

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2012 年 9 月 7 日 (07.09.2012)



(10) 国际公布号
WO 2012/116562 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2011/083420
- (22) 国际申请日: 2011 年 12 月 3 日 (03.12.2011)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201110047892.7 2011 年 2 月 28 日 (28.02.2011) CN
- (71) 申请人 (对除美国外的所有指定国): 北京百度网讯科技有限公司 (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) [CN/CN]; 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。
- (72) 发明人; 及
- (75) 发明人/申请人 (仅对美国): 赵世奇 (ZHAO, Shiqi) [CN/CN]; 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 吴华 (WU, Hua) [CN/CN]; 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。 王海峰 (WANG, Haifeng) [CN/CN]; 中国北京市海淀区上地十街 10 号百度大厦 2 层, Beijing 100085 (CN)。

(74) 代理人: 深圳市威世博知识产权代理事务所 (普通合伙) (CHINA WISPRO INTELLECTUAL PROPERTY LLP.); 中国广东省深圳南山区高新区粤兴三道 8 号中国地质大学产学研基地中地大楼 A806, Guangdong 518057 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

[见续页]

(54) Title: METHOD AND DEVICE FOR CROSS-LANGUAGE QUERY

(54) 发明名称: 一种跨语言搜索的方法和装置

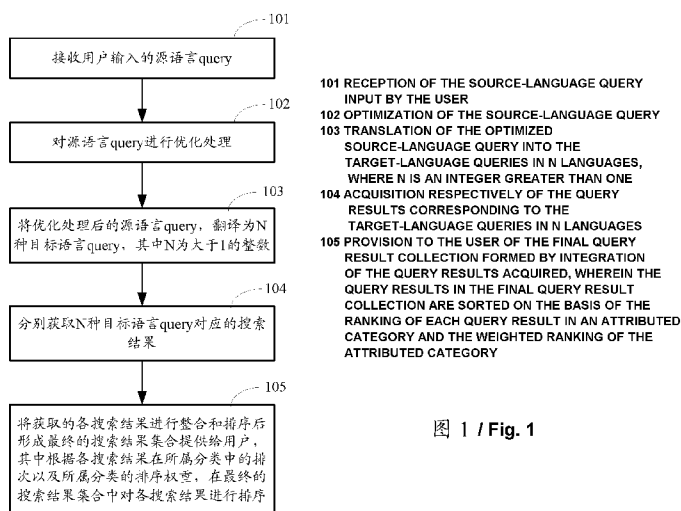


图 1 / Fig. 1

(57) Abstract: Provided in the present invention are a method and device for cross-language query. The method comprises: A. receiving a source-language query input by a user; B. translating the source-language query into target-language queries in N languages, where N is an greater than one; C. acquiring respectively query results corresponding to the target-language queries in N languages; D. providing the user with a final query result collection formed by integrating the query results acquired in step C, wherein the query results in the final query result collection are sorted on the basis of the ranking of each query result in an attributed category and the weighted ranking of the attributed category. The present invention allows for the query results to comprise files of multiple languages, providing the user with improved and increased number of query results.

[见续页]



WO 2012/116562 A1



本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

(57) 摘要:

本发明提供了一种跨语言搜索的方法和装置, 其中方法包括: A、接收用户输入的源语言搜索请求(query); B、将所述源语言 query 翻译为 N 种目标语言 query, N 为大于 1 的整数; C、分别获取所述 N 种目标语言 query 对应的搜索结果; D、将步骤 C 获取的搜索结果进行整合后形成最终的搜索结果集合提供给用户; 其中在所述最终的搜索结果集合中, 根据各搜索结果在所属分类中的排次以及所属分类的排序权重, 对各搜索结果进行排序。通过本发明能够实现包含多语言文档的搜索结果, 为用户提供更优、更多的搜索结果。

一种跨语言搜索的方法和装置

本申请要求了申请日为2011年02月28日，申请号为201110047892.7发明名称为“一种跨语言搜索的方法和装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本发明涉及互联网技术领域，特别涉及一种跨语言搜索的方法和装置。

背景技术

随着互联网信息的不断增长，人们对于信息搜索提出了更高的要求，不再满足于在同一种语种文档集中搜索，而要求获取多种语种文档。例如，如果用户输入的搜索词（query）为“贝克汉姆 图片”，则中文文档集中的搜索可能并不能最大程度地满足用户需求，欧美网站的英文文档集中可能具有更优、更多的搜索结果。

当从多语种文档集中进行搜索的需求越来越高时，为了获得更多、更全面、更准确的信息，同时为了跨越语言障碍，人们希望能够以一种自己熟悉的语言描述query，而搜索结果中能够包括多语言的文档，即进行两语种之间的跨语言搜索。

发明内容

有鉴于此，本发明提供了一种跨语言搜索的方法和装置，以便于实现包含多语言文档的搜索结果，为用户提供更优、更多的搜索结果。

具体技术方案如下：

一种跨语言搜索的方法，该方法包括：

- A、接收用户输入的源语言搜索请求 query；
- B、将所述源语言 query 翻译为 N 种目标语言 query，N 为大于 1 的整数；
- C、分别获取所述 N 种目标语言 query 对应的搜索结果；
- D、将步骤 C 获取的搜索结果进行整合后形成最终的搜索结果集合提供给用户；

其中在所述最终的搜索结果集合中，根据各搜索结果在所属分类中的排次以及所属分类的排序权重，对各搜索结果进行排序。

在步骤 B 中，针对每一种目标语言，将所述源语言 query 对应的该种目标语言的翻译结果中，翻译分值最高的一种翻译结果作为目标语言 query；

翻译结果 e 的翻译分值由以下因素中的至少一种确定：翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

较优地，所述步骤 B 具体包括：

B1、对所述源语言 query 进行优化处理，所述优化处理包括 query 纠错处理和 query 扩展处理中的任一种或组合；

B2、将优化处理后的源语言 query 翻译为 N 种目标语言 query。

其中，如果所述优化处理仅包括 query 纠错处理，则对所述用户输入的源语言 query 进行 query 纠错处理后得到包含 $n1$ 个 query 的源语言 query 集合 $Q1$ ， $n1$ 为预设的正整数；

所述步骤 B2 具体为：针对每一种目标语言，分别利用所述 $Q1$ 中的各 query 进行翻译，确定翻译分值总和最高的翻译结果作为目标语言

query; 其中, 翻译结果的翻译分值总和为 $\sum_{i=1}^{n1} P(e|q_i)$, $P(e|q_i)$ 为 Q1 中 q_i 被翻译为 e 的翻译分值;

翻译结果 e 对应的翻译分值由以下因素中的至少一种确定: 翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

如果所述优化处理仅包括 query 扩展处理, 则对所述用户输入的源语言 query 进行 query 扩展处理后得到包含 $n2$ 个 query 的源语言 query 集合 Q2, $n2$ 为预设的正整数;

所述步骤 B2 具体为: 针对每一种目标语言, 分别利用所述 Q2 中的各 query 进行翻译, 确定翻译分值总和最高的翻译结果作为目标语言 query; 其中, 翻译结果的翻译分值总和为 $\sum_{i=1}^{n2} P(e|q_i)$, $P(e|q_i)$ 为 Q2 中 q_i 被翻译为 e 的翻译分值;

翻译结果 e 对应的翻译分值由以下因素中的至少一种确定: 翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

如果所述优化处理既包括 query 纠错处理又包括 query 扩展处理, 则对所述用户输入的源语言 query 进行 query 纠错处理和 query 扩展处理后得到包含 n 个 query 的源语言 query 集合 Q, n 为预设的正整数;

所述步骤 B2 具体为: 针对每一种目标语言, 分别利用所述 Q 中的各 query 进行翻译, 确定翻译分值总和最高的翻译结果作为目标语言 query; 其中, 翻译结果的翻译分值总和为 $\sum_{i=1}^n P(e|q_i)$, $P(e|q_i)$ 为 Q 中 q_i 被翻译为 e 的翻译分值;

翻译结果 e 的翻译分值由以下因素中的至少一种确定: 翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

其中, 对所述用户输入的源语言 query 进行 query 纠错处理后和 query 扩展处理后得到包含 n 个 query 的源语言 query 集合 Q 具体包括:

对所述用户输入的源语言 query 进行 query 纠错处理后得到包含 $n1$ 个 query 的源语言 query 集合 $Q1$, $n1$ 为预设的正整数, 将所述 $Q1$ 中的各 query 分别进行 query 扩展处理, 得到包含 n 个 query 的源语言 query 集合 Q ; 或者,

对所述用户输入的源语言 query 进行 query 扩展处理后得到包含 $n2$ 个 query 的源语言 query 集合 $Q2$, $n2$ 为预设的正整数, 将所述 $Q2$ 中的各 query 分别进行 query 纠错处理, 得到包含 n 个 query 的源语言 query 集合 Q ; 或者,

对所述用户输入的源语言 query 同时进行 query 纠错处理和 query 扩展处理后, 分别得到包含 $n1$ 个 query 的源语言 query 集合 $Q1$ 和包含 $n2$ 个 query 的源语言 query 集合 $Q2$, 将所述 $Q1$ 和 $Q2$ 取并集后, 得到包含 n 个 query 的源语言 query 集合 Q 。

对所述用户输入的源语言 query 进行 query 纠错处理具体包括:

利用所述用户输入的源语言 query 查找纠错训练语料, 判断纠错训练语料中是否存在与所述用户输入的源语言 query 相同的错误 query, 如果是, 则确定与所述用户输入的源语言 query 相同的错误 query 所对应的所有正确 query, 从确定的所有正确 query 中选择对应纠错概率排在前 $n1$ 个的正确 query 构成源语言 query 集合 $Q1$; 否则, 所述 $Q1$ 中仅包括所述用户输入的源语言 query;

其中,所述纠错训练语料包括:预先从搜索日志中收集的错误 query 和对应正确 query 构成的 query 对,以及错误 query 被纠错为对应正确 query 的纠错概率。

对所述用户输入的源语言 query 进行 query 扩展处理具体包括:

将所述用户输入的源语言 query 进行分词处理,通过查找源语言的复述资源确定分词处理后得到的各词语的同义词,利用分词处理后得到的各词语及各词语的同义词进行组合,取组合得到的 query 中扩展分值排在前 n^2 个的 query 构成所述 Q2;

query 的扩展分值由创建所述复述资源中该 query 的统计次数确定。

另外,所述步骤 C 还包括:

获取所述源语言 query 对应的搜索结果。

对于对源语言 query 进行优化处理的情况,所述步骤 C 还包括:从优化处理后得到的源语言 query 集合中选择一个 query,获取选择的 query 对应的搜索结果。

所述从优化处理后得到的源语言 query 集合中选择一个 query 时,使用的选择策略包括:

对优化处理后得到的源语言 query 集合中的各 query 逐一进行搜索,直至找到搜索效果满足预设要求的 query,选择该搜索效果满足预设要求的 query; 或者,

对优化处理后得到的源语言 query 集合中的各 query 进行搜索,选择搜索效果最优的 query。

具体地,步骤 D 中所述整合包括:对步骤 C 获取的搜索结果进行合并和去重。

所述根据各搜索结果在所属分类中的排次以及所属分类的排序权重，对各搜索结果进行排序具体包括：

利用各搜索结果在所属分类中的排次以及所属分类的排序权重，对各搜索结果进行打分，按照打分结果从高到低对各搜索结果进行排序；

其中，搜索结果 Rst 的打分结果 $score(Rst)$ 为： $score(Rst)=$

$$\frac{1}{m} \sum_{i=1}^m w_i \frac{1}{rank_i(Rst)},$$

m 为搜索结果分类的总个数， w_i 为第 i 种分类的排序权重， $rank_i(Rst)$ 为 Rst 在第 i 种分类中的排次。

特别地，搜索结果所属分类为：搜索结果对应的语言；

第 i 种分类的排序权重的确定方法具体为：

S1、提取所述用户输入的源语言 query 的特征；

S2、将步骤 S1 提取的特征与各语言的特征向量进行相似度计算，确定相似度超过预设的相似度阈值的语言为所述用户输入的源语言 query 的映射语言；

S3、对于搜索结果 Rst，如果 Rst 所属分类为映射语言，则该所属分类的排序权重 w_i 为第一设定值 a ；如果 Rst 所属分类为源语言且源语言不是该所属分类的映射语言，则该所属分类的排序权重 w_i 为第二设定值 b ；如果 Rst 所属分类既不是映射语言也不是源语言，则该所属分类的排序权重 w_i 为第三设定值 c ；

其中， $a > b > c$ ，各语言的特征向量是预先对各语言的已有资源进行挖掘所训练出来的。

一种跨语言搜索的装置，该装置包括：用户侧交互单元、翻译处理单元、搜索处理单元和结果整合单元；

所述用户侧交互单元，用于接收用户输入的源语言搜索请求 query，

将所述结果整合单元整合后形成的搜索结果集合提供给所述用户；

所述翻译处理单元，用于将所述源语言 query 翻译为 N 种目标语言 query，N 为大于 1 的整数；

所述搜索处理单元，用于分别获取所述 N 种目标语言 query 对应的搜索结果；

所述结果整合单元，用于将所述搜索处理单元获取的搜索结果进行整合后形成最终的搜索结果集合；其中，在所述最终的搜索结果集合中，根据各搜索结果在所属分类中的排次以及所属分类的排序权重，对各搜索结果进行排序。

所述翻译处理单元针对每一种目标语言，将所述源语言 query 对应的该种目标语言的翻译结果中，翻译分值最高的一种翻译结果作为目标语言 query；

翻译结果 e 的翻译分值由以下因素中的至少一种确定：翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

较优地，该装置还包括：优化处理单元；

所述优化处理单元，用于对所述用户输入的源语言 query 进行优化处理后提供给所述翻译处理单元，所述优化处理包括 query 纠错处理和 query 扩展处理中的任一种或组合；

所述翻译处理单元将所述优化处理单元进行优化处理后的源语言 query 翻译为 N 种目标语言 query。

如果所述优化处理仅包括 query 纠错处理，则所述优化处理单元对所述用户输入的源语言 query 进行 query 纠错处理后得到包含 $n1$ 个 query

的源语言 query 集合 Q1, n1 为预设的正整数;

所述翻译处理单元针对每一种目标语言, 分别利用所述 Q1 中的各 query 进行翻译, 确定翻译分值总和最高的翻译结果作为目标语言 query;

其中, 翻译结果的翻译分值总和为 $\sum_{i=1}^{n1} P(e|q_i)$, $P(e|q_i)$ 为 Q1 中 q_i 被翻译为 e 的翻译分值;

翻译结果 e 对应的翻译分值由以下因素中的至少一种确定: 翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

如果所述优化处理仅包括 query 扩展处理, 则所述优化处理单元对所述用户输入的源语言 query 进行 query 扩展处理后得到包含 n2 个 query 的源语言 query 集合 Q2, n2 为预设的正整数;

所述翻译处理单元针对每一种目标语言, 分别利用所述 Q2 中的各 query 进行翻译, 确定翻译分值总和最高的翻译结果作为目标语言 query;

其中, 翻译结果的翻译分值总和为 $\sum_{i=1}^{n2} P(e|q_i)$, $P(e|q_i)$ 为 Q2 中 q_i 被翻译为 e 的翻译分值;

翻译结果 e 对应的翻译分值由以下因素中的至少一种确定: 翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

如果所述优化处理既包括 query 纠错处理又包括 query 扩展处理, 则所述优化处理单元对所述用户输入的源语言 query 进行 query 纠错处理和 query 扩展处理后得到包含 n 个 query 的源语言 query 集合 Q, n 为预设的正整数;

所述翻译处理单元针对每一种目标语言，分别利用所述 Q 中的各 query 进行翻译，确定翻译分值总和最高的翻译结果作为目标语言 query；其中，翻译结果的翻译分值总和为 $\sum_{i=1}^n P(e|q_i)$ ， $P(e|q_i)$ 为 Q 中 q_i 被翻译为 e 的翻译分值；

翻译结果 e 的翻译分值由以下因素中的至少一种确定：翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

具体地，所述优化处理单元可以包括：第一纠错模块和第一扩展模块；

所述第一纠错模块，用于对所述用户输入的源语言 query 进行 query 纠错处理后得到包含 $n1$ 个 query 的源语言 query 集合 $Q1$ ， $n1$ 为预设的正整数；

所述第一扩展模块，用于将所述 $Q1$ 中的各 query 分别进行 query 扩展处理，得到包含 n 个 query 的源语言 query 集合 Q 。

或者，所述优化处理单元具体包括：第二扩展模块和第二纠错模块；

所述第二扩展模块，用于对所述用户输入的源语言 query 进行 query 扩展处理后得到包含 $n2$ 个 query 的源语言 query 集合 $Q2$ ， $n2$ 为预设的正整数；

所述第二纠错模块，用于将所述 $Q2$ 中的各 query 分别进行 query 纠错处理，得到包含 n 个 query 的源语言 query 集合 Q 。

或者，所述优化处理单元具体包括：第三纠错模块、第三扩展模块和合并处理模块；

所述第三纠错模块，用于对所述用户输入的源语言 query 进行 query 纠错处理，得到包含 $n1$ 个 query 的源语言 query 集合 $Q1$ ；

所述第三扩展模块,用于对所述用户输入的源语言 query 进行 query 扩展处理,得到包含 n_2 个 query 的源语言 query 集合 Q_2 ;

所述合并处理模块,用于将所述 Q_1 和 Q_2 取并集后,得到包含 n 个 query 的源语言 query 集合 Q 。

所述优化处理单元具体利用所述用户输入的源语言 query 查找纠错训练语料,判断纠错训练语料中是否存在与所述用户输入的源语言 query 相同的错误 query,如果是,则确定与所述用户输入的源语言 query 相同的错误 query 所对应的所有正确 query,从确定的所有正确 query 中选择对应纠错概率排在前 n_1 个的正确 query 构成源语言 query 集合 Q_1 ;否则,所述 Q_1 中仅包括所述用户输入的源语言 query;

其中,所述纠错训练语料包括:预先从搜索日志中收集的错误 query 和对应正确 query 构成的 query 对,以及错误 query 被纠错为对应正确 query 的纠错概率。

所述优化处理单元具体将所述用户输入的源语言 query 进行分词处理,通过查找源语言的复述资源确定分词处理后得到的各词语的同义词,利用分词处理后得到的各词语及各词语的同义词进行组合,取组合得到的 query 中扩展分值排在前 n_2 个的 query 构成所述 Q_2 ;

query 的扩展分值由创建所述复述资源中该 query 的统计次数确定。

另外,所述搜索处理单元,还用于获取所述源语言 query 对应的搜索结果。

对应于包括优化处理单元的情况,该装置还包括:

源请求选择单元,用于从所述优化处理单元进行优化处理后得到的源语言 query 集合中选择一个 query;

所述搜索处理单元，还用于获取所述源请求选择单元选择的 query 对应的搜索结果。

其中，所述源请求选择单元采用的选择策略包括：

对所述优化处理单元进行优化处理后得到的源语言 query 集合中的各 query 逐一进行搜索，直至找到搜索效果满足预设要求的 query，选择该搜索效果满足预设要求的 query；或者，

对所述优化处理单元进行优化处理后得到的源语言 query 集合中的各 query 进行搜索，选择搜索效果最优的 query。

具体地，所述结果整合单元包括：合并处理模块、去重处理模块和排序处理模块；

所述合并处理模块，用于将所述搜索处理单元获取的搜索结果进行合并处理；

所述去重处理单元，用于将所述合并处理模块合并处理后的搜索结果进行去重处理得到搜索结果集合；

所述排序处理模块，用于在所述搜索结果集合中，根据各搜索结果在所属分类中的排次以及所属分类的排序权重，对各搜索结果进行排序。

所述排序处理模块具体利用所述搜索结果集合中各搜索结果在所属分类中的排次以及所属分类的排序权重，对各搜索结果进行打分，按照打分结果从高到低对各搜索结果进行排序；

其中，搜索结果 Rst 的打分结果 $score(Rst)$ 为： $score(Rst) =$

$$\frac{1}{m} \sum_{i=1}^m w_i \frac{1}{rank_i(Rst)},$$

m 为搜索结果分类的总个数， w_i 为第 i 种分类的排序权重， $rank_i(Rst)$ 为 Rst 在第 i 种分类中的排次。

重， $rank_i(Rst)$ 为 Rst 在第 i 种分类中的排次。

搜索结果所属分类为：搜索结果对应的语言；

所述排序处理模块在确定第 i 种分类的排序权重时，具体执行以下操作：提取所述用户输入的源语言 query 的特征；将提取的特征与各语言的特征向量进行相似度计算，确定相似度超过预设的相似度阈值的语言为所述用户输入的源语言 query 的映射语言；对于搜索结果 Rst，如果 Rst 所属分类为映射语言，则该所属分类的排序权重 w_i 为第一设定值 a；如果 Rst 所属分类为源语言且源语言不是该所属分类的映射语言，则该所属分类的排序权重 w_i 为第二设定值 b；如果 Rst 所属分类既不是映射语言也不是源语言，则该所属分类的排序权重 w_i 为第三设定值 c；

其中， $a > b > c$ ，各语言的特征向量是预先对各语言的已有资源进行挖掘所训练出来的。

由以上技术方案可以看出，本发明将源语言 query 翻译为多种目标语言 query 后，获取 N 种目标语言 query 对应的搜索结果，对搜索结果进行整合后形成最终的搜索结果集合提供给用户，其中在最终的搜索结果集合中，根据各搜索结果在所属分类中的排次以及所属分类的排序权重，对各搜索结果进行排序。通过本发明实现了包含多语言文档的搜索结果，为用户提供了更优、更多的搜索结果。

附图说明

图 1 为本发明实施例一提供的方法流程图；

图 2 为本发明实施例二提供的示意图；

图 3 为本发明实施例三提供的装置结构图；

图 4 中的(a)(b)(c)为本发明实施例三提供的优化处理单元的三种结构图。

具体实施方式

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。

实施例一、

图 1 为本发明实施例一提供的方法流程图，如图 1 所示，该方法可以包括以下步骤：

步骤 101：接收用户输入的源语言 query。

该方法可以在搜索引擎的服务器端实现，在用户输入源语言 query 后，浏览器将用户输入的源语言 query 发送给搜索引擎的服务器端。

步骤 102：对源语言 query 进行优化处理。

本步骤为可选步骤，对源语言 query 进行的优化处理可以包括：query 纠错处理和 query 扩展处理中的任一种或组合。其中 query 纠错处理的目的是为了提高源语言 query 被正确翻译的可能性，query 扩展处理的目的是为了扩大 query 的资源召回率。

其中，query 纠错处理旨在纠正 query 中的错误，主要是拼写错误，可以基于噪声信道模型实现。具体包括：预先从搜索日志中收集大量的错误 query 和对应正确 query 构成的 query 对，并计算错误 query 被纠错为正确 query 的概率，即纠错概率 $P(p|p')$ ，组成纠错训练语料；针对用户输入的源语言 query 查询纠错训练语料，判断纠错训练语料中是否存在与源语言 query 相同的错误 query，如果是，则确定与源语言 query 相同的错误 query 所对应的所有正确 query，从中选择对应纠错概率排在前 $n1$ 个的正确 query 构成源语言 query 进行纠错处理后的源语言 query 集合 $Q1$ ， $n1$ 为预设的正整数；如果否，则无需对源语言 query 进行纠错处理，最终得到的 $Q1$ 中仅包括用户输入的源语言 query。

其中, $P(p'|p) = P(p|p')P(p')/P(p)$ 。

$P(p'|p)$ 为源语言 query p 被纠错为 p' 的纠错概率, $P(p|p')$ 为搜索日志中 p' 被错写成 p 的概率, $P(p')$ 是 p' 在搜索日志中出现的次数, $P(p)$ 是 p 在搜索日志中出现的次数。

query 扩展处理的目的在于扩展出同义 query, 以避免由于用户输入的源语言 query 中包含专用名词或新词所带来的翻译困难的问题。例如, 如果用户输入的源语言 query 为“诺基亚 n8 美图”, 其中“美图”多出现在网络用语中, 在一般的翻译模型中不常出现, 这就会给后续翻译源语言 query 带来难度, 但倘若将其进行同义扩展为“图片”, 就会降低翻译难度。

在进行同义扩展时, 可以将接收到的源语言 query 扩展为一个 query 集合 $Q2$, 假设源语言 query 为 q , 扩展出的 $Q2$ 为 $\{q_1, q_2, \dots, q_{n2}\}$, 其中, $n2$ 为预设的正整数, 扩展出的 query 集合 $Q2$ 中包含 q 以及由 q 扩展出的同义 query。

具体地, 可以首先将源语言 query 进行分词处理, 通过查找源语言的复述资源, 例如同义词词典, 确定分词处理后得到的各词语的同义词, 利用分词处理后得到的各词语及其同义词进行组合后, 选取组合得到的 query 中扩展分值在前 $n2$ 个的 query 构成集合 $Q2$ 。其中, query 的扩展分值可以由创建复述资源中 query 出现的统计次数确定。需要特别说明的是, 复述资源不限于词, 也可以为短语, 甚至为句子, 例如基于词典注释的替换、语序变换、句子结构变换、句子拆分与合并或基于推理的复述得到的资源, 只要描述的事物相同, 表达的含义相同, 都可以认为是复述资源。

例如, 如果 n_2 为 2, 用户输入的源语言 query 为“中国 美食”, 进行分词处理后得到“中国”和“美食”, “中国”的同义词有“中华”; “美食”的同义词有“佳肴”, 进行组合后得到的 query 为“中国 美食”, “中国 佳肴”, “中华 美食”, “中华 佳肴”, 选取扩展分值排在前 2 个的 query 构成 $Q_2\{\text{“中国 美食”, “中华 美食”}\}$ 。

如果上述优化处理包括: query 纠错处理以及 query 扩展处理, 可以首先进行 query 纠错处理, 得到源语言 query 集合 Q_1 , 对 query 纠错处理后得到的 Q_1 中的各 query 分别进行 query 扩展处理, 最终得到集合 $Q\{q_1, q_2, \dots, q_n\}$, n 为预设的正整数; 也可以首先进行 query 扩展处理, 得到源语言 query 集合 Q_2 , 然后对 query 扩展处理后得到的 Q_2 中的各 query 分别进行 query 纠错处理, 最终得到集合 $Q\{q_1, q_2, \dots, q_n\}$, 或者, query 纠错处理以及 query 扩展处理可以同时进行, 将 query 纠错处理以及 query 扩展处理得到的集合 Q_1 和 Q_2 取并集后, 得到集合 $Q\{q_1, q_2, \dots, q_n\}$ 。

步骤 103: 将优化处理后的源语言 query, 翻译为 N 种目标语言 query, 其中 N 为大于 1 的整数。

本发明实施例中可以预先将 N 种常用的语言种类作为目标语言, 例如, 预先将常用的英文、法文、日文、德文等设置为目标语言。具体设置哪几种目标语言可以根据需求灵活设置, 本发明对此并不加限制。

如果没有对用户输入的源语言 query 执行步骤 102 中的优化处理, 则直接将用户输入的源语言 query 翻译为 N 种目标语言 query。

在对源语言 query 翻译成第 m 种目标语言时, 可能会得到多种不同的翻译结果, 针对这种情况, 可以选择多种不同的翻译结果中翻译分值

最高的一种翻译结果。即如果源语言 query 为 q ，则针对第 m 种目标语言的翻译结果 $\hat{e}$ 为： $\hat{e} = \arg \max_e P(e|q)$ ，其中， $P(e|q)$ 表示 q 被翻译为 e 的翻译分值。

其中，翻译结果的翻译分值可以由以下两种因素中的至少一种确定：翻译所使用的翻译语料库中该翻译结果出现的统计次数以及翻译结果中各词的组合概率。

如果对用户输入的源语言 query 执行了步骤 102 中的优化处理，则将优化处理后的源语言 query 翻译为 N 种目标语言 query。如果在进行优化处理后得到了扩展出的 query 集合 Q ，则针对每一种目标语言，分别将 Q 中的各 query 都翻译为目标语言 query，从中确定出一个翻译分值总和最高的翻译结果作为该种目标语言的翻译结果。

即针对每一种目标语言，其翻译结果 $\hat{e}$ 为： $\hat{e} = \arg \max_e \sum_{i=1}^n P(e|q_i)$ 。 (1)

其中 $P(e|q_i)$ 表示 q 扩展出的集合 Q 中 q_i 被翻译为 e 的翻译分值，如果 q_i 不能被翻译为 e ，则对应翻译分值为零， $\sum_{i=1}^n P(e|q_i)$ 表示集合 Q 中能够被翻译为 e 的所有 q_i 的翻译分值总和。例如，用户输入的源语言 query 为“中国 美食”，优化处理后得到的集合 Q 为{“中国 美食”，“中国 佳肴”，“中华 美食”，“中华 佳肴”}，针对目标语言为英文的情况，“中国 美食”、“中华 美食”和“中华 佳肴”都被翻译为“Chinese cuisine”，则如果这三个 query 被翻译为“Chinese cuisine”的翻译分值总和最大，则将“Chinese cuisine”作为英文的翻译结果。

需要说明的是，如果在优化处理时仅进行了 query 纠错处理，则 Q 就是步骤 102 中所述的 $Q1$ ，公式 (1) 中 n 即为 $n1$ ；如果在优化处理时

仅进行了 query 扩展处理, 则 Q 就是步骤 102 中所述的 Q2, 公式 (1) 中 n 即为 n2。

另外, 在对用户输入的源语言 query 或者由源语言 query 扩展出的 Q 中各 query 进行翻译时, 可以通过匹配翻译模型的方式实现。更优地, 翻译模型中可以包含预先利用挖掘出的专用名词、新词所训练出的翻译模型。

步骤 104: 分别获取 N 种目标语言 query 对应的搜索结果。

在翻译得到 N 种目标语言 query 时, 分别在对应目标语言的资源库中进行搜索。在每一种目标语言的资源库中进行搜索时, 可以是普通搜索, 即在对应目标语言的非结构化资源库中进行搜索; 也可以是垂直搜索, 即在对应目标语言的结构化资源库中进行搜索。

另外, 除了将源语言 query 翻译为 N 种目标语言 query 并获取目标语言 query 对应的搜索结果之外, 还可以同时获取源语言 query 的搜索结果。具体地, 如果不对源语言 query 进行优化处理, 则直接利用用户输入的源语言 query 获取搜索结果。如果对源语言 query 进行了优化处理, 则从优化处理后得到的源语言 query 集合 Q 中选择一个 query, 利用该选择的 query 在源语言的资源库中进行搜索获取对应的搜索结果。

在从源语言 query 集合 Q 中选择 query 时, 采用的选择的策略可以包括但不限于:

对集合 Q 中的各 query 逐一进行搜索, 直至找到搜索效果满足预设要求的 query, 选择该搜索效果满足预设要求的 query 作为目标语言 query; 或者, 对集合 Q 中的各 query 进行搜索, 选择搜索效果最优的 query 作为目标语言 query。

其中，搜索效果可以体现为搜索结果数量，或者，搜索结果中与源语言 query 的相关度满足预设相关度要求的搜索结果数量，或者，在设定时间内发布的搜索结果数量，或者，搜索结果的来源满足预设来源要求的搜索结果数量，等等。

步骤 105：将获取的各搜索结果进行整合和排序后形成最终的搜索结果集合提供给用户，其中根据各搜索结果在所属分类中的排次以及所属分类的排序权重，在最终的搜索结果集合中对各搜索结果进行排序。

由于从 N 种目标语言的资源库中都获得了搜索结果，或者更进一步从源语言的资源库中也获得了搜索结果，因此，需要对获得的各搜索结果进行整合，其中整合包括：搜索结果的合并和去重。

在对整合后的搜索结果进行排序时，可以利用各搜索结果所属分类的排序权重以及在所属分类中的排次为各搜索结果打分，然后按照打分结果从高到低进行排序。

具体地，搜索结果 Rst 的打分结果 $score(Rst)$ 为：

$$score(Rst) = \frac{1}{m} \sum_{i=1}^m w_i \frac{1}{rank_i(Rst)}, \quad (2)$$

其中，m 为搜索结果分类的总个数， w_i 为第 i 种分类的排序权重， $rank_i(Rst)$ 为 Rst 在第 i 种分类中的排次。

通常，搜索结果所属分类可以由其对应的语言确定，即搜索结果来源于哪种语言的搜索。由于在有些情况下，通过不同的语言得到的搜索结果可能存在重合，因此，公式 (2) 实际上是求加权平均值的处理。例如，某一个搜索结果既是中文的搜索结果也是英文的搜索结果，则将该搜索结果在英文搜索结果中的排次乘以英文的排序权重，再加上该搜索结果在中文搜索结果中的排次乘以中文的排序权重，获得的和再除以搜

索所使用语言的总个数，得到该搜索结果的打分结果。

本发明实施例中，可以预先对各语言（包括各目标语言和源语言）的已有资源进行挖掘，训练出特征向量；在确定各语言的排序权重时，提取源语言 query 的特征，将提取的特征与各语言的特征向量进行相似度计算，确定相似度超过预设的相似度阈值的语言为该源语言 query 的映射语言。

例如，对于按照语言进行分类，对日文对应的已有资源进行挖掘，训练出特征向量包括：东京、日本、相扑、和服、小泉、樱花、木村拓哉……。

当用户输入的源语言 query 为“木村拓哉有哪些著名电影”，则提取源语言 query 的特征为“木村拓哉 电影”，对提取的特征与各语言的特征向量进行相似度计算，确定与日文的分类的特征向量之间的相似度超过预设的相似度阈值，则确定该源语言 query 的映射语言为日文。

如果搜索结果所属分类为映射语言，则该分类的排序权重 w_i 为第一设定值 a；如果搜索结果所属分类为源语言且源语言不是该所属分类的映射语言，则所属分类的排序权重 w_i 为第二设定值 b；如果搜索结果所属分类既不是映射语言也不是源语言，则所属分类的排序权重 w_i 为第三设定值 c。其中， $a > b > c$ ，例如可以取 $b=1$ ， $a>1, c<1$ 。

另外，搜索结果在其所属分类中的排次可以由搜索结果与搜索所使用 query 之间的相关度、搜索结果所来源的站点权重、搜索结果的发布时间等中的一种或任意组合确定。

至此实施例一所示流程结束。下面通过实施例二中的一个具体实例对上述方法进行描述。

实施例二、

实施例二对应的示意图如图 2 所示, 假设用户输入中文 query: “小贝 美图”。

对该 query 进行优化处理, 经过纠错处理后确定该 query 正确无需纠错, 然后查找同义词词典, 利用词语“小贝”可以扩展出“贝克汉姆”、词语“美图”可以扩展出“图片”, 最终扩展后得到集合 Q 为: {“小贝 图片”, “贝克汉姆 图片”, “小贝 美图”, “贝克汉姆 美图”}。

假设搜索引擎的服务器端预先设置了 3 种目标语言: 英文、日文和法文, 另外还需要针对源语言(中文)进行搜索, 即最终利用 4 种语言进行搜索。其中, 在翻译成英文时, “小贝 图片”和“小贝 美图”可以翻译成“Beckham picture”、“Cockles picture”, “小贝 图片”和“小贝 美图”翻译成“Beckham picture”对应的翻译分值分别为 8 和 6, “小贝 图片”和“小贝 美图”翻译成“Cockles picture”对应的翻译分值分别为 1 和 2; “贝克汉姆 图片”和“贝克汉姆 美图”均可以翻译成“Beckham picture”, 对应的翻译分值分别为 9 和 7。

计算“Beckham picture”的翻译分值总和为 $8+6+9+7=30$, 计算“Cockles picture”的翻译分值总和为 $1+2=3$ 。

最终选取英文的翻译结果为“Beckham picture”。

其他目标语言的翻译结果类似, 不再举例描述。

然后, 分别利用各目标语言的翻译结果, 作为目标语言 query 分别到对应目标语言的资源库中进行搜索, 得到各搜索结果。

针对源语言, 从集合 Q 中选择搜索效果最好的 query “贝克汉姆 图片”, 并获取该选择的 query 的搜索结果。

将中文、英文、日文和法文对应的搜索结果进行整合。

在对整合后的搜索结果进行排序时，按照公式(2)的方式进行。在对源语言 query “小贝 美图”进行特征提取以及相似度匹配后，确定映射语言为英文，则取英文的排序权重为 2，中文的排序权重为 1，日文和法文的排序权重为 0.5。

对于其中一个搜索结果 Page1，如果该 Page1 在中文检索结果、英文检索结果和日文检索结果中都存在，且在中文检索结果中的排次为 2，在英文检索结果中的排次为 5，在日文检索结果中的排次为 8，则该 Page1

的打分结果 $score(\text{Page1})$ 为： $score(\text{Page1}) = \frac{1}{4} \times (1 \times \frac{1}{2} + 2 \times \frac{1}{5} + 0.5 \times \frac{1}{8}) = 0.24$ 。

实施例三、

图 3 为本发明实施例提供的装置结构图，如图 3 所示，该装置可以包括：用户侧交互单元 300、翻译处理单元 310、搜索处理单元 320 和结果整合单元 330。

用户侧交互单元 300，用于接收用户输入的源语言搜索请求 query，将结果整合单元 330 整合后形成的搜索结果集合提供给用户。

翻译处理单元 310，用于将源语言 query 翻译为 N 种目标语言 query，N 为大于 1 的整数。

搜索处理单元 320，用于分别获取 N 种目标语言 query 对应的搜索结果。

结果整合单元 330，用于将搜索处理单元 320 获取的搜索结果进行整合后形成最终的搜索结果集合；其中，在最终的排序结果中，根据各搜索结果在所属分类中的排次以及所属分类的排序权重，对各搜索结果进行排序。

具体地，翻译处理单元 310 可以针对每一种目标语言，将源语言 query 对应的该种目标语言的翻译结果中，翻译分值最高的一种翻译结果作为目标语言 query。其中，翻译结果 e 的翻译分值由以下因素中的至少一种确定：翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

较优地，该装置还可以包括：优化处理单元 340。

优化处理单元 340，用于对用户输入的源语言 query 进行优化处理后提供给翻译处理单元 310，其中优化处理可以包括 query 纠错处理和 query 扩展处理中的任一种或组合。

翻译处理单元 310 将优化处理单元 340 进行优化处理后的源语言 query 翻译为 N 种目标语言 query。

如果装置中包含优化处理单元 340，则根据优化处理的形式可以存在以下三种情况：

第一种情况：如果优化处理仅包括 query 纠错处理，则优化处理单元 340 对用户输入的源语言 query 进行 query 纠错处理后得到包含 $n1$ 个 query 的源语言 query 集合 $Q1$ ， $n1$ 为预设的正整数。

此时，翻译处理单元 310 针对每一种目标语言，分别利用 $Q1$ 中的各 query 进行翻译，确定翻译分值总和最高的翻译结果作为目标语言 query；其中，翻译结果的翻译分值总和为 $\sum_{i=1}^{n1} P(e|q_i)$ ， $P(e|q_i)$ 为 $Q1$ 中 q_i 被翻译为 e 的翻译分值。

翻译结果 e 对应的翻译分值由以下因素中的至少一种确定：翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

第二种情况: 如果优化处理仅包括 query 扩展处理, 则优化处理单元 340 对用户输入的源语言 query 进行 query 扩展处理后得到包含 n_2 个 query 的源语言 query 集合 Q_2 , n_2 为预设的正整数。

翻译处理单元 310 针对每一种目标语言, 分别利用 Q_2 中的各 query 进行翻译, 确定翻译分值总和最高的翻译结果作为目标语言 query; 其中, 翻译结果的翻译分值总和为 $\sum_{i=1}^{n_2} P(e|q_i)$, $P(e|q_i)$ 为 Q_2 中 q_i 被翻译为 e 的翻译分值。

翻译结果 e 对应的翻译分值由以下因素中的至少一种确定: 翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

第三种情况: 如果优化处理既包括 query 纠错处理又包括 query 扩展处理, 则优化处理单元 340 对用户输入的源语言 query 进行 query 纠错处理和 query 扩展处理后得到包含 n 个 query 的源语言 query 集合 Q , n 为预设的正整数。

翻译处理单元 310 针对每一种目标语言, 分别利用 Q 中的各 query 进行翻译, 确定翻译分值总和最高的翻译结果作为目标语言 query; 其中, 翻译结果的翻译分值总和为 $\sum_{i=1}^n P(e|q_i)$, $P(e|q_i)$ 为 Q 中 q_i 被翻译为 e 的翻译分值。

翻译结果 e 的翻译分值由以下因素中的至少一种确定: 翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

在第三种情况中, 优化处理单元 340 可以存在三种结构:

第一种结构: 如图 4 中的(a)所示, 优化处理单元 340 可以具体包括:

第一纠错模块 401 和第一扩展模块 402。

第一纠错模块 401, 用于对用户输入的源语言 query 进行 query 纠错处理后得到包含 n_1 个 query 的源语言 query 集合 Q_1 , n_1 为预设的正整数。

第一扩展模块 402, 用于将 Q_1 中的各 query 分别进行 query 扩展处理, 得到包含 n 个 query 的源语言 query 集合 Q 。

第二种结构: 如图 4 中的(b)所示, 优化处理单元 340 可以具体包括: 第二扩展模块 411 和第二纠错模块 412。

第二扩展模块 411, 用于对用户输入的源语言 query 进行 query 扩展处理后得到包含 n_2 个 query 的源语言 query 集合 Q_2 , n_2 为预设的正整数。

第二纠错模块 412, 用于将 Q_2 中的各 query 分别进行 query 纠错处理, 得到包含 n 个 query 的源语言 query 集合 Q 。

第三种结构: 如图 4 中的(c)所示, 优化处理单元 340 可以具体包括: 第三纠错模块 421、第三扩展模块 422 和合并处理模块 423。

第三纠错模块 421, 用于对用户输入的源语言 query 进行 query 纠错处理, 得到包含 n_1 个 query 的源语言 query 集合 Q_1 。

第三扩展模块 422, 用于对用户输入的源语言 query 进行 query 扩展处理, 得到包含 n_2 个 query 的源语言 query 集合 Q_2 。

合并处理模块 423, 用于将 Q_1 和 Q_2 取并集后, 得到包含 n 个 query 的源语言 query 集合 Q 。

具体地, 图 3 中所示的优化处理单元 340 进行的 query 纠错处理可以具体为: 利用用户输入的源语言 query 查找纠错训练语料, 判断纠错

训练语料中是否存在与用户输入的源语言 query 相同的错误 query, 如果是, 则确定与用户输入的源语言 query 相同的错误 query 所对应的所有正确 query, 从确定的所有正确 query 中选择对应纠错概率排在前 n1 个的正确 query 构成源语言 query 集合 Q1; 否则, Q1 中仅包括用户输入的源语言 query。

其中, 纠错训练语料包括: 预先从搜索日志中收集的错误 query 和对应正确 query 构成的 query 对, 以及错误 query 被纠错为对应正确 query 的纠错概率。

优化处理单元 340 进行的 query 扩展处理可以具体为: 将用户输入的源语言 query 进行分词处理, 通过查找源语言的复述资源, 例如同义词词典, 确定分词处理后得到的各词语的同义词, 利用分词处理后得到的各词语及各词语的同义词进行组合, 取组合得到的 query 中扩展分值排在前 n2 个的 query 构成 Q2。

其中, query 的扩展分值由创建复述资源中该 query 的统计次数确定。需要特别说明的是, 复述资源不限于词, 也可以为短语, 甚至为句子, 例如基于词典注释的替换、语序变换、句子结构变换、句子拆分与合并或基于推理的复述得到的资源, 只要描述的事物相同, 表达的含义相同, 都可以认为是复述资源。

另外, 为了同时获取到源语言的搜索结果, 上述搜索处理单元 320 还可以用于获取源语言 query 对应的搜索结果。

如果对源语言 query 进行了优化处理, 则该装置可以进一步包括: 源请求选择单元 350, 用于从优化处理单元 340 进行优化处理后得到的源语言 query 集合中选择一个 query。

此时，搜索处理单元 320 在针对源语言获取搜索结果时，具体是获取源请求选择单元 350 选择的 query 对应的搜索结果。

源请求选择单元 350 采用的选择策略可以包括但不限于：对优化处理单元 340 进行优化处理后得到的源语言 query 集合中的各 query 逐一进行搜索，直至找到搜索效果满足预设要求的 query，选择该搜索效果满足预设要求的 query；或者，对优化处理单元 340 进行优化处理后得到的源语言 query 集合中的各 query 进行搜索，选择搜索效果最优的 query。

另外，结果整合单元 330 包括：合并处理模块 331、去重处理模块 332 和排序处理模块 333。

合并处理模块 331，用于将搜索处理单元 320 获取的搜索结果进行合并处理。

去重处理单元 332，用于将合并处理模块 331 合并处理后的搜索结果进行去重处理得到搜索结果集合。

排序处理模块 333，用于在上述搜索结果集合中，根据各搜索结果在所属分类中的排次以及所属分类的排序权重，对各搜索结果进行排序。

具体地，排序处理模块 333 利用搜索结果集合中各搜索结果在所属分类中的排次以及所属分类的排序权重，对各搜索结果进行打分，按照打分结果从高到低对各搜索结果进行排序。

其中，搜索结果 Rst 的打分结果 $score(Rst)$ 为：

$$score(Rst) = \frac{1}{m} \sum_{i=1}^m w_i \frac{1}{rank_i(Rst)},$$

m 为搜索结果分类的总个数， w_i 为第 i

种分类的排序权重， $rank_i(Rst)$ 为 Rst 在第 i 种分类中的排次。

上述搜索结果所属分类可以为：搜索结果对应的语言，该语言可以包括目标语言，还可以包括源语言。

排序处理模块 333 在确定第 i 种分类的排序权重时，具体执行以下操作：提取用户输入的源语言 query 的特征；将提取的特征与各语言的特征向量进行相似度计算，确定相似度超过预设的相似度阈值的语言为用户输入的源语言 query 的映射语言；对于搜索结果 Rst，如果 Rst 所属分类为映射语言，则该所属分类的排序权重 w_i 为第一设定值 a ；如果 Rst 所属分类为源语言且源语言不是该所属分类的映射语言，则该所属分类的排序权重 w_i 为第二设定值 b ；如果 Rst 所属分类既不是映射语言也不是源语言，则该所属分类的排序权重 w_i 为第三设定值 c 。其中， $a > b > c$ ，各语言的特征向量是预先对各语言的已有资源进行挖掘所训练出来的。

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。

权利要求书

1、一种跨语言搜索的方法，其特征在于，该方法包括：

A、接收用户输入的源语言搜索请求 query；

B、将所述源语言 query 翻译为 N 种目标语言 query，N 为大于 1 的整数；

C、分别获取所述 N 种目标语言 query 对应的搜索结果；

D、将步骤 C 获取的搜索结果进行整合后形成最终的搜索结果集合提供给用户；

其中在所述最终的搜索结果集合中，根据各搜索结果在所属分类中的排次以及所属分类的排序权重，对各搜索结果进行排序。

2、根据权利要求 1 所述的方法，其特征在于，在步骤 B 中，针对每一种目标语言，将所述源语言 query 对应的该种目标语言的翻译结果中，翻译分值最高的一种翻译结果作为目标语言 query；

翻译结果 e 的翻译分值由以下因素中的至少一种确定：翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

3、根据权利要求 1 所述的方法，其特征在于，所述步骤 B 具体包括：

B1、对所述源语言 query 进行优化处理，所述优化处理包括 query 纠错处理和 query 扩展处理中的任一种或组合；

B2、将优化处理后的源语言 query 翻译为 N 种目标语言 query。

4、根据权利要求 3 所述的方法，其特征在于，如果所述优化处理仅包括 query 纠错处理，则对所述用户输入的源语言 query 进行 query 纠错处理后得到包含 $n1$ 个 query 的源语言 query 集合 $Q1$ ， $n1$ 为预设的正整

数;

所述步骤 B2 具体为: 针对每一种目标语言, 分别利用所述 Q1 中的各 query 进行翻译, 确定翻译分值总和最高的翻译结果作为目标语言 query; 其中, 翻译结果的翻译分值总和为 $\sum_{i=1}^{n1} P(e|q_i)$, $P(e|q_i)$ 为所述 Q1 中 q_i 被翻译为 e 的翻译分值;

翻译结果 e 对应的翻译分值由以下因素中的至少一种确定: 翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

5、根据权利要求 3 所述的方法, 其特征在于, 如果所述优化处理仅包括 query 扩展处理, 则对所述用户输入的源语言 query 进行 query 扩展处理后得到包含 $n2$ 个 query 的源语言 query 集合 Q2, $n2$ 为预设的正整数;

所述步骤 B2 具体为: 针对每一种目标语言, 分别利用所述 Q2 中的各 query 进行翻译, 确定翻译分值总和最高的翻译结果作为目标语言 query; 其中, 翻译结果的翻译分值总和为 $\sum_{i=1}^{n2} P(e|q_i)$, $P(e|q_i)$ 为所述 Q2 中 q_i 被翻译为 e 的翻译分值;

翻译结果 e 对应的翻译分值由以下因素中的至少一种确定: 翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

6、根据权利要求 3 所述的方法, 其特征在于, 如果所述优化处理既包括 query 纠错处理又包括 query 扩展处理, 则对所述用户输入的源语言 query 进行 query 纠错处理和 query 扩展处理后得到包含 n 个 query 的

源语言 query 集合 Q , n 为预设的正整数;

所述步骤 B2 具体为: 针对每一种目标语言, 分别利用所述 Q 中的各 query 进行翻译, 确定翻译分值总和最高的翻译结果作为目标语言 query; 其中, 翻译结果的翻译分值总和为 $\sum_{i=1}^n P(e|q_i)$, $P(e|q_i)$ 为 Q 中 q_i 被翻译为 e 的翻译分值;

翻译结果 e 的翻译分值由以下因素中的至少一种确定: 翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

7、根据权利要求 6 所述的方法, 其特征在于, 对所述用户输入的源语言 query 进行 query 纠错处理后和 query 扩展处理后得到包含 n 个 query 的源语言 query 集合 Q 具体包括:

对所述用户输入的源语言 query 进行 query 纠错处理后得到包含 $n1$ 个 query 的源语言 query 集合 $Q1$, $n1$ 为预设的正整数, 将所述 $Q1$ 中的各 query 分别进行 query 扩展处理, 得到包含 n 个 query 的源语言 query 集合 Q ; 或者,

对所述用户输入的源语言 query 进行 query 扩展处理后得到包含 $n2$ 个 query 的源语言 query 集合 $Q2$, $n2$ 为预设的正整数, 将所述 $Q2$ 中的各 query 分别进行 query 纠错处理, 得到包含 n 个 query 的源语言 query 集合 Q ; 或者,

对所述用户输入的源语言 query 同时进行 query 纠错处理和 query 扩展处理后, 分别得到包含 $n1$ 个 query 的源语言 query 集合 $Q1$ 和包含 $n2$ 个 query 的源语言 query 集合 $Q2$, 将所述 $Q1$ 和 $Q2$ 取并集后, 得到包含 n 个 query 的源语言 query 集合 Q 。

8、根据权利要求 3、4 或 7 所述的方法, 其特征在于, 对所述用户

输入的源语言 query 进行 query 纠错处理具体包括:

利用所述用户输入的源语言 query 查找纠错训练语料, 判断纠错训练语料中是否存在与所述用户输入的源语言 query 相同的错误 query, 如果是, 则确定与所述用户输入的源语言 query 相同的错误 query 所对应的所有正确 query, 从确定的所有正确 query 中选择对应纠错概率排在前 n1 个的正确 query 构成源语言 query 集合 Q1, n1 为预设的正整数; 否则, 所述 Q1 中仅包括所述用户输入的源语言 query;

其中, 所述纠错训练语料包括: 预先从搜索日志中收集的错误 query 和对应正确 query 构成的 query 对, 以及错误 query 被纠错为对应正确 query 的纠错概率。

9、根据权利要求 3、5 或 7 所述的方法, 其特征在于, 对所述用户输入的源语言 query 进行 query 扩展处理具体包括:

将所述用户输入的源语言 query 进行分词处理, 通过查找源语言的复述资源确定分词处理后得到的各词语的同义词, 利用分词处理后得到的各词语及各词语的同义词进行组合, 取组合得到的 query 中扩展分值排在前 n2 个的 query 构成所述 Q2, n2 为预设的正整数;

query 的扩展分值由创建所述复述资源中该 query 的统计次数确定。

10、根据权利要求 1 至 7 任一权项所述的方法, 其特征在于, 所述步骤 C 还包括:

获取所述源语言 query 对应的搜索结果。

11、根据权利要求 4、5、6 或 7 所述的方法, 其特征在于, 所述步骤 C 还包括: 从优化处理后得到的源语言 query 集合中选择一个 query, 获取所述选择的 query 对应的搜索结果。

12、根据权利要求 11 所述的方法，其特征在于，所述从优化处理后得到的源语言 query 集合中选择一个 query 时，使用的选择策略包括：

对优化处理后得到的源语言 query 集合中的各 query 逐一进行搜索，直至找到搜索效果满足预设要求的 query，选择该搜索效果满足预设要求的 query；或者，

对优化处理后得到的源语言 query 集合中的各 query 进行搜索，选择搜索效果最优的 query。

13、根据权利要求 1 所述的方法，其特征在于，步骤 D 中所述整合包括：对步骤 C 获取的搜索结果进行合并和去重。

14、根据权利要求 1 所述的方法，其特征在于，所述根据各搜索结果在所属分类中的排次以及所属分类的排序权重，对各搜索结果进行排序具体包括：

利用各搜索结果在所属分类中的排次以及所属分类的排序权重，对各搜索结果进行打分，按照打分结果从高到低对各搜索结果进行排序；

其中，搜索结果 Rst 的打分结果 $score(Rst)$ 为： $score(Rst)=$

$$\frac{1}{m} \sum_{i=1}^m w_i \frac{1}{rank_i(Rst)},$$

m 为搜索结果分类的总个数， w_i 为第 i 种分类的排序权重， $rank_i(Rst)$ 为 Rst 在第 i 种分类中的排次。

重， $rank_i(Rst)$ 为 Rst 在第 i 种分类中的排次。

15、根据权利要求 14 所述的方法，其特征在于，搜索结果所属分类为：搜索结果对应的语言；

第 i 种分类的排序权重的确定方法具体为：

S1、提取所述用户输入的源语言 query 的特征；

S2、将步骤 S1 提取的特征与各语言的特征向量进行相似度计算，确定相似度超过预设的相似度阈值的语言为所述用户输入的源语言

query 的映射语言;

S3、对于搜索结果 Rst, 如果 Rst 所属分类为映射语言, 则该所属分类的排序权重 w_i 为第一设定值 a; 如果 Rst 所属分类为源语言且源语言不是该所属分类的映射语言, 则该所属分类的排序权重 w_i 为第二设定值 b; 如果 Rst 所属分类既不是映射语言也不是源语言, 则该所属分类的排序权重 w_i 为第三设定值 c;

其中, $a > b > c$, 各语言的特征向量是预先对各语言的已有资源进行挖掘所训练出来的。

16、一种跨语言搜索的装置, 其特征在于, 该装置包括: 用户侧交互单元、翻译处理单元、搜索处理单元和结果整合单元;

所述用户侧交互单元, 用于接收用户输入的源语言搜索请求 query, 将所述结果整合单元整合后形成的搜索结果集合提供给所述用户;

所述翻译处理单元, 用于将所述源语言 query 翻译为 N 种目标语言 query, N 为大于 1 的整数;

所述搜索处理单元, 用于分别获取所述 N 种目标语言 query 对应的搜索结果;

所述结果整合单元, 用于将所述搜索处理单元获取的搜索结果进行整合后形成最终的搜索结果集合; 其中, 在所述最终的搜索结果集合中, 根据各搜索结果在所属分类中的排次以及所属分类的排序权重, 对各搜索结果进行排序。

17、根据权利要求 16 所述的装置, 其特征在于, 所述翻译处理单元针对每一种目标语言, 将所述源语言 query 对应的该种目标语言的翻译结果中, 翻译分值最高的一种翻译结果作为目标语言 query;

翻译结果 e 的翻译分值由以下因素中的至少一种确定：翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

18、根据权利要求 16 所述的装置，其特征在于，该装置还包括：优化处理单元；

所述优化处理单元，用于对所述用户输入的源语言 query 进行优化处理后提供给所述翻译处理单元，所述优化处理包括 query 纠错处理和 query 扩展处理中的任一种或组合；

所述翻译处理单元将所述优化处理单元进行优化处理后的源语言 query 翻译为 N 种目标语言 query。

19、根据权利要求 18 所述的装置，其特征在于，如果所述优化处理仅包括 query 纠错处理，则所述优化处理单元对所述用户输入的源语言 query 进行 query 纠错处理后得到包含 $n1$ 个 query 的源语言 query 集合 $Q1$ ， $n1$ 为预设的正整数；

所述翻译处理单元针对每一种目标语言，分别利用所述 $Q1$ 中的各 query 进行翻译，确定翻译分值总和最高的翻译结果作为目标语言 query；其中，翻译结果的翻译分值总和为 $\sum_{i=1}^{n1} P(e|q_i)$ ， $P(e|q_i)$ 为 $Q1$ 中 q_i 被翻译为 e 的翻译分值；

翻译结果 e 对应的翻译分值由以下因素中的至少一种确定：翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

20、根据权利要求 18 所述的装置，其特征在于，如果所述优化处理仅包括 query 扩展处理，则所述优化处理单元对所述用户输入的源语言

query 进行 query 扩展处理后得到包含 n_2 个 query 的源语言 query 集合 Q_2 , n_2 为预设的正整数;

所述翻译处理单元针对每一种目标语言, 分别利用所述 Q_2 中的各 query 进行翻译, 确定翻译分值总和最高的翻译结果作为目标语言 query; 其中, 翻译结果的翻译分值总和为 $\sum_{i=1}^{n_2} P(e|q_i)$, $P(e|q_i)$ 为 Q_2 中 q_i 被翻译为 e 的翻译分值;

翻译结果 e 对应的翻译分值由以下因素中的至少一种确定: 翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

21、根据权利要求 18 所述的装置, 其特征在于, 如果所述优化处理既包括 query 纠错处理又包括 query 扩展处理, 则所述优化处理单元对所述用户输入的源语言 query 进行 query 纠错处理和 query 扩展处理后得到包含 n 个 query 的源语言 query 集合 Q , n 为预设的正整数;

所述翻译处理单元针对每一种目标语言, 分别利用所述 Q 中的各 query 进行翻译, 确定翻译分值总和最高的翻译结果作为目标语言 query; 其中, 翻译结果的翻译分值总和为 $\sum_{i=1}^n P(e|q_i)$, $P(e|q_i)$ 为 Q 中 q_i 被翻译为 e 的翻译分值;

翻译结果 e 的翻译分值由以下因素中的至少一种确定: 翻译所使用的翻译语料库中翻译结果 e 的统计次数以及翻译结果 e 中各词的组合概率。

22、根据权利要求 21 所述的装置, 其特征在于, 所述优化处理单元具体包括: 第一纠错模块和第一扩展模块;

所述第一纠错模块, 用于对所述用户输入的源语言 query 进行 query

纠错处理后得到包含 $n1$ 个 query 的源语言 query 集合 $Q1$, $n1$ 为预设的正整数;

所述第一扩展模块, 用于将所述 $Q1$ 中的各 query 分别进行 query 扩展处理, 得到包含 n 个 query 的源语言 query 集合 Q 。

23、根据权利要求 21 所述的装置, 其特征在于, 所述优化处理单元具体包括: 第二扩展模块和第二纠错模块;

所述第二扩展模块, 用于对所述用户输入的源语言 query 进行 query 扩展处理后得到包含 $n2$ 个 query 的源语言 query 集合 $Q2$, $n2$ 为预设的正整数;

所述第二纠错模块, 用于将所述 $Q2$ 中的各 query 分别进行 query 纠错处理, 得到包含 n 个 query 的源语言 query 集合 Q 。

24、根据权利要求 21 所述的装置, 其特征在于, 所述优化处理单元具体包括: 第三纠错模块、第三扩展模块和合并处理模块;

所述第三纠错模块, 用于对所述用户输入的源语言 query 进行 query 纠错处理, 得到包含 $n1$ 个 query 的源语言 query 集合 $Q1$;

所述第三扩展模块, 用于对所述用户输入的源语言 query 进行 query 扩展处理, 得到包含 $n2$ 个 query 的源语言 query 集合 $Q2$;

所述合并处理模块, 用于将所述 $Q1$ 和 $Q2$ 取并集后, 得到包含 n 个 query 的源语言 query 集合 Q 。

25、根据权利要求 19 或 21 所述的装置, 其特征在于, 所述优化处理单元具体利用所述用户输入的源语言 query 查找纠错训练语料, 判断纠错训练语料中是否存在与所述用户输入的源语言 query 相同的错误 query, 如果是, 则确定与所述用户输入的源语言 query 相同的错误 query

所对应的所有正确 query，从确定的所有正确 query 中选择对应纠错概率排在前 n1 个的正确 query 构成源语言 query 集合 Q1；否则，所述 Q1 中仅包括所述用户输入的源语言 query；

其中，所述纠错训练语料包括：预先从搜索日志中收集的错误 query 和对应正确 query 构成的 query 对，以及错误 query 被纠错为对应正确 query 的纠错概率。

26、根据权利要求 20 或 21 所述的装置，其特征在于，所述优化处理单元具体将所述用户输入的源语言 query 进行分词处理，通过查找源语言的复述资源确定分词处理后得到的各词语的同义词，利用分词处理后得到的各词语及各词语的同义词进行组合，取组合得到的 query 中扩展分值排在前 n2 个的 query 构成所述 Q2；

query 的扩展分值由创建所述复述资源中该 query 的统计次数确定。

27、根据权利要求 16 至 24 任一权项所述的装置，其特征在于，所述搜索处理单元，还用于获取所述源语言 query 对应的搜索结果。

28、根据权利要求 19、20、21 或 22 任一权项所述的装置，其特征在于，该装置还包括：

源请求选择单元，用于从所述优化处理单元进行优化处理后得到的源语言 query 集合中选择一个 query；

所述搜索处理单元，还用于获取所述源请求选择单元选择的 query 对应的搜索结果。

29、根据权利要求 28 所述的装置，其特征在于，所述源请求选择单元采用的选择策略包括：

对所述优化处理单元进行优化处理后得到的源语言 query 集合中的

各 query 逐一进行搜索，直至找到搜索效果满足预设要求的 query，选择该搜索效果满足预设要求的 query；或者，

对所述优化处理单元进行优化处理后得到的源语言 query 集合中的各 query 进行搜索，选择搜索效果最优的 query。

30、根据权利要求 16 所述的装置，其特征在于，所述结果整合单元包括：合并处理模块、去重处理模块和排序处理模块；

所述合并处理模块，用于将所述搜索处理单元获取的搜索结果进行合并处理；

所述去重处理单元，用于将所述合并处理模块合并处理后的搜索结果进行去重处理得到搜索结果集合；

所述排序处理模块，用于在所述搜索结果集合中，根据各搜索结果在所属分类中的排次以及所属分类的排序权重，对各搜索结果进行排序。

31、根据权利要求 30 所述的装置，其特征在于，所述排序处理模块具体利用所述搜索结果集合中各搜索结果在所属分类中的排次以及所属分类的排序权重，对各搜索结果进行打分，按照打分结果从高到低对各搜索结果进行排序；

其中，搜索结果 Rst 的打分结果 $score(Rst)$ 为： $score(Rst) =$

$$\frac{1}{m} \sum_{i=1}^m w_i \frac{1}{rank_i(Rst)},$$

m 为搜索结果分类的总个数， w_i 为第 i 种分类的排序权重， $rank_i(Rst)$ 为 Rst 在第 i 种分类中的排次。

重， $rank_i(Rst)$ 为 Rst 在第 i 种分类中的排次。

32、根据权利要求 31 所述的装置，其特征在于，搜索结果所属分类为：搜索结果对应的语言；

所述排序处理模块在确定第 i 种分类的排序权重时，具体执行以下操作：提取所述用户输入的源语言 query 的特征；将提取的特征与各语

言的特征向量进行相似度计算，确定相似度超过预设的相似度阈值的语言为所述用户输入的源语言 query 的映射语言；对于搜索结果 Rst，如果 Rst 所属分类为映射语言，则该所属分类的排序权重 w_i 为第一设定值 a；如果 Rst 所属分类为源语言且源语言不是该所属分类的映射语言，则该所属分类的排序权重 w_i 为第二设定值 b；如果 Rst 所属分类既不是映射语言也不是源语言，则该所属分类的排序权重 w_i 为第三设定值 c；

其中， $a > b > c$ ，各语言的特征向量是预先对各语言的已有资源进行挖掘所训练出来的。

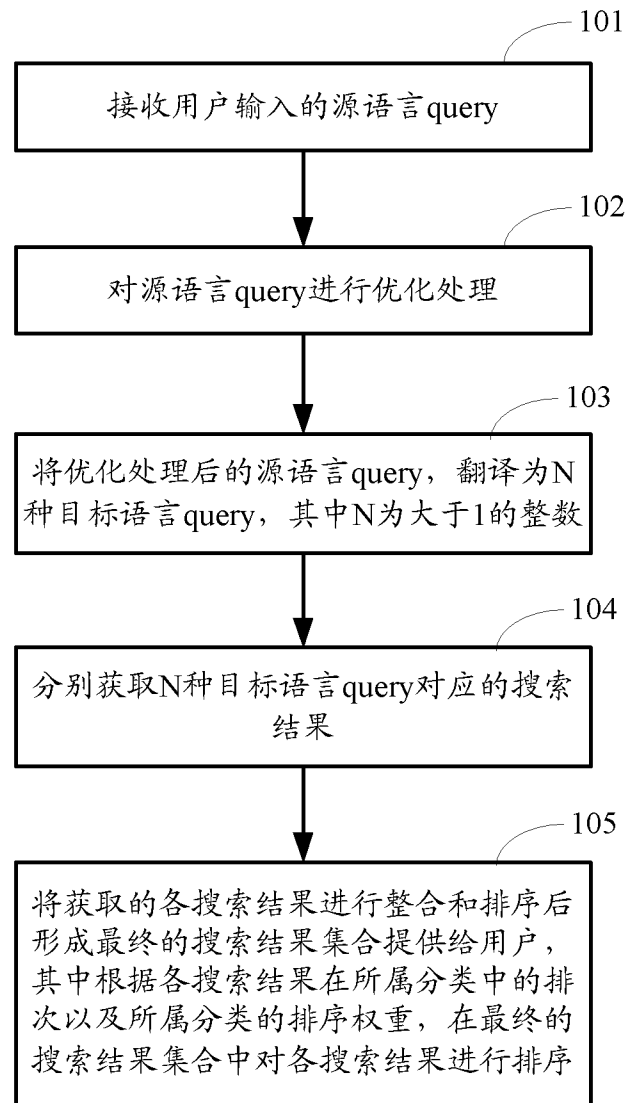
说明书附图

图 1

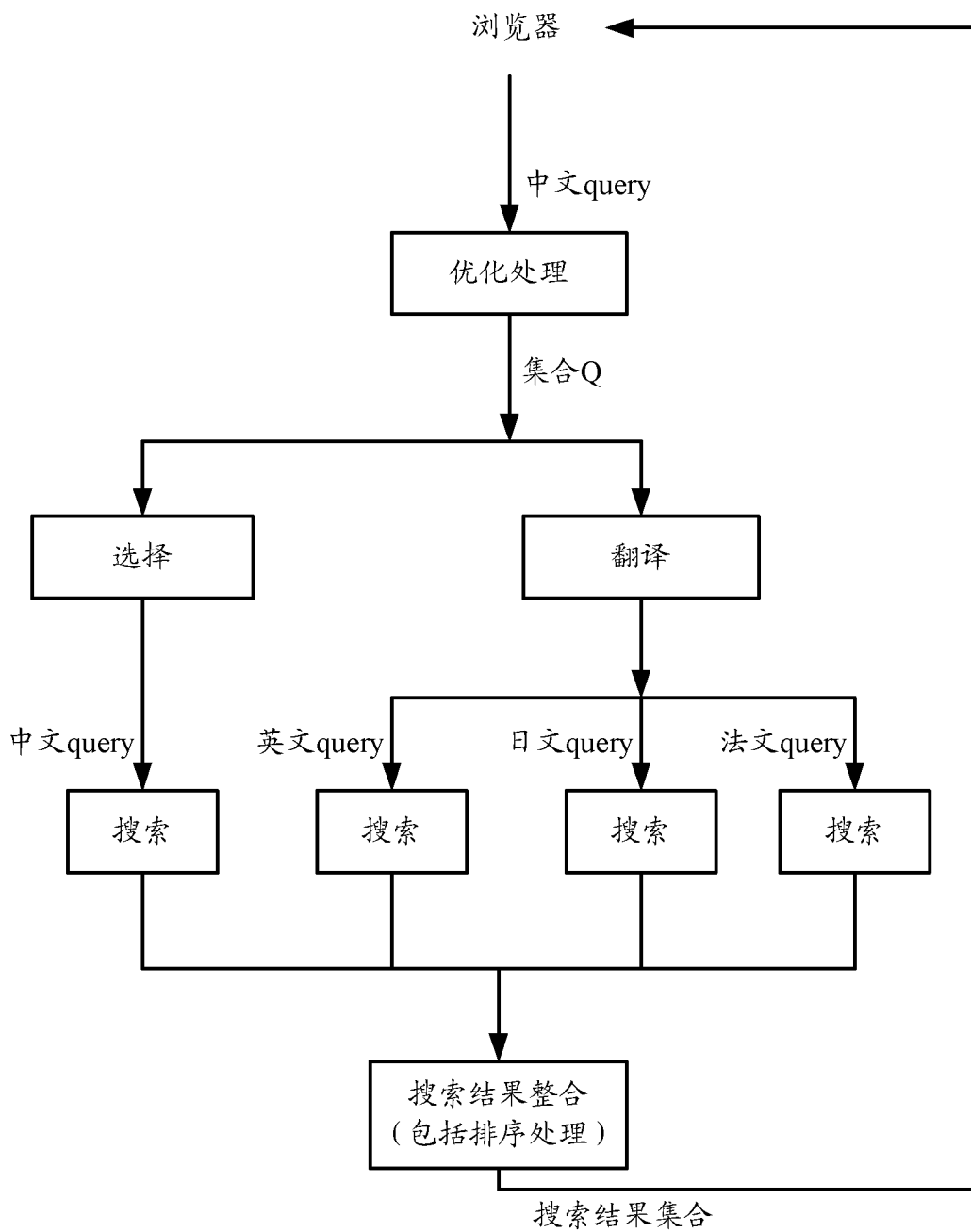


图 2

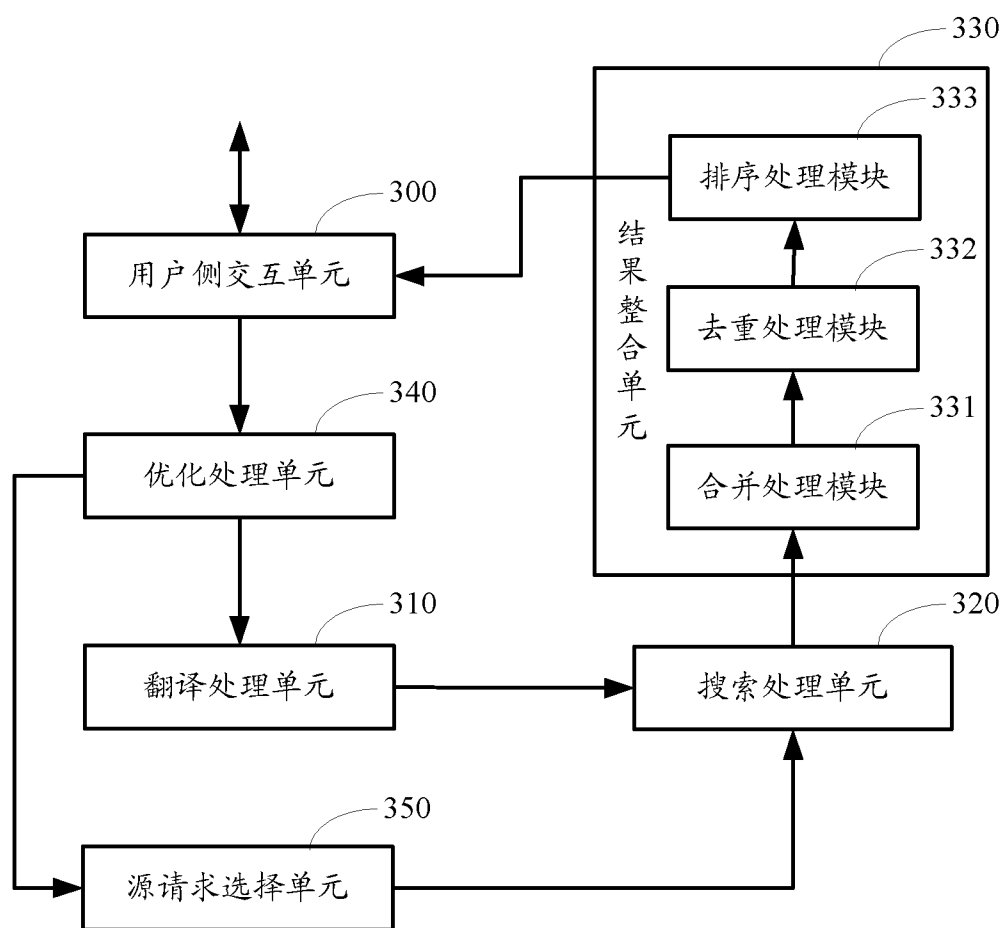


图 3

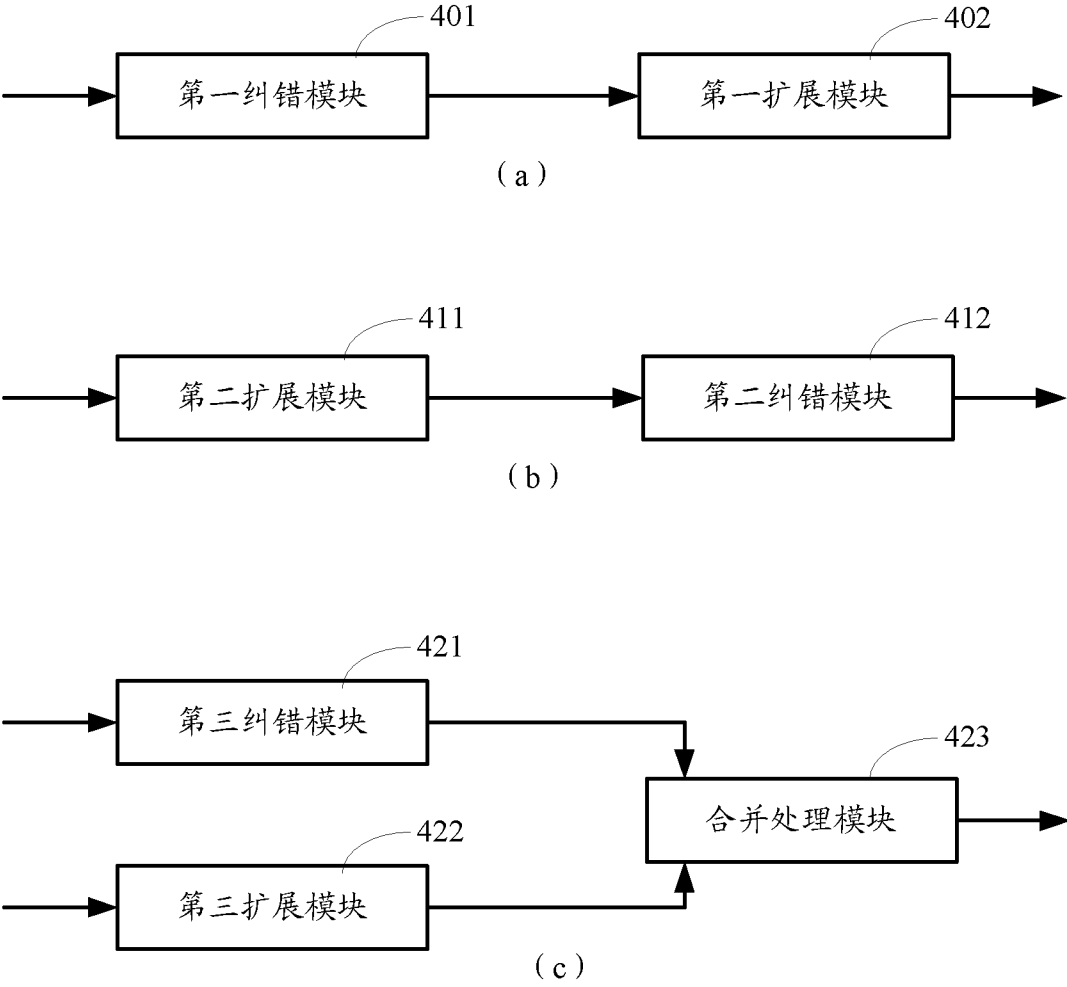


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2011/083420**A. CLASSIFICATION OF SUBJECT MATTER**

G06F 17/30 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: G06F 17/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CPRSABS, CNKI, CNTXT, DWPI, SIPOABS: gather, (multi or cross) w langu+, search, query, retriev+, translat+, integrat+, set, value, weight, frequency, times

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 101271461 A (TOSHIBA CORP.), 24 September 2008 (24.09.2008), description, paragraphs 31-80	1, 10, 13, 14, 16, 27, 30, 31
Y		2-3, 9, 17-18
A		4-8, 11-12, 15, 19-26, 28-29, 32
Y	CN 101743544 A (GOOGLE INC.), 16 June 2010 (16.06.2010), description, paragraphs 25 and 36	2-3, 9, 17-18
A	CN 101868797 A (GOOGLE INC.), 20 October 2010 (20.10.2010), the whole document	1-32

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:	“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
“A” document defining the general state of the art which is not considered to be of particular relevance	“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
“E” earlier application or patent but published on or after the international filing date	“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	“&” document member of the same patent family
“O” document referring to an oral disclosure, use, exhibition or other means	
“P” document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 22 February 2012 (22.02.2012)	Date of mailing of the international search report 15 March 2012 (15.03.2012)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer YAN, Shiyong Telephone No.: (86-10) 62411767

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2011/083420

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 101271461 A	24.09.2008	US 2008235202 A1	25.09.2008
		JP 2008234656 A	02.10.2008
		CN 101271461 B	13.07.2011
CN 101743544 A	16.06.2010	US 2008288474 A1	20.11.2008
		WO 2008144538 A1	27.11.2008
		KR 20100022467 A	02.03.2010
		EP 2165278 A1	24.03.2010
		INMUMNP 200902199 E	18.06.2010
		JP 2010527494 A	12.08.2010
CN 101868797 A	20.10.2010	WO 2009039524 A1	26.03.2009
		US 2009083243 A1	26.03.2009
		US 2009193003 A1	30.07.2009
		EP 2201484 A1	30.06.2010

国际检索报告

国际申请号

PCT/CN2011/083420

A. 主题的分类

G06F17/30(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

IPC:G06F17/-

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)):CPRSABS,CNKI,CNTEXT,DWPI,SIPOABS:多语言, 跨语言, 检索, 搜索, 查找, 查询, 翻译, 整合, 集合, 分值, 权值, 频率, 次数, (multi or cross) w langu+, search, query, retriev+, translat+, integrat+, set, value, weight, frequency, times

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN101271461A(株式会社东芝), 24.9 月 2008(24.09.2008), 说明书 31-80 段	1,10,13,14,16,27,30,31
Y		2-3,9,17-18
A		4-8,11-12,15,19-26,28-29,32
Y	CN101743544A(谷歌公司), 16.6 月 2010(16.06.2010), 说明书 25 段, 36 段	2-3,9,17-18
A	CN101868797A(谷歌公司), 20.10 月 2010(20.10.2010), 全文	1-32

☐ 其余文件在 C 栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

22.2 月 2012(22.02.2012)

国际检索报告邮寄日期

15.3 月 2012 (15.03.2012)

ISA/CN 的名称和邮寄地址:

中华人民共和国国家知识产权局

中国北京市海淀区蓟门桥西土城路 6 号 100088

传真号: (86-10)62019451

受权官员

颜世莹

电话号码: (86-10) 62411767

国际检索报告

关于同族专利的信息

国际申请号

PCT/CN2011/083420

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
CN 101271461 A	24.09.2008	US 2008235202 A1	25.09.2008
		JP 2008234656 A	02.10.2008
		CN 101271461 B	13.07.2011
CN 101743544 A	16.06.2010	US 2008288474 A1	20.11.2008
		WO 2008144538 A1	27.11.2008
		KR 20100022467 A	02.03.2010
		EP 2165278 A1	24.03.2010
		INMUMNP200902199 E	18.06.2010
CN 101868797 A	20.10.2010	JP 2010527494 A	12.08.2010
		WO 2009039524 A1	26.03.2009
		US 2009083243 A1	26.03.2009
		US 2009193003 A1	30.07.2009
		EP 2201484 A1	30.06.2010