



(12) 发明专利申请

(10) 申请公布号 CN 102360358 A

(43) 申请公布日 2012. 02. 22

(21) 申请号 201110295552. 6

(22) 申请日 2011. 09. 28

(71) 申请人 百度在线网络技术(北京)有限公司

地址 100085 北京市海淀区上地十街 10 号

百度大厦

(72) 发明人 傅强 江申

(74) 专利代理机构 深圳市威世博知识产权代理

事务所(普通合伙) 44280

代理人 何青瓦 李庆波

(51) Int. Cl.

G06F 17/30(2006. 01)

G06Q 30/00(2012. 01)

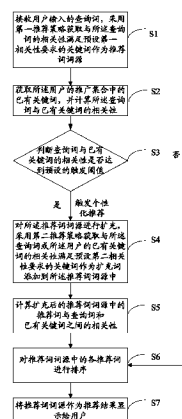
权利要求书 3 页 说明书 13 页 附图 2 页

(54) 发明名称

关键词推荐方法及系统

(57) 摘要

本发明提供了一种关键词推荐方法及系统,该方法包括:接收用户输入的查询词,采用第一推荐策略获取与所述查询词的相关性满足预设第一相关性要求的关键词作为推荐词词源;计算所述查询词与已有关键词的相关性;判断所述查询词与已有关键词的相关性是否达到预设的触发阈值,若是,则触发个性化推荐,若否,则得到关键词的推荐结果;其中,所述触发个性化推荐具体包括:对所述推荐词词源进行扩充,采用第二推荐策略获取与所述查询词或所述用户的已有关键词的相关性满足预设第二相关性要求的关键词添加到所述推荐词词源中。相较于现有技术,本发明在推荐时引入更多用户信息,判断用户需求,防止出现盲目的个性化推荐,同时又保证推荐结果的多样性。



1. 一种关键词推荐方法,其特征在于,包括:

接收用户输入的查询词,采用第一推荐策略获取与所述查询词的相关性满足预设第一相关性要求的关键词作为推荐词词源;

获取所述用户的推广集合中的已有关键词,并计算所述查询词与已有关键词的相关性;

判断所述查询词与已有关键词的相关性是否达到预设的触发阈值,若是,则触发个性化推荐,若否,则将所述推荐词词源作为推荐结果推荐给所述用户;

其中,所述触发个性化推荐具体包括:

对所述推荐词词源进行扩充,采用第二推荐策略获取与所述查询词或所述用户的已有关键词的相关性满足第二相关性要求的关键词作为扩充词添加到所述推荐词词源中;

将扩充后的推荐词词源作为推荐结果推荐给所述用户。

2. 根据权利要求1所述的方法,其特征在于,所述计算查询词与已有关键词的相关性具体包括:

对所述查询词进行分词处理,得到查询词集合 T_b ;

利用查询词集合 T_b 对已有关键词的命中情况,计算查询词与已有关键词之间的相关系数 x ,该相关系数 x 表征所述查询词与已有关键词之间的相关性。

3. 根据权利要求2所述的方法,其特征在于,所述计算查询词与已有关键词之间的相关系数 x 包括:

所述相关系数 x 等于所有查询词集合 T_b 中命中已有关键词的元素的总字节长度与查询词的字节长度的比值;

或者,所述相关系数 x 等于:查询词集合 T_b 中命中已有关键词的各个元素的字节长度乘以各个元素对应的权值之和,与查询词集合 T_b 中所有元素的字节长度乘以各个元素对应的权值之和的比值;

或者,所述相关系数 x 等于:查询词集合 T_b 中命中已有关键词的各个元素的字节长度乘以各个元素对应的已有关键词的权值之和,与查询词集合 T_b 的各个元素的字节长度乘以各个元素对应的已有关键词的权值之和的比值。

4. 根据权利要求1所述的方法,其特征在于,所述推广集合由所述用户的单个或多个推广单元构成。

5. 根据权利要求1所述的方法,其特征在于,所述采用第一推荐策略获取的关键词包括:

与所述查询词的字面相关性满足预设要求的关键词;

与所述查询词的语义相关性满足预设要求的关键词;或者

与所述查询词属于同一个行业类目或存在同一个购买者的关键词。

6. 根据权利要求1所述的方法,其特征在于,在所述个性化推荐之前,还包括对所述已有关键词进行过滤,去掉与输入的查询词相关性较低的已有关键词。

7. 根据权利要求1所述的方法,其特征在于,所述采用第二推荐策略获取的扩充词包括:与所述查询词或所述用户的已有关键词满足预设第二相关性要求的同行已有关键词,或者,与所述查询词或所述用户的已有关键词属于同一聚类且满足预设第二相关性要求的搜索关键词。

8. 根据权利要求 1 所述的方法,其特征在于,在对所述推荐词词源进行扩充之后,还包括:

计算扩充后的推荐词词源中的推荐词与查询词和已有关键词之间的相关性;
根据所述相关性计算结果对扩充后的推荐词词源中的各推荐词进行排序。

9. 根据权利要求 8 所述的方法,其特征在于,所述计算推荐词与查询词和已有关键词之间的相关性的方法是:

$$\mathcal{R}(w, q, u) = (1 - \lambda) \times \mathcal{R}(w, q) + \lambda \times \mathcal{R}(w, u)$$

$\mathcal{R}(w, q, u)$ 是推荐词 w 与查询词 q 和已有关键词 u 的相关性, $\mathcal{R}(w, q)$ 是推荐词 w 与查询词 q 的相关性, $\mathcal{R}(w, u)$ 是推荐词 w 与已有关键词 u 的相关性, λ 为调整因子,由查询词 q 与已有关键词 u 之间的相关性确定。

10. 一种关键词推荐系统,其特征在于,包括:

词源获取模块,用于接收用户输入的查询词,并采用第一推荐策略获取与所述查询词的相关性满足预设第一相关性要求的关键词,得到推荐词词源;

相关性计算模块,用于获取所述用户的推广集合中的已有关键词,并计算查询词与已有关键词的相关性;

判断模块,用于判断所述相关性计算模块计算得到的相关性是否达到预设的触发阈值,若是,则触发个性化推荐,触发词源扩充模块,若否,则触发推荐结果模块;

词源扩充模块,用于受到触发后,对所述推荐词词源进行扩充,采用第二推荐策略获取与查询词或已有关键词的相关性满足第二相关性要求的关键词作为扩充词添加到所述推荐词词源中,并触发推荐结果模块;

推荐结果模块,用于受到触发后,将推荐词词源作为推荐结果推荐并显示给所述用户。

11. 根据权利要求 10 所述的系统,其特征在于,所述相关性计算模块先对所述查询词进行分词处理,得到查询词集合 T_b ,再利用查询词集合 T_b 对已有关键词的命中情况,计算查询词与所有已有关键词之间的相关系数 x ,该相关系数 x 表征所述查询词与已有关键词之间的相关性。

12. 根据权利要求 11 所述的系统,其特征在于,所述相关性计算模块计算查询词与已有关键词之间的相关系数 x 包括:

所述相关系数 x 等于所有查询词集合 T_b 中命中已有关键词的元素的总字节长度与查询词的字节长度的比值;

或者,所述相关系数 x 等于:查询词集合 T_b 中命中已有关键词的各个元素的字节长度乘以各个元素对应的权值之和,与查询词集合 T_b 中所有元素的字节长度乘以各个元素对应的权值之和的比值;

或者,所述相关系数 x 等于:查询词集合 T_b 中命中已有关键词的各个元素的字节长度乘以各个元素对应的已有关键词的权值之和,与查询词集合 T_b 的各个元素的字节长度乘以各个元素对应的已有关键词的权值之和的比值。

13. 根据权利要求 10 所述的系统,其特征在于,所述推广集合由所述用户的单个或多个推广单元构成。

14. 根据权利要求 10 所述的系统,其特征在于,所述词源获取模块采用第一推荐策略

获取的关键词包括：

- 与所述查询词的字面相关性满足预设要求的关键词；
- 与所述查询词的语义相关性满足预设要求的关键词；或者
- 与所述查询词属于同一个行业类目或存在同一个购买者的关键词。

15. 根据权利要求 10 所述的系统,其特征在于,还包括:已有关键词过滤模块,用于对已有关键词进行过滤,去掉与输入的查询词相关性低的已有关键词,并将过滤后的已有关键词提供给所述词源扩充模块使用。

16. 根据权利要求 10 所述的系统,其特征在于,所述词源扩充模块采用第二推荐策略获取的扩充词包括:与所述查询词或所述用户的已有关键词满足预设第二相关性要求的同行已有关键词,或者,与所述查询词或所述用户的已有关键词属于同一聚类且满足预设第二相关性要求的搜索关键词。

17. 根据权利要求 10 所述的系统,其特征在于,该系统还包括:

第二相关性计算模块,用于计算扩充后的推荐词词源中的推荐词与查询词和已有关键词之间的相关性;

排序模块,用于根据所述第二相关性计算模块的计算结果对扩充后的推荐词词源中各推荐词进行排序。

18. 根据权利要求 17 所述的系统,其特征在于,所述第二相关性计算模块计算推荐词与查询词和已有关键词之间的相关性的公式是:

$$\mathcal{R}(w, q, u) = (1 - \lambda) \times \mathcal{R}(w, q) + \lambda \times \mathcal{R}(w, u)$$

$\mathcal{R}(w, q, u)$ 是推荐词 w 与查询词 q 和已有关键词 u 的相关性, $\mathcal{R}(w, q)$ 是推荐词 w 与查询词 q 的相关性, $\mathcal{R}(w, u)$ 是推荐词 w 与已有关键词 u 的相关性, λ 为调整因子,由查询词 q 与已有关键词 u 之间的相关性确定。

关键词推荐方法及系统

【技术领域】

[0001] 本发明涉及互联网应用技术,尤其涉及一种关键词推荐方法及系统。

【背景技术】

[0002] 网络推广技术是随搜索引擎的发展而发展起来的一种技术,例如百度推广、Google 关键字广告 (AdWords) 和广告联盟 (AdSense) 等。当在搜索引擎中输入关键词检索信息时,系统投放与输入的关键词相匹配的网络广告,以纯文本的方式设置于搜索结果页面的右侧空白处,或者,以相应关键词匹配的广告出现在搜索结果中,并处于较前的位置。此类广告是按照用户购买的关键词进行推广的,其关键词的选择对于推广计划是否成功有着密不可分的关系。为了帮助用户在新建或完善推广单元时选取更好的关键词,有效地实现网络营销,推广网站都会有辅助推荐工具进行关键词的推荐。特别是在用户的推广计划、推广单元层级不断增多的情况下,辅助推荐工具的作用尤为明显,可以帮助用户快速完成关键词的选取。

[0003] 现有的关键词推荐的主要形式是根据用户的输入内容,推荐出与输入内容在字面、结构、语义上相关的关键词,类似于搜索技术,其推荐结果质量的好坏往往与用户输入内容密切相关。对于目标不明确的输入内容,会使得推荐结果覆盖业务范围较广泛,可能只有一小部分结果符合该用户的实际需求,这样会增加用户的筛选难度,降低实际的推荐质量。

[0004] 在实际使用的过程中,由于用户输入内容的不确定性,并不是每次查询都需要个性化推荐,现有的推荐方法由于所依赖的信息不足,并不能判断是否需要,而盲目地进行个性化推荐导致需求判断错误,推荐结果丧失多样性,影响实际的推荐质量。

【发明内容】

[0005] 为了解决上述问题,本发明提供了一种关键词推荐方法及系统,以便于防止出现盲目的个性化推荐,同时又保证推荐结果的多样性。

[0006] 具体技术方案如下:

[0007] 本发明实施例提供的一种关键词推荐方法,包括:

[0008] 接收用户输入的查询词,采用第一推荐策略获取与所述查询词的相关性满足预设第一相关性要求的关键词作为推荐词词源;

[0009] 获取所述用户的推广集合中的已有关键词,并计算所述查询词与已有关键词的相关性;

[0010] 判断所述查询词与已有关键词的相关性是否达到预设的触发阈值,若是,则触发个性化推荐,若否,则将所述推荐词词源作为推荐结果推荐给所述用户;

[0011] 其中,所述触发个性化推荐具体包括:

[0012] 对所述推荐词词源进行扩充,采用第二推荐策略获取与所述查询词或所述用户的已有关键词的相关性满足第二相关性要求的关键词作为扩充词添加到所述推荐词词源

中；

[0013] 将扩充后的推荐词词源作为推荐结果推荐给所述用户。

[0014] 根据本发明之一优选实施例，所述计算查询词与已有关键词的相关性具体包括：

[0015] 对所述查询词进行分词处理，得到查询词集合 T_b ；

[0016] 利用查询词集合 T_b 对已有关键词的命中情况，计算查询词与已有关键词之间的相关系数 x ，该相关系数 x 表征所述查询词与已有关键词之间的相关性。

[0017] 根据本发明之一优选实施例，所述计算查询词与已有关键词之间的相关系数 x 包括：

[0018] 所述相关系数 x 等于所有查询词集合 T_b 中命中已有关键词的元素的总字节长度与查询词的字节长度的比值；

[0019] 或者，所述相关系数 x 根据查询词集合 T_b 中命中已有关键词的各元素的权值进行计算，所述相关系数 x 等于：查询词集合 T_b 中命中已有关键词的各个元素的字节长度乘以各个元素对应的权值之和，与查询词集合 T_b 中所有元素的字节长度乘以各个元素对应的权值之和的比值；

[0020] 或者，所述相关系数 x 根据查询词集合 T_b 中的元素命中的已有关键词的权值进行计算，所述相关系数 x 等于：查询词集合 T_b 中命中已有关键词的各个元素的字节长度乘以各个元素对应的已有关键词的权值之和，与查询词集合 T_b 的各个元素的字节长度乘以各个元素对应的已有关键词的权值之和的比值。

[0021] 根据本发明之一优选实施例，所述推广集合由所述用户的单个或多个推广单元构成。

[0022] 根据本发明之一优选实施例，所述采用第一推荐策略获取的关键词包括：

[0023] 与所述查询词的字面相关性满足预设要求的关键词；

[0024] 与所述查询词的语义相关性满足预设要求的关键词；或者

[0025] 与所述查询词属于同一个行业类目或存在同一个购买者的关键词。

[0026] 根据本发明之一优选实施例，在所述个性化推荐之前，还包括对所述已有关键词进行过滤，去掉与输入的查询词相关性较低的已有关键词。

[0027] 根据本发明之一优选实施例，所述采用第二推荐策略获取的扩充词包括：与所述查询词或所述用户的已有关键词满足预设第二相关性要求的同行已有关键词，或者，与所述查询词或所述用户的已有关键词属于同一聚类且满足预设第二相关性要求的搜索关键词。

[0028] 根据本发明之一优选实施例，在对所述推荐词词源进行扩充之后，还包括：

[0029] 计算扩充后的推荐词词源中的推荐词与查询词和已有关键词之间的相关性；

[0030] 根据所述相关性计算结果对扩充后的推荐词词源中的各推荐词进行排序。

[0031] 根据本发明之一优选实施例，所述计算推荐词与查询词和已有关键词之间的相关性的方法是：

[0032]

$$\mathcal{R}(w, q, u) = (1 - \lambda) \times \mathcal{R}(w, q) + \lambda \times \mathcal{R}(w, u)$$

[0033] $\mathcal{R}(w, q, u)$ 是推荐词 w 与查询词 q 和已有关键词 u 的相关性， $\mathcal{R}(w, q)$ 是推荐词 w 与查询词 q 的相关性， $\mathcal{R}(w, u)$ 是推荐词 w 与已有关键词 u 的相关性， λ 为调

整因子,由查询词 q 与已有关键词 u 之间的相关性确定。

[0034] 相应地,本发明实施例提供一种关键词推荐系统,包括

[0035] 词源获取模块,用于接收用户输入的查询词,并采用第一推荐策略获取与所述查询词的相关性满足预设第一相关性要求的关键词,得到推荐词词源;

[0036] 相关性计算模块,用于获取所述用户的推广集合中的已有关键词,并计算查询词与已有关键词的相关性;

[0037] 判断模块,用于判断所述相关性计算模块计算得到的相关性是否达到预设的触发阈值,若是,则触发个性化推荐,触发词源扩充模块,若否,则触发推荐结果模块;

[0038] 词源扩充模块,用于受到触发后,对所述推荐词词源进行扩充,采用第二推荐策略获取与查询词或已有关键词的相关性满足第二相关性要求的关键词作为扩充词添加到所述推荐词词源中,并触发推荐结果模块;

[0039] 推荐结果模块,用于受到触发后,将推荐词词源作为推荐结果推荐并显示给所述用户。

[0040] 根据本发明之一优选实施例,所述相关性计算模块先对所述查询词进行分词处理,得到查询词集合 T_b ,再利用查询词集合 T_b 对已有关键词的命中情况,计算查询词与所有已有关键词之间的相关系数 x ,该相关系数 x 表征所述查询词与已有关键词之间的相关性。

[0041] 根据本发明之一优选实施例,所述相关性计算模块计算查询词与已有关键词之间的相关系数 x 包括:

[0042] 所述相关系数 x 等于所有查询词集合 T_b 中命中已有关键词的元素的总字节长度与查询词的字节长度的比值;

[0043] 或者,所述相关系数 x 等于:查询词集合 T_b 中命中已有关键词的各个元素的字节长度乘以各个元素对应的权值之和,与查询词集合 T_b 中所有元素的字节长度乘以各个元素对应的权值之和的比值;

[0044] 或者,所述相关系数 x 等于:查询词集合 T_b 中命中已有关键词的各个元素的字节长度乘以各个元素对应的已有关键词的权值之和,与查询词集合 T_b 的各个元素的字节长度乘以各个元素对应的已有关键词的权值之和的比值。

[0045] 根据本发明之一优选实施例,所述推广集合由所述用户的单个或多个推广单元构成。

[0046] 根据本发明之一优选实施例,所述词源获取模块采用第一推荐策略获取的关键词包括:

[0047] 与所述查询词的字面相关性满足预设要求的关键词;

[0048] 与所述查询词的语义相关性满足预设要求的关键词;或者

[0049] 与所述查询词属于同一个行业类目或存在同一个购买者的关键词。

[0050] 根据本发明之一优选实施例,还包括:已有关键词过滤模块,用于对已有关键词进行过滤,去掉与输入的查询词相关性低的已有关键词,并将过滤后的已有关键词提供给所述词源扩充模块使用。

[0051] 根据本发明之一优选实施例,所述词源扩充模块采用第二推荐策略获取的扩充词包括:与所述查询词或所述用户的已有关键词满足预设第二相关性要求的同行已有关键词,或者,与所述查询词或所述用户的已有关键词属于同一聚类且满足预设第二相关性要

求的搜索关键词。

[0052] 根据本发明之一优选实施例,该系统还包括:

[0053] 第二相关性计算模块,用于计算扩充后的推荐词词源中的推荐词与查询词和已有关键词之间的相关性;

[0054] 排序模块,用于根据所述第二相关性计算模块的计算结果对扩充后的推荐词词源中各推荐词进行排序。

[0055] 根据本发明之一优选实施例,所述第二相关性计算模块计算推荐词与查询词和已有关键词之间的相关性的公式是:

[0056]

$$\mathcal{R}(w, q, u) = (1 - \lambda) \times \mathcal{R}(w, q) + \lambda \times \mathcal{R}(w, u)$$

[0057] $\mathcal{R}(w, q, u)$ 是推荐词 w 与查询词 q 和已有关键词 u 的相关性, $\mathcal{R}(w, q)$ 是推荐词 w 与查询词 q 的相关性, $\mathcal{R}(w, u)$ 是推荐词 w 与已有关键词 u 的相关性, λ 为调整因子,由查询词 q 与已有关键词 u 之间的相关性确定。

[0058] 由以上技术方案可以看出,本发明提供了一种关键词推荐方法及系统,在推荐时引入更多用户信息,利用已有信息提取出用户实际查询需求的问题,令推荐结果可更好的满足不同用户的实际需求,并通过触发过程的限制,只对可识别的查询进行个性化推荐,防止出现盲目的个性化,控制个性化信息的权值,同时保证了推荐结果的多样性,可实现针对不同使用用户的个性化推荐。

【附图说明】

[0059] 图1是本发明实施例提供的关键词推荐方法的流程图;

[0060] 图2是本发明实施例提供的计算查询词与已有关键词的相关性的流程图;

[0061] 图3是本发明实施例提供的关键词推荐系统的框图。

【具体实施方式】

[0062] 为了使本发明的目的、技术方案和优点更加清楚,下面结合附图和具体实施例对本发明进行详细描述。

[0063] 在网络推广营销中,推广单元是管理关键词和创意的小单位,推广单元的建立与关键词的选择息息相关。关键词可以是产品词、通俗词、地域词、品牌词等,在建立推广单元时,一般将意义相近、结构相同的关键词纳入同一个推广单元中,以便有针对性地撰写创意。推广单元信息包括产品、关键词、创意或者网页。同一个用户可以为其不同的产品建立不同的推广单元,在各个推广单元中选取关键词,添加创意及产品信息。

[0064] 本发明是针对用户在新建或维护推广单元时选取的关键词进行个性化推荐,可以在用户输入内容不明确的情况下,根据推广单元中用户已有信息进行个性化推荐。

[0065] 如图1所示,为本发明实施例提供的关键词推荐方法的流程图,包括:

[0066] S1:接收用户输入的查询词,采用第一推荐策略获取与所述查询词的相关性满足预设第一相关性要求的关键词作为推荐词词源。

[0067] 用户输入的查询词的内容可以是与某一产品或服务相关的词、句或短语,比如“机

票”，“在线游戏”等查询词。

[0068] 所述采用第一推荐策略获取的关键词包括：

[0069] 与所述查询词的字面相关性满足预设要求的关键词。其中，所述字面相关性的计算方法可以采用文本信息相关性的计算方法得到。

[0070] 比如，计算关键词的文本信息相关性的方法是：假设一个文章集合 $\{W\}$ ，总文章数目为 N ，其中含有关键词 A 的文章总数为 N_a ，含有关键词 B 的文章总数是 N_b ，含有 $\{A+B\}$ 的文章总数是 N_{ab} ，那么相关性 $R_{AB} = N_{ab} / (N_a + N_b - N_{ab})$ 。

[0071] 选择满足预设要求的关键词时，例如可以选择相关性大于一定阈值的关键词或者排在前 $N1$ 个的关键词， $N1$ 为预设正整数。比如，与查询词“在线字典”在字面上相关的有“英语在线字典”“新华在线字典”“日语在线字典”“中华在线字典”等词。选取前 $N1 = 2$ 个关键词，即“英语在线字典”“新华在线字典”作为推荐词词源。

[0072] 相类似地，还有与所述查询词的语义相关性满足预设相关性要求的关键词。所述语义相关性是指在潜在含义上有联系的关键词。比如，对于查询词“清华大学”，可以有“清华大学”“清华”“THU”“Tsinghua”等关键词。

[0073] 或者，还可以推荐与所述查询词属于同一个行业类目的关键词。比如，“二极管”属于电子元器件类目，相关的词有“三极管”“光耦”等关键词。

[0074] 或者，还可以推荐与所述查询词存在同一个购买者的关键词。通常如果一个购买者购买的关键词之间是有相互关联，成套或组合使用的。比如，某个型号的手机与其电池。或者，是具有相同或相似功能和效果的词，通常具有可替代关系的，比如用于盛放液体的玻璃杯和陶瓷杯。

[0075] $S2$ ：获取所述用户的推广集合中的已有关键词，并计算所述查询词与已有关键词的相关性。

[0076] 所述推广集合由所述用户的单个或多个推广单元构成，也可以是该用户的所有推广单元。

[0077] 所述计算查询词与已有关键词之间的相关性，如图 2 所示，具体包括：

[0078] $S201$ 、对所述查询词进行分词处理，得到查询词集合 T_b 。

[0079] $S202$ 、利用查询词集合 T_b 对已有关键词的命中情况，确定查询词与已有关键词之间的相关系数 x ，该相关系数 x 表征所述查询词与已有关键词之间的相关性。

[0080] 所述相关系数 x 可以采用但不限于以下三种方式：

[0081] 第一种方式：采用较为简单的方式进行计算，所述相关系数 x 等于所有查询词集合 T_b 中命中已有关键词的元素的总字节长度与查询词的字节长度的比值。

[0082] 举个例子，对于查询词“多功能电脑桌”，先分词得到查询词集合 $T_b \{ \text{多功能, 电脑, 桌} \}$ ，已有关键词假设有“办公电脑桌”、“台式机电脑桌”，则查询词集合 T_b 中命中已有关键词的元素为“电脑”和“桌”，其总字节长度为 6，查询词的字节长度为 12，则相关系数为 0.5。如果已有关键词有“网吧电脑桌”、“多功能家具”，则查询词集合 T_b 中命中已有关键词的元素为“多功能”、“电脑”和“桌”，其总字节长度为 12，查询词的字节长度为 12，则相关系数为 1。

[0083] 第二种方式：所述相关系数 x 还可以根据查询词集合 T_b 中命中已有关键词的各元素的权值进行计算。首先为查询词集合 T_b 中的元素赋予不同的权值：确定查询词集合 T_b 中

的核心元素,为核心元素赋予较高的权值,其中核心元素可以但不限于根据表意能力进行确定,则相关系数 x 等于查询词集合 T_b 中命中已有关键词的各个元素的字节长度乘以各个元素对应的权值之和,与查询词集合 T_b 中所有元素的字节长度乘以各个元素对应的权值之和的比值。

[0084] 上述例子中,查询词“多功能电脑桌”,假设已有关键词有“办公电脑桌”、“台式机电脑桌”,确定查询词集合 T_b { 多功能, 电脑, 桌 } 中“电脑”和“桌”为核心元素,权值均为 2,“多功能”的权值为 1,其中命中已有关键词的元素为“电脑”和“桌”,则相关系数 $x = (2 \times 4 + 2 \times 2) \div (2 \times 4 + 2 \times 2 + 1 \times 6) = 2/3$ 。

[0085] 第三种方式:所述相关系数 x 还可以根据查询词集合 T_b 中的元素命中的已有关键词的权值进行计算,已有关键词的权值可以根据消费率或者点击率等因素确定,体现已有关键词的重要程度。所述相关系数 x 等于:查询词集合 T_b 中的命中已有关键词的各个元素的字节长度乘以各个元素对应的已有关键词的权值之和,与查询词集合 T_b 的各个元素的字节长度乘以各个元素对应的已有关键词的权值之和的比值。其中各个元素对应的已有关键词的权值取对应已有关键词的权值中较高的值。

[0086] 例如,查询词“多功能电脑桌”,已有关键词有“电脑桌设计”、“定制电脑桌”“网吧电脑桌”,已有关键词分别的权值是 1、1 和 2,则查询词集合 T_b 中命中已有关键词中的元素有“电脑”和“桌”,取对应已有关键词的权值较高的值作为元素对应的权值。因而得到相关系数 $x = (2 \times 4 + 2 \times 2) \div (2 \times 4 + 2 \times 2 + 1 \times 6) = 2/3$ 。

[0087] S3:判断所述查询词与已有关键词的相关性是否达到预设的触发阈值 x_0 ,若是,则触发个性化推荐,执行步骤 S4,若否,则直接对所述推荐词词源根据推荐词的相关性进行排序,得到关键词的推荐结果,结束流程。

[0088] 所述触发阈值 x_0 为预设的数值,根据实际的应用场景,可以设定不同的触发阈值,当查询词与已有关键词的相关性在 $[x_0, 1]$ 区间时才触发个性化推荐。例如 x_0 取值 0.2,则只有在查询词与已有关键词的相关性大于或等于 0.2 时,才触发个性化推荐。触发阈值设定得越低,越容易触发个性化推荐。

[0089] 对所述推荐词词源根据推荐词的相关性进行排序具体包括:

[0090] 计算推荐词词源中的推荐词与查询词之间的相关性 $\mathcal{R}(w, q)$,并据此相关性分配权值,进行排序后得到推荐结果显示给用户。

[0091] 其中,所述推荐词与查询词之间的相关性 $\mathcal{R}(w, q)$ 的计算方式可以采用空间向量模型的计算方式,如现有的内积相似度、余弦相似度等方法来计算。

[0092] 优选地,在个性化推荐之前,还包括对所述已有关键词进行过滤,去掉与输入的查询词相关性较低的已有关键词,提高效率,也便于控制后续添加扩充词的范围。

[0093] S4、对所述推荐词词源进行扩充,采用第二推荐策略获取与所述查询词或所述用户的已有关键词的相关性满足预设第二相关性要求的关键词作为扩充词添加到所述推荐词词源中。

[0094] 所述采用第二推荐策略获取的扩充词包括:与所述查询词或所述用户的已有关键词满足预设第二相关性要求的同行已有关键词,或者,与所述查询词属于同一聚类且满足预设第二相关性要求的搜索关键词。

[0095] 所述同行已有关键词是指同行业的客户已经购买的关键词,这些关键词之间通常具有一定的语义相关性。

[0096] 举个例子,用户输入的查询词“英语培训”,对于同一行业的其他用户,如韦博、新东方等客户,选择购买的关键词可能是“雅思口语”“托福英语听力培训”“职场英语培训”“专业英语”等,分别计算与查询词“英语培训”的之间的相关性,可以选择满足预设相关性要求的同行已有关键词作为扩充词,例如可以选择相关性大于一定阈值的关键词或者排在前 N_2 个的关键词, N_2 为预设正整数。比如,阈值设为 0.3,假设相关性大于 0.3 的关键词有“托福英语听力培训”“职场英语培训”“专业英语”,则将该些关键词作为扩充词,添加到推荐词词源中。

[0097] 所述同一聚类是指根据网民的搜索习惯进行聚类的策略,是通过统计网民的搜索日志,将能够得到同一类搜索结果的关键词聚类成同一聚类。比如,网民要进行购买鲜花,根据网民有搜索过的关键词可能有“鲜花预定”“鲜花速递”“鲜花工坊”“鲜花店”等,将这些关键词聚类于同一聚类。再比如,“淘宝”“拍拍”“当当”“有啊”等聚类于“购物”聚类,“优酷”“土豆”“迅雷”“酷 6”等聚类归于“视频”聚类。

[0098] 根据同一聚类中关键词与所述查询词的相关性,可以选择满足预设第二相关性要求的关键词作为扩充词,例如可以选择相关性大于一定阈值的关键词或者排在前 N_2 个的关键词, N_2 为预设正整数。比如,对于查询词鲜花,可以选取排在前 $N_2 = 2$ 个关键词,就有“鲜花预定”“鲜花速递”作为扩充词。

[0099] S5、计算扩充后的推荐词词源中的推荐词与查询词和已有关键词之间的相关性。

[0100] 每个推荐词与查询词和已有关键词之间的相关性的计算公式为:

[0101]

$$\mathcal{R}(w, q, u) = (1 - \lambda) \times \mathcal{R}(w, q) + \lambda \times \mathcal{R}(w, u)$$

[0102] $\mathcal{R}(w, q, u)$ 为推荐词 w 与查询词 q 和已有关键词 u 的相关性, $\mathcal{R}(w, q)$ 是推荐词 w 与查询词 q 的相关性, $\mathcal{R}(w, u)$ 是推荐词 w 与已有关键词 u 的相关性, λ 为调整因子,由查询词与已有关键词之间的相关性确定,查询词与已有关键词之间的相关性越大,调整因子 λ 的值越高。其中, $\mathcal{R}(w, u)$ 和 $\mathcal{R}(w, q)$ 的计算方式均可以采用空间向量模型的计算方式,如现有的内积相似度、余弦相似度等方法来计算。

[0103] S6、根据步骤 S5 的相关性计算结果,对扩充后的推荐词词源中的各推荐词进行排序。

[0104] 在本步骤中根据相关性计算结果进行的排序,能够将结果中与已有关键词具有较高相关性的推荐词进行排序提前,优先进行展示。其中,推荐词与查询词之间的相关性能够体现出推荐词词源与查询词之间的相关性。

[0105] S7、依据步骤 S6 的排序结果,将扩充后的推荐词词源作为推荐结果显示给用户。

[0106] 在实际使用的过程中,在得到关键词的推荐结果时,还可以对所述推荐结果中的推荐词进行一些过滤,去掉与产品整体策略不符的推荐词,去掉与推广单元中已有关键词重复的推荐词,保证已有的关键词不会在推荐结果中出现。

[0107] 实施例 1、

[0108] 以百度推广中的关键词推荐为例,当用户选择在具体的推广单元里添加关键词

时,使用关键词推荐工具来获取推荐词,则会利用该推广单元的已有关键词作为辅助推荐信息来进行个性化推荐。

[0109] 由于同一推广单元的已有关键词一般为同一业务且结构类似的关键词,当用户在具体推广单元内进行选词时往往需求较为明确,因此利用具体推广单元中的已有关键词作为查询的需求补充信息,对推荐词词源进行扩充。

[0110] 如某一推广单元中的已有关键词为: A_1, A_2, \dots, A_n , 接收到用户输入的查询词为 B , 则计算查询词 B 与该推广单元中已有关键词 A_1, A_2, \dots, A_n 之间的相关性,具体是:

[0111] 首先对查询词 B 进行分词处理,得到查询词集合 T_B 。

[0112] 利用查询词集合 T_B 对已有关键词的命中情况,确定查询词 B 和已有关键词之间的相关系数 x ,所述相关系数 x 等于所有查询词集合 T_B 中的元素出现在已有关键词中的总字节长度除以查询词的字节长度,该相关系数 x 表征所述查询词与已有关键词之间的相关性。

[0113] 判断所述相关系数 x 大于预设的触发阈值 x_0 ,则触发个性化推荐。

[0114] 如果相关系数小于 x_0 ,不触发个性化推荐,采用第一推荐策略获取与所述查询词 B 的相关性满足预设第一相关性要求的关键词作为推荐词词源,并根据推荐词词源中各推荐词与查询词 B 的相关性进行排序,将结果展示给用户。

[0115] 如果相关系数等于或大于 x_0 ,则触发个性化推荐,包括:

[0116] 对推荐词词源扩充,除了采用第一推荐策略获取的推荐词词源以外,还包括与查询词 B 或已有关键词 A_1, A_2, \dots, A_n 满足预设第二相关性要求的同行已有关键词 C_1, C_2, \dots , 以及与所述查询词 B 属于同一聚类且满足预设第二相关性要求的搜索关键词 D_1, D_2, \dots 。

[0117] 计算推荐词词源中每一个推荐词与查询词和已有关键词之间的相关性,采用的计算公式为:

[0118]

$$\mathcal{R}(w, q, u) = (1 - \lambda) \times \mathcal{R}(w, q) + \lambda \times \mathcal{R}(w, u)$$

[0119] 其中, $\mathcal{R}(w, q, u)$ 为推荐词 w 与查询词 q 和已有关键词 u 的相关性, $\mathcal{R}(w, q)$ 是推荐词 w 与查询词 q 的相关性, $\mathcal{R}(w, u)$ 是推荐词 w 与已有关键词 u 的相关性。 λ 为调整因子,由查询词 q 与已有关键词 u 之间的相关性确定。

[0120] 根据查询词 B 与已有关键词 A_1, A_2, \dots, A_n 之间的相关系数 x ,确定调整因子 λ ,根据确定调整因子 λ ,所述查询词与已有关键词之间的相关性越大,调整因子 λ 的值越高,则推荐词中引入已有关键词信息权值越高,对推荐结果的调整则越显著。

[0121] 其函数关系 $\lambda = f(x)$ 可根据实际的应用需求采用不同的函数关系。例如,可以采用函数关系式为 $\lambda = f(x) = \frac{0.5}{\exp(1-x)}$,确定已有关键词对整体推荐结果排序的影响。

[0122] 在完成所述每个推荐词与查询词和已有关键词之间的相关性计算之后,对所述推荐词词源中的推荐词根据相关性计算结果进行排序,能够将结果中与已有关键词相关的推荐词进行排序提前,优先进行展示,同时产生更多类似的推荐结果,以更好的满足用户对此类关键词的需求。

[0123] 最后,根据推荐词在推荐词词源中的排序结果,将扩充后的推荐词词源作为推荐结果 Q_1, Q_2, \dots, Q_n ,显示给用户。

[0124] 上述实施例是针对单个推广单元进行个性化推荐的说明,在实际方案中,由于用户已有关键词的多样性,往往涉及各种不同的应用场景,本发明还可以将已有关键词的获取单元扩大至该用户的多个推广单元,甚至是该用户账户内的所有已有关键词。

[0125] 对于多个推广单元的已有关键词,由于已有关键词涉及的行业和种类可能较多,有些可能与输入的查询词相差很远。为了规范引入的已有关键词和节省存储空间及处理时间,优选地在判断为触发个性化推荐之后,还包括一过滤已有关键词的步骤,判断每个已有关键词与查询词之间的相关性,并将相关性小于某一预设值的已有关键词作为噪音排除掉,去掉与查询词相关性低的已有关键词。后续的过程与上述单个推广单元的一致。

[0126] 相应地,如图 3 所示,为本发明实施例提供的关键词推荐系统的框图,包括:

[0127] 词源获取模块 101,用于接收用户输入的查询词,并采用第一推荐策略获取与所述查询词的相关性满足预设第一相关性要求的关键词,得到推荐词词源。

[0128] 用户输入的查询词的内容可以是与某一产品或服务相关的词或短语,比如输入“机票”,“在线游戏”等。

[0129] 所述第一推荐策略获取的关键词包括:

[0130] 与所述查询词的字面相关性满足预设要求的关键词。其中,所述字面相关性的计算方法可以采用文本信息相关性的计算方法得到。

[0131] 比如,计算关键词的文本信息相关性的方法是:假设一个文章集合 $\{W\}$,总文章数目为 N ,其中含有关键词 A 的文章总数为 N_a ,含有关键词 B 的文章总数是 N_b ,含有 $\{A+B\}$ 的文章总数是 N_{ab} ,那么相关性 $R_{AB} = N_{ab} / (N_a + N_b - N_{ab})$ 。

[0132] 选择满足预设要求的关键词时,例如可以选择相关性大于一定阈值的关键词或者排在前 $N1$ 个的关键词, $N1$ 为预设正整数。比如,与查询词“在线字典”在字面上相关的有“英语在线字典”“新华在线字典”“日语在线字典”“中华在线字典”等词。选取前 $N1 = 2$ 个关键词,即“英语在线字典”“新华在线字典”作为推荐词词源。

[0133] 相类似地,还有与所述查询词的语义相关性满足预设相关性要求的关键词。所述语义相关性是指在潜在含义上有联系的关键词。比如,对于查询词“清华大学”,可以有“清华大学”“清华”“THU”“Tsinghua”等关键词。

[0134] 或者,还可以推荐与所述查询词属于同一个行业类目的关键词。比如,“二极管”属于电子元器件类目,相关的词有“三极管”“光耦”等关键词。

[0135] 或者,还可以推荐与所述查询词存在同一个购买者的关键词。通常存在同一个购买者的关键词之间是有相互关联,成套或组合使用的。比如,某个型号的手机与其电池。或者,也可以是具有相同或相似功能和效果的词,通常具有可替代关系的,比如用于盛放液体的玻璃杯和陶瓷杯。

[0136] 相关性计算模块 102,用于获取所述用户的推广集合中的已有关键词,并计算查询词与已有关键词的相关性。

[0137] 所述推广集合由所述用户的单个或多个推广单元构成,也可以是该用户的所有推广单元。

[0138] 具体地,相关性计算模块 102 先对所述查询词进行分词处理,得到查询词集合 T_B ;再利用查询词集合 T_B 对已有关键词的命中情况,计算查询词与所有已有关键词之间的相关系数 x ,该相关系数 x 表征所述查询词与已有关键词之间的相关性。所述相关系数 x 可以采

用但不限于以下三种方式：

[0139] 第一种方式：采用较为简单的方式进行计算，所述相关系数 x 等于所有查询词集合 T_b 中命中已有关键词的元素的总字节长度与查询词的字节长度的比值。

[0140] 举个例子，对于查询词“多功能电脑桌”，先分词得到查询词集合 T_b { 多功能，电脑，桌 }，已有关键词假设有“办公电脑桌”、“台式机电脑桌”，则查询词集合 T_b 中命中已有关键词的元素为“电脑”和“桌”，其总字节长度为 6，查询词的字节长度为 12，则相关系数为 0.5。如果已有关键词有“网吧电脑桌”、“多功能家具”，则查询词集合 T_b 中命中已有关键词的元素为“多功能”、“电脑”和“桌”，其总字节长度为 12，查询词的字节长度为 12，则相关系数为 1。

[0141] 第二种方式：所述相关系数 x 还可以根据查询词集合 T_b 中命中已有关键词的各元素的权值进行计算。首先为查询词集合 T_b 中的元素赋予不同的权值：确定查询词集合 T_b 中的核心元素，为核心元素赋予较高的权值，其中核心元素可以但不限于根据表意能力进行确定，则相关系数 x 等于查询词集合 T_b 中命中已有关键词的各个元素的字节长度乘以各个元素对应的权值之和，与查询词集合 T_b 中所有元素的字节长度乘以各个元素对应的权值之和的比值。

[0142] 上述例子中，查询词“多功能电脑桌”，假设已有关键词有“办公电脑桌”、“台式机电脑桌”，确定查询词集合 T_b { 多功能，电脑，桌 } 中“电脑”和“桌”为核心元素，权值均为 2，“多功能”的权值为 1，其中命中已有关键词的元素为“电脑”和“桌”，则相关系数 $x = (2 \times 4 + 2 \times 2) \div (2 \times 4 + 2 \times 2 + 1 \times 6) = 2/3$ 。

[0143] 第三种方式：所述相关系数 x 还可以根据查询词集合 T_b 中的元素命中的已有关键词的权值进行计算，已有关键词的权值可以根据消费率或者点击率等因素确定，体现已有关键词的重要程度。所述相关系数 x 等于：查询词集合 T_b 中的命中已有关键词的各个元素的字节长度乘以各个元素对应的已有关键词的权值之和，与查询词集合 T_b 的各个元素的字节长度乘以各个元素对应的已有关键词的权值之和的比值。其中各个元素对应的已有关键词的权值取对应已有关键词的权值中较高的值。

[0144] 例如，查询词“多功能电脑桌”，已有关键词有“电脑桌设计”、“定制电脑桌”“网吧电脑桌”，已有关键词分别的权值是 1、1 和 2，则查询词集合 T_b 中命中已有关键词中的元素有“电脑”和“桌”，取对应已有关键词的权值较高的值作为元素对应的权值。因而得到相关系数 $x = (2 \times 4 + 2 \times 2) \div (2 \times 4 + 2 \times 2 + 1 \times 6) = 2/3$ 。

[0145] 判断模块 103，用于判断所述相关性计算模块计算得到的相关性是否达到预设的触发阈值 x_0 ，若是，则触发个性化推荐，触发词源扩充模块 104，若否，则直接触发排序模块 106，对推荐词词源根据推荐词的相关性进行排序，再触发推荐结果模块 107，根据所述排序模块 106 的排序结果，将推荐词词源作为推荐结果推荐并显示给所述用户。

[0146] 所述触发阈值 x_0 为预设的数值，根据实际的应用场景，可以设定不同的触发阈值，当查询词与已有关键词的相关性在 $[x_0, 1]$ 区间时才触发个性化推荐。例如 x_0 取值 0.2，则只有在查询词与已有关键词的相关性大于或等于 0.2 时，才触发个性化推荐。触发阈值设定得越低，越容易触发个性化推荐。

[0147] 优选地，系统还包括一已有关键词过滤模块（图未示），用于对已有关键词进行过滤，去掉与输入的查询词相关性低的已有关键词，并将过滤后的已有关键词提供给所述词

源扩充模块使用。

[0148] 词源扩充模块 104, 用于受到触发后, 对所述推荐词词源进行扩充, 采用第二推荐策略获取与所述查询词或所述用户的已有关键词的相关性满足预设第二相关性要求的关键词作为扩充词添加到所述推荐词词源中。

[0149] 在词源扩充模块 104 中, 所述第二推荐策略获取的扩充词包括: 与所述查询词或所述用户的已有关键词满足预设第二相关性要求的同行已有关键词, 或者, 与所述查询词属于同一聚类且满足预设第二相关性要求的搜索关键词。

[0150] 所述同行已有关键词是指同行业的客户已经购买的关键词, 这些关键词之间通常具有一定的语义相关性。

[0151] 举个例子, 用户输入的查询词“英语培训”, 对于同一行业的其他用户, 如韦博、新东方等客户, 选择购买的关键词可能是“雅思口语”“托福英语听力培训”“职场英语培训”“专业英语”等, 分别计算与查询词“英语培训”的之间的相关性, 可以选择满足预设第二相关性要求的同行已有关键词作为扩充词, 例如可以选择相关性大于一定阈值的关键词或者排在前 N_2 个的关键词, N_2 为预设正整数。比如, 阈值设为 0.3, 假设相关性大于 0.3 的关键词有“托福英语听力培训”“职场英语培训”“专业英语”, 则将该些关键词作为扩充词, 添加到推荐词词源中。

[0152] 所述同一聚类是指根据网民的搜索习惯进行聚类的策略, 是通过统计网民的搜索日志, 将能够得到同一类搜索结果的关键词聚类成同一聚类。比如, 网民要进行购买鲜花, 根据网民有搜索过的关键词可能有“鲜花预定”“鲜花速递”“鲜花工坊”“鲜花店”等, 将这些关键词聚类于同一聚类。再比如, “淘宝”“拍拍”“当当”“有啊”等聚类于“购物”聚类, “优酷”“土豆”“迅雷”“酷 6”等聚类归于“视频”聚类。

[0153] 根据同一聚类中关键词与所述查询词的相关性, 可以选择满足预设第二相关性要求的关键词作为扩充词, 例如可以选择相关性大于一定阈值的关键词或者排在前 N_2 个的关键词, N_2 为预设正整数。比如, 对于查询词鲜花, 可以选取排在前 $N_2 = 2$ 个关键词, 就有“鲜花预定”“鲜花速递”作为扩充词。

[0154] 第二相关性计算模块 105, 用于计算扩充后的推荐词词源中的推荐词与查询词和已有关键词之间的相关性。

[0155] 具体地, 第二相关性计算模块 105 计算推荐词与查询词和已有关键词之间的相关性的公式是:

[0156]

$$\mathcal{R}(w, q, u) = (1 - \lambda) \times \mathcal{R}(w, q) + \lambda \times \mathcal{R}(w, u)$$

[0157] $\mathcal{R}(w, q, u)$ 是推荐词 w 与查询词 q 和已有关键词 u 的相关性, $\mathcal{R}(w, q)$ 是推荐词 w 与查询词 q 的相关性, $\mathcal{R}(w, u)$ 是推荐词 w 与已有关键词 u 的相关性, λ 为调整因子, 由查询词 q 与已有关键词 u 之间的相关性确定。所述查询词与已有关键词之间的相关性越大, 调整因子 λ 的值越高。

[0158] 排序模块 106, 用于根据所述第二相关性计算模块 105 的计算结果, 对扩充后的推荐词词源中各推荐词进行排序。

[0159] 排序模块 106 根据相关性计算结果进行排序, 能够将结果中与已有关键词具有较高相关性的推荐词进行排序提前, 优先进行展示。其中, 推荐词与查询词之间的相关性能够

体现出推荐词词源与查询词之间的相关性。

[0160] 推荐结果模块 107, 根据所述排序模块 106 的排序结果, 将推荐词词源作为推荐结果推荐并显示给所述用户。

[0161] 在实际使用的过程中, 在推荐结果模块 107 之前还可设置推荐词过滤模块, 对所述推荐结果中的推荐词进行一些过滤, 得到关键词的推荐结果, 去掉与产品整体策略不符的推荐词, 去掉与推广单元中已有关键词重复的推荐词, 保证已有的关键词不会在推荐结果中出现。

[0162] 继续以百度推广中的关键词推荐为例, 当用户选择在具体的推广单元里添加关键词时, 使用关键词推荐工具来获取推荐词, 则会利用该推广单元的已有关键词作为辅助推荐信息来进行个性化推荐。

[0163] 由于同一推广单元的已有关键词一般为同一业务且结构类似的关键词, 当用户在具体推广单元内进行选词时往往需求较为明确, 因此利用具体推广单元中的已有关键词作为查询的需求补充信息, 对推荐词词源进行扩充。

[0164] 如某一推广单元中的已有关键词为: A_1, A_2, \dots, A_n , 词源获取模块 101 接收到用户输入的查询词为 B 后, 相关性计算模块 102 则计算查询词 B 与该推广单元中已有关键词 A_1, A_2, \dots, A_n 之间的相关性, 具体是: 相关性计算模块 102 先对查询词 B 进行分词处理, 得到查询词集合 T_B ; 再利用查询词集合 T_B 对已有关键词的命中情况, 确定查询词 B 和已有关键词之间的相关系数 x , 所述相关系数 x 等于所有查询词集合 T_B 中的元素出现在已有关键词中的总字节长度除以查询词的字节长度, 该相关系数 x 表征所述查询词与已有关键词之间的相关性。

[0165] 判断模块 103 判断所述相关系数是否大于预设的触发阈值 x_0 , 如果是, 则触发个性化推荐。

[0166] 如果相关系数小于 x_0 , 不触发个性化推荐, 采用第一推荐策略获取与所述查询词 B 的相关性满足预设相关性要求的关键词作为推荐词词源, 并由排序模块 106 直接进行排序, 推荐结果模块 107 将结果展示给用户。

[0167] 如果相关系数大于 x_0 , 则触发个性化推荐, 先经过词源扩充模块 104 进行扩充推荐词的词源, 除了采用第一推荐策略获取的推荐词词源以外, 还包括与查询词 B 或已有关键词 A_1, A_2, \dots, A_n 满足预设第二相关性要求的同行已有关键词 C_1, C_2, \dots , 以及与所述查询词 B 属于同一聚类且满足预设第二相关性要求的搜索关键词 D_1, D_2, \dots 。

[0168] 第二相关性计算模块 105 计算推荐词词源中每一个推荐词与查询词和已有关键词之间的相关性, 采用的计算公式为:

[0169]

$$\mathcal{R}(w, q, u) = (1 - \lambda) \times \mathcal{R}(w, q) + \lambda \times \mathcal{R}(w, u)$$

[0170] 其中, $\mathcal{R}(w, q, u)$ 为推荐词 w 与查询词 q 和已有关键词 u 的相关性, $\mathcal{R}(w, q)$ 是推荐词 w 与查询词 q 的相关性, $\mathcal{R}(w, u)$ 是推荐词 w 与已有关键词 u 的相关性。 λ 为调整因子, 由查询词 q 与已有关键词 u 之间的相关性确定。

[0171] 根据查询词 B 与已有关键词 A_1, A_2, \dots, A_n 之间的相关系数 x , 确定调整因子 λ , 根据确定调整因子 λ , 所述查询词与已有关键词之间的相关性越大, 调整因子 λ 的值越高,

则推荐词中引入已有关键词信息权值越高,对推荐结果的调整则越显著。

[0172] 其函数关系 $\lambda = f(x)$ 可根据实际的应用需求采用不同的函数关系。例如,可以采用函数关系式为 $\lambda = f(x) = \frac{0.5}{\exp(1-x)}$,确定已有关键词对整体推荐结果排序的影响。

[0173] 排序模块 106 对词源扩充模块 104 得到的词源中的推荐词进行排序,根据第二相关性计算模块 105 计算得到的每个推荐词与查询词和已有关键词之间的相关性计算结果进行排序,能够将结果中与已有关键词相关的推荐词进行排序提前,优先进行展示,同时产生更多类似的推荐结果,以更好的满足用户对此类关键词的需求。

[0174] 最后,推荐结果模块 107 根据推荐词在推荐词词源中的排序结果,将扩充后的推荐词词源推荐结果 Q_1, Q_2, \dots, Q_n , 显示给用户。

[0175] 上述实施例是针对单个推广单元进行个性化推荐的说明,在实际方案中,由于用户已有关键词的多样性,往往涉及各种不同的应用场景,本发明还可以将已有关键词的获取单元扩大至该用户的多个推广单元,甚至是该用户账户内的所有已有关键词。

[0176] 对于多个推广单元的已有关键词,由于已有关键词涉及的行业和种类可能较多,有些可能与输入的查询词相差很远。为了规范引入的已有关键词和节省存储空间及处理时间,优选地,系统还包括一已有关键词过滤模块,对已有关键词进行过滤,去掉与输入的查询词相关性低的已有关键词,并将过滤后的已有关键词提供给所述词源扩充模块 104 使用。该模块先判断每个已有关键词与查询词之间的相关性,并将相关性小于某一预设值的已有关键词作为噪音排除掉,去掉与查询词相关性低的已有关键词。在过滤完后,词源扩充模块 104 才进行扩充推荐词的词源。

[0177] 本发明提供的关键词推荐方法是利用已有信息提取出用户实际查询需求的问题,令推荐结果可更好的满足不同用户的实际需求,并通过触发过程的限制,只对可识别的查询进行个性化推荐,防止出现盲目的个性化,控制个性化信息的权值,同时保证了推荐结果的多样性,间接地使潜在客户更容易匹配到用户的相关业务,帮助用户在新建或维护推广单元时选取好的关键词,起到很好的个性化推荐效果。

[0178] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明保护的范围之内。

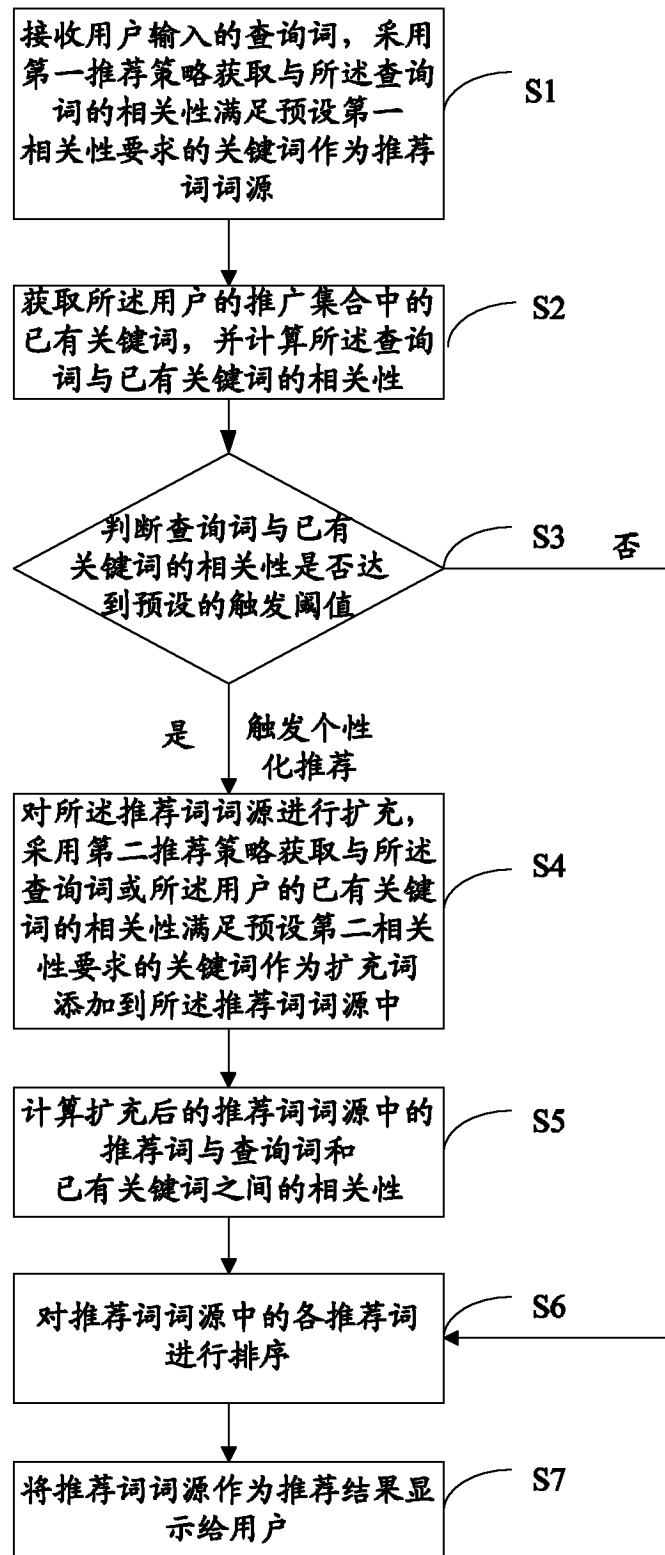


图 1

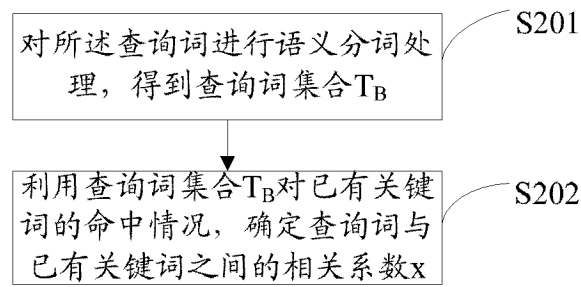


图 2

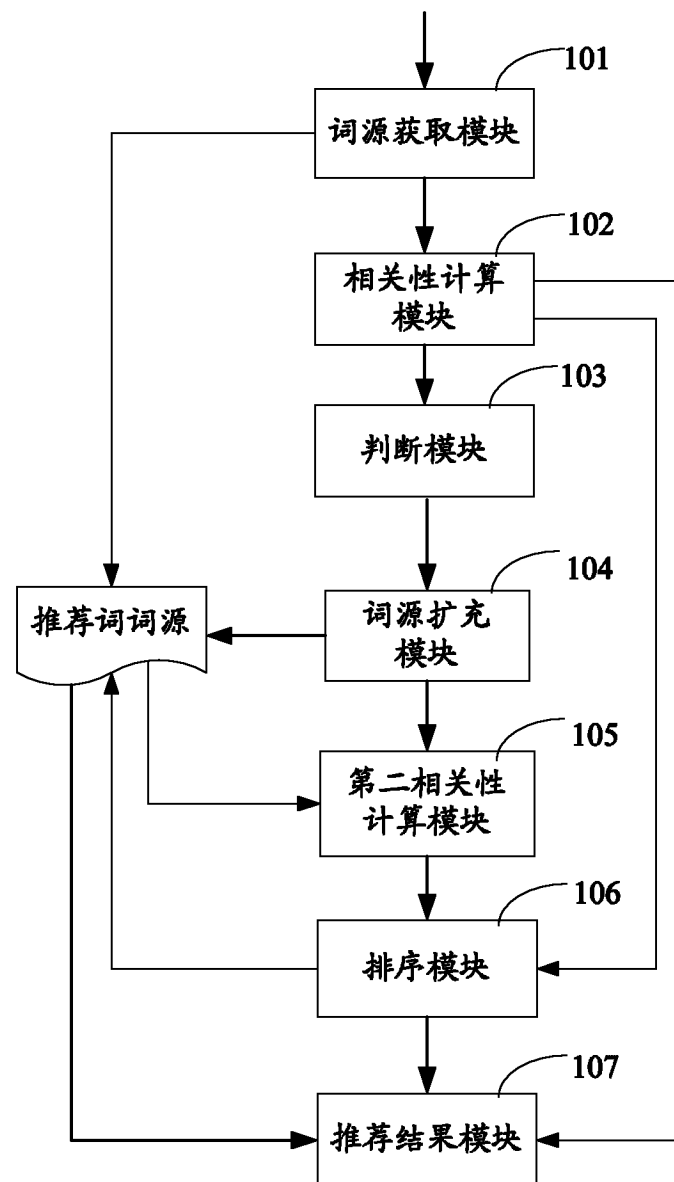


图 3