数据开发平台/计算平台员工请联系李海波，分机：39508，Email： ricky.lihb@alibaba-inc.com；

阿里云事业群飞天/CTO/服务支持部与海外业务事业部/数据引擎/技术保障员工请联系岳强，分机：39750，Email： [qiang.yueq@alibaba-inc.com](mailto:qiang.yueq@alibaba-inc.com)；

安全部员工请联系李想（劳动），分机：43307，Email： laodong.lx@alibaba-inc.com；

云OS事业群/共享业务事业部（除国际UED外）员工请联系胡娟娟（涓月），分机：75242，Email：[juanjuan.hjj@alibaba-inc.com](mailto:juanjuan.hjj@alibaba-inc.com) ；

1688事业部（B2B-CBU）/国际事业部（B2B-ICBU）/共享业务事业部国际UED/国际B2C员工请联系邓荣（朱公），分机：75247，Email：[dengrong.dr@alibaba-inc.com](mailto:dengrong.dr@alibaba-inc.com)。

**专利卷号（阿里巴巴集团法务部填写）：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**以下内容请发明人填写：**

**发明完成日（或者方案构思完成的日期）：\_\_\_\_\_\_\_20160203\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**发明完成地：\_\_\_\_\_\_\_\_\_浙江省杭州市\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**发明名称：\_\_\_\_\_一种基于场景知识库的关联搜索词的推荐方法和装置\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**发明人姓名：\_\_谢朋峻、Qiu Long、孙亢、郎君\_\_\_**

**发明人花名：\_\_程辰、认意、华胄、上乘\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**发明人工号：\_051276、85274、58597、71372\_\_\_\_\_\_\_**

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

**技术联系人：\_\_\_\_\_\_\_程辰（谢朋峻）\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**电话：\_\_\_\_\_\_\_\_\_15988480878\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**E-MAIL（请填写工作邮箱）：\_\_\_\_\_chengchen.xpj@taobao.com\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**所属公司/部门：\_\_\_搜索事业部-Data Alibaba-语言技术\_\_\_\_**

**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。**

**专利申请技术交底书正文**

**术语解释：**

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。**

已经计划实施，产品名字：一站购。

1. **阿里集团哪些竞争对手的业务/产品和本发明相关？请列出竞争对手的名称和相关业务/产品的名称（如有多个请一并列出，此项为必填项）。**

无

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？**

无

1. **本发明要解决的技术问题是什么？**

基于场景知识库，识别用户当前query背后的场景意图，向用户推荐场景下关联的搜索关键词。

1. **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））

当前的搜索系统下，对用户推荐搜索关键词的方法一般有三种：

1）下拉推荐



图-1 下拉推荐的搜索词推荐

2）相关搜索



图-2 相关搜索中的搜索词推荐

3）基于知识图谱的宽泛query的推荐，见本人2014年申请的专利《一种针对知识性购物query的实体挖掘和推荐装置》

1. **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）

方法1）下拉推荐是基于统计大量用户历史搜索关键词，在线根据输入关键词进行前缀匹配，推荐匹配的最热门的相关搜索词，如图-1所示，用户搜索“宝宝推车”后，推荐的是“轻便折叠”、“玩具”等对当前query做细化的搜索词。

方法2）相关搜索是对大量用户在session内query变换做统计，得到搜索某个query 后最可能搜索的其他query。如图-2所示，

方法3）是对知识型宽泛query，基于已有的知识图谱，对当前query进行细化；比如用户搜索“杭州特产”，可以推荐“西湖龙井”、“萧山萝卜干”等

方法1）和方法2）都是统计历史用户行为，来进行搜索词推荐的方法，它往往受制于用户的认知，具有马太效应；

方法3）是对知识型query基于知识图谱进行解释和细化，它能给用户很好的体验，但是它面向的只是“北京特产”、“送男友礼物”这种宽泛的知识型query。

上面三种方法都是对用户搜索的query进行细化，缺乏对用户query意图的分析和扩展。

用户搜索一些query的时候，从query中我们可以分析出用户当前所处的人生阶段或者爱好，从而可以向用户推荐当前场景下的其他关联的搜索关键词。

比如用户搜索“宝宝推车”，我们可以预测用户很大概率处于“宝宝刚出生”的人生阶段，搜“宝宝推车”是想购买“婴儿出行用品”，我们推荐给用户“婴儿出行必备”，如图-3所示。



图-3 婴儿出行必备

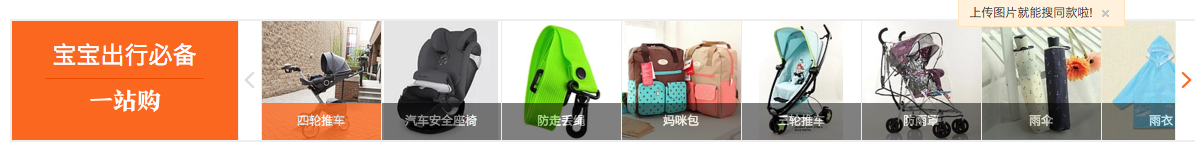
用户点击后，我们以分楼层的方式推荐给用户“婴儿出行必备”相关联的“汽车安全座椅”、“宝宝背带”、“宝宝腰凳”等多个搜索关键词。如图-4所示

图-4 宝宝出行相关的搜索关键词

再比如用户搜索“登山帐篷”的时候，我们可以预测用户很大概率是一个“登山爱好者”，我们可以推荐“登山装备”给用户，如图-5所示。



图-5 登山装备

用户点击后，我们以分楼层的方式推荐给用户“登山必备”相关联的“防潮垫”、“双肩背包”、“登山杖”等关联的搜索关键词。如图-6所示。

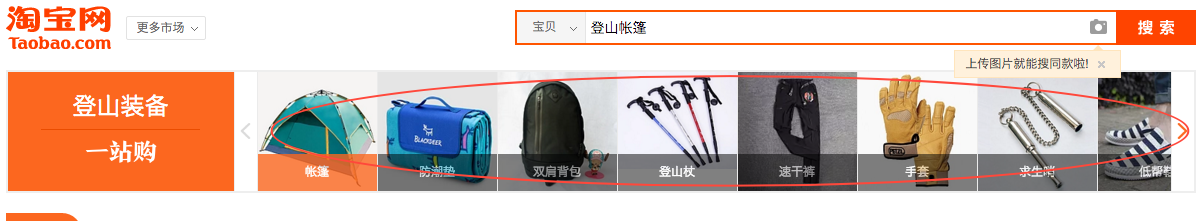
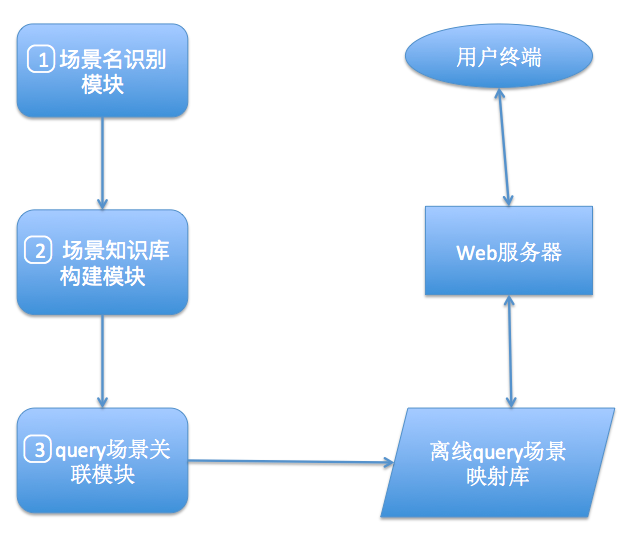


图-6 登山装备关联的搜索词

如上述的方式，我们独创性地构建电商场景知识库，基于知识库识别用户的query意图，推荐场景关联的搜索词给用户。

1. **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：



1. **离线部分**

**模块1 场景名识别模块**

为了构建场景知识库，第一步需要识别出场景名。场景名的获取来源是用户每天的上千万的搜索关键词。场景名识别是要从搜索query中识别出类似“宝宝出行”、“卧室装修”、“骑行”、“乒乓球用品”等描述某个兴趣爱好或者阶段需求的关键词。

识别场景名的方法包括以下几步：

1. 获取宽泛query构成场景名候选集合。宽泛query主要考虑query在商品类目和商品属性上的点击分散性，分散性比较高的query留下来作为场景名候选集合；
2. 构建场景名识别的机器学习分类器

首先，需要人工来标注一定数据量的“场景名query”和“非场景名query”作为分类的样本

然后，计算query的文本特征包括“分词词频”、“分词词性”等，统计特征包括query的点击商品的类目、query的成交转化率等。

3）有了步骤2）的分类器，我们对步骤1）的所有候选场景名进行预测，获取最终可用的“场景名”列表。

**模块2 场景知识库构建模块**

有了场景名后，我们需要挖掘场景名下包含的品类。比如对于场景名“骑行”，我们需要挖掘出骑行需要用到的“口罩”、“水壶”、“头盔”、“骑行服”等。从而构建好“骑行”知识库。

知识库的构建步骤分为：

1. 计算所有场景名query一年内点击、购买的商品；
2. **对所有商品**进行打标,识别出商品的核心产品词。

### 比如对于商品“正品SAHOO防风头盔 冬季自行车 山地车骑行装备 单车头盔”，通过打标和产品词识别，我们能抽取出核心产品词“头盔“；

1. 统计每个场景下query点击、购买的所有商品，提取所有核心产品词和对应的频度，就能得到表-1

|  |  |
| --- | --- |
| 场景名 | 点击购买的核心产品词集合 |
| 骑行 | 车灯：19000；骑行服：6900；头巾：5699；腿套：3843；头盔：2344;… |
| 滑雪 | 滑雪衣：6500；滑雪板：4311；滑雪裤：3222；手套：2333 |
| 。。。 |  |

表1 统计场景query覆盖的核心产品词

1. 对每个场景，截取排名前10的核心产品词；沉淀到“场景知识库”，如图-7

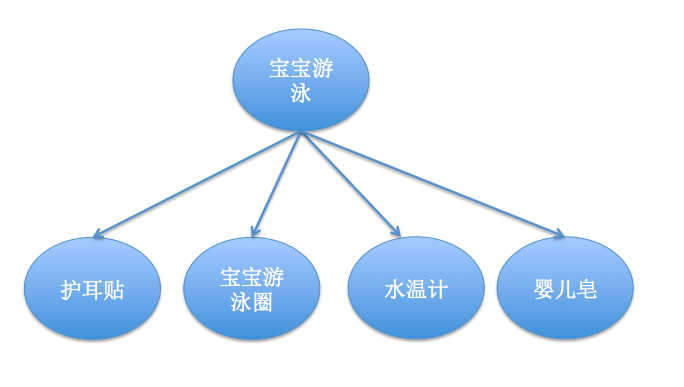


图7 宝宝游泳场景知识库

**模块3 query场景关联模块**

有了类似图-7的场景知识库，我们就可以针对用户具体的query来关联场景，并基于场景来推荐关联的搜索词了。

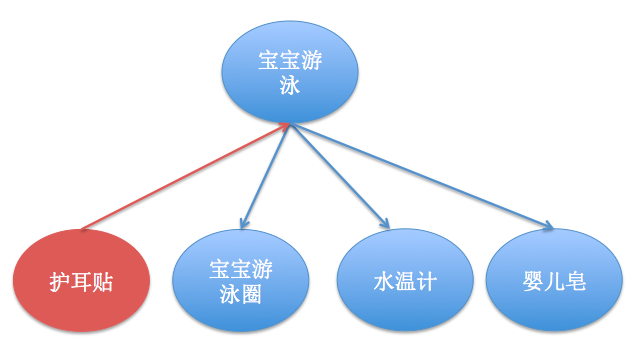


图8 query关联知识库

如图-8所示，当用户搜索“护耳贴”的时候，基于知识库的匹配，我们能预测用户很大可能是在购买“宝宝游泳”用品，我们可以推荐相关的“宝宝游泳圈”、“水温计”、“婴儿皂”给用户。从而得到离线的query场景映射表，如表2

|  |  |  |
| --- | --- | --- |
| 搜索query | 映射的场景 | 推荐的关联搜索词 |
| 护耳贴 | 宝宝游泳 | 宝宝游泳圈、水温计、婴儿皂。。 |
| 台球杆 | 台球装备 | 台球、批头、巧克粉、手套。。 |
| 春联 | 春节年货 | 鞭炮、窗花、红包。。 |

表2 离线场景query映射表

1. **在线部分**

用户在线搜索的时候，根据搜索的query访问web服务器；web服务器根据query去匹配“离线场景query映射表”，如果能命中，就返回“对应的场景”；用户点击对应场景后，返回“场景下搜索词”；

1. **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。

1. **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）

基于场景知识库，识别用户当前query背后的场景意图，向用户推荐场景下关联的搜索关键词。

1. **能实现本发明目的的其他替代方案**（请尽量多列举可实现本发明目的的其他替代方案，以便充分公开本发明，并防止别人轻易绕开我们的专利，例如，假如你发明了一种左右滑动的解锁方式，请把上下滑动的解锁方式、圆弧轨道滑动解锁方式等方案也补充在此处）
2. **本发明是否涉及软件开源？**（若涉及软件开源，请阐明是哪个开源项目及软件开源时间）

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，在您的专利提案通过审核正式递交至专利局后及该专利获得授权后,我们将按照阿里巴巴集团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。