

(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 2012/116561 A1

- (51) 国际专利分类号:
G06F 17/30 (2006.01)
- (21) 国际申请号: PCT/CN2011/083411
- (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:
201110049883.1 2011 年 3 月 2 日 (02.03.2011) 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): 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.); 中国广东省深圳市南山区高新

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)。

本国际公布:

— 包括国际检索报告 (条约第 21 条(3))。

(54) Title: WEB PAGE RANKING METHOD AND SYSTEM IN CROSS-LANGUAGE SEARCH

(54) 发明名称: 一种跨语言搜索中的网页排序方法和系统

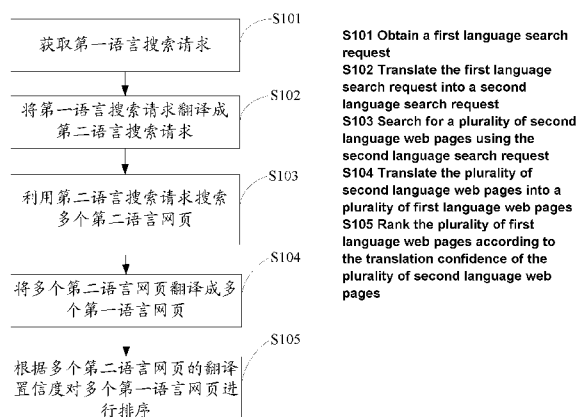


图 1 / Fig. 1

(57) Abstract: Provided are a web page ranking method and system in the cross-language search. The web page ranking method in the cross-language search includes: obtaining a first language search request; translating the first language search request into a second language search request; searching to obtain a plurality of second language web pages using the second language search request; translating the plurality of second language web pages into a plurality of first language web pages; and ranking the plurality of first language web pages according to the translation confidence of the plurality of second language web pages. In this way, the search results after translation are ranked according to the translation confidence, thus improving user experience.

[见续页]



WO 2012/116561 A1



(57) 摘要:

本发明提供了一种跨语言搜索中的网页排序方法及系统。该跨语言搜索中的网页排序方法包括：获取第一语言搜索请求；将第一语言搜索请求翻译成第二语言搜索请求；利用第二语言搜索请求搜索得到多个第二语言网页；将多个第二语言网页翻译成多个第一语言网页；根据多个第二语言网页的翻译置信度对多个第一语言网页进行排序。通过上述方式，根据翻译置信度对翻译后的搜索结果进行排序，进而提高了用户体验。

一种跨语言搜索中的网页排序方法和系统

本申请要求了申请日为2011年03月02日,申请号为201110049883.1,发明名称为“一种跨语言搜索中的网页排序方法和系统”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

本发明涉及互联网领域,特别涉及一种跨语言搜索中的网页排序方法及系统。

背景技术

随着网络搜索技术的不断发展,为克服用户的语言障碍,跨语言搜索技术应运而生。在跨语言搜索(比如用中文搜索英文网页)时,首先,输入中文的搜索请求,并将中文的搜索请求翻译成英文的搜索请求,再利用英文的搜索请求搜索英文网页。然后,将搜索到的英文网页的内容翻译成中文呈现给读者。在搜索结果呈现的过程中,一般需要进行排序。在现有的跨语言搜索技术,主要通过英文的搜索请求与英文网页的相关度来进行排序。然而,由于跨语言搜索进行了翻译过程,因此可能导致翻译质量较差的结果排在前面,导致用户体验不佳。

发明内容

有鉴于此,本发明所要解决的技术问题是提供一种跨语言搜索中的网页排序方法及系统,以提高用户体验。

本发明为解决技术问题而采用的技术方案是：提供一种跨语言搜索中的网页排序方法，所述跨语言搜索中的网页排序方法包括：

- a. 获取第一语言搜索请求；
- b. 将所述第一语言搜索请求翻译成第二语言搜索请求；
- c. 利用所述第二语言搜索请求搜索得到多个第二语言网页；
- d. 将所述多个第二语言网页翻译成多个第一语言网页；
- e. 根据所述多个第二语言网页的翻译置信度对所述多个第一语言网页进行排序。

根据本发明之一优选实施例，在所述步骤 e 中，在排序后的所述多个第一语言网页中，所述翻译置信度越高的所述第一语言网页的位置越靠前。

根据本发明之一优选实施例，所述步骤 e 包括：

- e11. 利用语言模型计算所述第二语言网页的翻译困惑度；
- e12. 根据第一语言网页所对应第二语言网页的所述翻译困惑度对所述多个第一语言网页进行排序；

其中所述语言模型是利用翻译所述第二语言网页时使用的双语语料库中的源语言语料生成的。

根据本发明之一优选实施例，在所述步骤 e11 中，通过以下公式计算所述翻译困惑度：

$$P = 2^{-\sum_{i=1}^I p(x_i) \log p(x_i)}$$

其中， P 为所述翻译困惑度， x_i 为所述第二语言网页中的第 i 个句子， $1 \leq i \leq I$ ， I 为所述第二语言网页中的句子数量， $p(x_i)$ 为通过所述语言模型计算获得的 x_i 的出现概率。

根据本发明之一优选实施例，所述语言模型为 n-gram 语言模型。

根据本发明之一优选实施例，所述步骤 e 包括：

e21. 统计所述第二语言网页在翻译过程中的调序次数；

e22. 根据第一语言网页所对应第二语言网页的所述调序次数对所述多个第一语言网页进行排序。

根据本发明之一优选实施例，所述步骤 e 包括：

e31. 分别计算各第二语言网页与利用翻译所述第二语言网页时使用的双语语料库中的源语言语料进行聚类得到的多个文档的最大相似度；

e32. 根据第一语言网页所对应第二语言网页的所述最大相似度对所述多个第一语言网页进行排序。

根据本发明之一优选实施例，在步骤 e31 中，从所述多个文档中获取多个主题，并计算每一所述文档属于每一所述主题的概率，以形成多个第一向量；计算所述第二语言网页属于每一所述主题的概率，以形成第二向量，计算所述多个第一向量与所述第二向量的相似度，并选择所述相似度中最大的作为所述最大相似度。

根据本发明之一优选实施例，在步骤 e31 中，根据如下公式计算所述最大相似度：

$$H = \max_{m=1}^M \frac{\sum_{n=1}^N p(t_n | d_s) \times p(t_n | d_m)}{\sqrt{\sum_{n=1}^N (p(t_n | d_s))^2} \sqrt{\sum_{n=1}^N (p(t_n | d_m))^2}}$$

其中， H 为所述最大相似度， t_n 为第 n 个主题， $1 \leq n \leq N$ ， N 为所述主题的数量， d_m 为第 m 个文档， $1 \leq m \leq M$ ， M 为所述文档的数量， $p(t_n | d_m)$ 为 d_m 属于 t_n 的概率， d_s 为所述第二语言网页， $p(t_n | d_s)$ 为 d_s 属于 t_n 的概率。

根据本发明之一优选实施例，所述步骤 e 包括：

e41. 统计所述第二语言网页在翻译过程中包含的未登录词的个数；

e42. 根据第一语言网页所对应第二语言网页的所述未登录词的个数对所述多个第一语言网页进行排序。

根据本发明之一优选实施例，所述步骤 e 包括：

e51. 计算所述第二语言网页所包含句子在翻译过程中的平均翻译评分；

e52. 根据第一语言网页所对应第二语言网页的所述平均翻译评分对所述多个第一语言网页进行排序。

根据本发明之一优选实施例，所述步骤 c 包括：

e61. 统计所述第二语言网页在翻译过程中的翻译规则使用次数；

e62. 根据第一语言网页所对应第二语言网页的所述翻译规则使用次数对所述多个第一语言网页进行排序。

本发明为解决技术问题而采用的技术方案是：提供一种跨语言搜索中的网页排序系统，所述跨语言搜索中的网页排序系统包括：

搜索请求获取单元，用于获取第一语言搜索请求；

第一翻译单元，用于将所述第一语言搜索请求翻译成第二语言搜索请求；

搜索单元，用于利用所述第二语言搜索请求搜索得到多个第二语言网页；

第二翻译单元，用于将所述多个第二语言网页翻译成多个第一语言网页；

排序单元，用于根据所述多个第二语言网页的翻译置信度对所述多

个第一语言网页进行排序。

根据本发明之一优选实施例，在经所述排序单元排序后的所述多个第一语言网页中，所述翻译置信度越高的所述第一语言网页的位置越靠前。

根据本发明之一优选实施例，所述排序单元包括：

源语言语料获取模块，用于获取翻译所述第二语言网页时使用的双语语料库中的源语言语料；

语言模型生成模块，用于利用所述源语言语料生成语言模型；

困惑度计算模块，用于利用所述语言模型计算所述第二语言网页的翻译困惑度；

第一排序模块，用于根据第一语言网页所对应第二语言网页的所述翻译困惑度对所述多个第一语言网页进行排序。

根据本发明之一优选实施例，所述困惑度计算模块通过以下公式计算所述翻译困惑度：

$$P = 2^{-\sum_{i=1}^I p(x_i) \log p(x_i)}$$

其中， P 为所述翻译困惑度， x_i 为所述第二语言网页中的第 i 个句子， $1 \leq i \leq I$ ， I 为所述第二语言网页中的句子数量， $p(x_i)$ 为通过所述语言模型计算获得的 x_i 的出现概率。

根据本发明之一优选实施例，所述语言模型为 n-gram 语言模型。

根据本发明之一优选实施例，所述排序单元包括：

调序次数统计模块，用于统计所述第二语言网页在翻译过程中的调序次数；

第二排序模块，用于根据第一语言网页所对应第二语言网页的所述

调序次数对所述多个第一语言网页进行排序。

根据本发明之一优选实施例，所述排序单元包括：

源语言语料获取模块，用于获取翻译所述第二语言网页时使用的双语语料库中的源语言语料；

聚类模块，用于将所述源语言语料聚类成多个文档；

相似度计算模块，用于计算所述第二语言网页与所述多个文档的最大相似度；

第三排序模块，用于根据第一语言网页所对应第二语言网页的所述最大相似度对所述多个第一语言网页进行排序。

根据本发明之一优选实施例，所述聚类模块从所述多个文档中获取多个主题，并计算每一所述文档属于每一所述主题的概率，以形成多个第一向量，所述相似度计算模块计算所述第二语言网页属于每一所述主题的概率，以形成第二向量，计算所述多个第一向量与所述第二向量的相似度，并选择所述相似度中最大的作为所述最大相似度。

根据本发明之一优选实施例，所述相似度计算模块根据如下公式计算所述最大相似度：

$$H = \max_{m=1}^M \frac{\sum_{n=1}^N p(t_n | d_s) \times p(t_n | d_m)}{\sqrt{\sum_{n=1}^N (p(t_n | d_s))^2} \sqrt{\sum_{n=1}^N (p(t_n | d_m))^2}}$$

其中， H 为所述最大相似度， t_n 为第 n 个主题， $1 \leq n \leq N$ ， N 为所述主题的数量， d_m 为第 m 个文档， $1 \leq m \leq M$ ， M 为所述文档的数量， $p(t_n | d_m)$ 为 d_m 属于 t_n 的概率， d_s 为所述第二语言网页， $p(t_n | d_s)$ 为 d_s 属于 t_n 的概率。

根据本发明之一优选实施例，所述排序单元包括：

未登录词统计模块，用于统计所述第二语言网页在翻译过程中包含的未登录词的个数；

第四排序模块，用于根据第一语言网页所对应第二语言网页的所述未登录词的个数对所述多个第一语言网页进行排序。

根据本发明之一优选实施例，所述排序单元包括：

翻译评分计算模块，用于计算所述第二语言网页所包含句子在翻译过程中的平均翻译评分；

第五排序模块，用于根据第一语言网页所对应第二语言网页的所述平均翻译评分对所述多个第一语言网页进行排序。

根据本发明之一优选实施例，所述排序单元包括：

规则使用次数统计模块，用于统计所述第二语言网页在翻译过程中的翻译规则使用次数；

第六排序模块，用于根据第一语言网页所对应第二语言网页的所述翻译规则使用次数对所述多个第一语言网页进行排序。

由以上技术方案可以看出，本发明提供的跨语言搜索中的网页排序方法及系统根据翻译置信度对翻译后的搜索结果进行排序，进而提高了用户体验。

附图说明

图 1 是本发明实施例的跨语言搜索中的网页排序方法的流程示意图；

图 2 是图 1 所示的跨语言搜索中的网页排序方法的排序过程的第一实施方式的流程示意图；

图 3 是图 1 所示的跨语言搜索中的网页排序方法的排序过程的第二

实施方式的流程图示意图；

图 4 是图 1 所示的跨语言搜索中的网页排序方法的排序过程的第三实施方式的流程图示意图；

图 5 是图 1 所示的跨语言搜索中的网页排序方法的排序过程的第四实施方式的流程图示意图；

图 6 是图 1 所示的跨语言搜索中的网页排序方法的排序过程的第五实施方式的流程图示意图；

图 7 是图 1 所示的跨语言搜索中的网页排序方法的排序过程的第六实施方式的流程图示意图；

图 8 是本发明实施例的跨语言搜索中的网页排序系统的示意框图；

图 9 是图 8 所示的跨语言搜索中的网页排序系统的排序单元的第一实施方式的示意框图；

图 10 是图 8 所示的跨语言搜索中的网页排序系统的排序单元的第二实施方式的示意框图；

图 11 是图 8 所示的跨语言搜索中的网页排序系统的排序单元的第三实施方式的示意框图；

图 12 是图 8 所示的跨语言搜索中的网页排序系统的排序单元的第四实施方式的示意框图；

图 13 是图 8 所示的跨语言搜索中的网页排序系统的排序单元的第五实施方式的示意框图；

图 14 是图 8 所示的跨语言搜索中的网页排序系统的排序单元的第六实施方式的示意框图。

具体实施方式

下面结合附图和实施例对本发明进行详细说明。

请参阅图 1，图 1 是本发明实施例的跨语言搜索中的网页排序方法的流程示意图。在本实施例中，该跨语言搜索中的网页排序方法主要包括以下几个步骤：

在步骤 S101 中，获取第一语言搜索请求。在本步骤中，用户可通过在浏览器的搜索框内输入想要搜索的第一语言搜索请求（Query），例如中文的搜索请求，并点击搜索按钮。该第一语言搜索请求经互联网传输至搜索引擎，并被搜索引擎所获取。

在步骤 S102 中，将第一语言搜索请求翻译成第二语言搜索请求。在本步骤中，可通过本领域公知的各种机器翻译手段将第一语言搜索请求翻译成第二语言搜索请求，例如，在利用中文搜索英文网页时，将中文的搜索请求翻译成英文的搜索请求。具体的机器翻译手段可包括基于单词、基于短语或句法的统计机器翻译等等。

在步骤 S103 中，利用第二语言搜索请求搜索多个第二语言网页。在本步骤中，通过搜索引擎搜索与第二语言搜索请求相关的多个第二语言网页，例如英文网页。

在步骤 S104 中，将多个第二语言网页翻译成多个第一语言网页。在本步骤中，可通过上文提到的各种机器翻译手段将第二语言网页中的网页内容翻译成第一语言，例如中文，进而实现跨语言搜索。

在步骤 S105 中，根据多个第二语言网页的翻译置信度对多个第一语言网页进行排序。在本步骤中，在排序后的多个第一语言网页中，翻译置信度越高的第一语言网页的位置越靠前，以将翻译质量好的网页结果

优先提供给用户，进而提高用户体验。在下文中将详细描述获得第二语言网页的翻译置信度的多种实施方式，本领域技术人员完全可以想到将本领域公知的其他翻译置信度获取方法应用到步骤 S105。

请参阅图 2，图 2 是图 1 所示的跨语言搜索中的网页排序方法的排序过程的第一实施方式的流程示意图。本实施方式主要包括以下几个步骤：

在步骤 S201 中，获取翻译第二语言网页时使用的双语语料库中的源语言语料。在机器翻译过程中，一般都利用双语语料库来训练翻译模型。该双语语料库包括多个双语例句对，每一双语例句对包括源语言例句以及与源语言例句对应的目标语言例句。在第二语言网页的翻译过程中，源语言为第二语言，目标语言为第一语言。双语语料库在机器翻译领域中被普遍使用，并可通过各种方式获得，在此不再赘述。

在步骤 S202 中，利用源语言语料生成语言模型，例如 n-gram 语言模型。

在步骤 S203 中，利用语言模型计算第二语言网页的翻译困惑度。具体来说，对于由 L 个词 $w_1, w_2, \dots, w_L$ 组成一个句子，通过语言模型可以计算出该句子的出现概率：

$$p(x_i) = p(w_1, w_2, \dots, w_L) = \prod_{l=1}^L p(w_l | w_{l-n}, \dots, w_{l-1})$$

其中， $p(w_l | w_{l-n}, \dots, w_{l-1})$ 表示词 w_l 与前面的 n 个词 $w_{l-n}, \dots, w_{l-1}$ 搭配的出现概率， n 为一正整数。例如，在 2-gram 语言模型中， $n = 2$ ，在 3-gram 语言模型中 $n = 3$ 。

对于包含有 I 个句子的第二语言网页来说，第二语言网页的翻译困惑度可通过如下公式计算：

$$P = 2^{-\sum_{i=1}^I p(x_i) \log p(x_i)}$$

其中, P 为第二语言网页的翻译困惑度, x_i 为第二语言网页中的第 i 个句子, $1 \leq i \leq I$, I 为第二语言网页中的句子数量, $p(x_i)$ 为通过上述语言模型计算获得的句子 x_i 的出现概率。在翻译过程中, 如果翻译困惑度越高, 表示翻译复杂度越高, 其翻译置信度越低。

在步骤 S204 中, 根据翻译困惑度对多个第一语言网页进行排序。其中, 在排序后的多个第一语言网页中, 翻译困惑度越高的第一语言网页的位置越靠后。

请参阅图 3, 图 3 是图 1 所示的跨语言搜索中的网页排序方法的排序过程的第二实施方式的流程示意图。本实施方式主要包括以下几个步骤:

在步骤 S301 中, 统计第二语言网页在翻译过程中的调序次数。在翻译过程中, 需要对源语言句子中的词或短语的翻译顺序进行调整, 这种调整即为调序。调序次数越多, 表示翻译复杂度越高, 其翻译置信度就越低。

在步骤 S302 中, 根据调序次数对多个第一语言网页进行排序。其中, 在排序后的多个第一语言网页中, 调序次数越多的第一语言网页的位置越靠后。

请参阅图 4, 图 4 是图 1 所示的跨语言搜索中的网页排序方法的排序过程的第三实施方式的流程示意图。本实施方式主要包括以下几个步骤:

在步骤 S401 中, 获取翻译第二语言网页时使用的双语语料库中的源语言语料。

在步骤 S402 中, 将源语言语料聚类成多个文档。具体来说, 利用聚类算法对源语言语料中的句子进行聚类, 然后把每个类的句子集合到一个文档, 进而形成多个文档。随后, 利用概率潜在语义分析 (Probabilistic Latent Semantic Analysis, PLSA) 或其他算法从该多个文档中获取多个主题, 并计算每一文档属于每一主题的概率, 以形成多个第一向量:

$$Vec(d_m) = (p(t_1 | d_m), p(t_2 | d_m), \dots, p(t_n | d_m), \dots, p(t_N | d_m)),$$

其中, t_n 为第 n 个主题, $1 \leq n \leq N$, N 为主题的数量, d_m 为第 m 个文档, $1 \leq m \leq M$, M 为文档的数量, $p(t_n | d_m)$ 为文档 d_m 属于主题 t_n 的概率。

在步骤 S403 中, 计算第二语言网页与多个文档的最大相似度。具体来说, 计算第二语言网页属于每一主题的概率, 以形成第二向量:

$$Vec(d_s) = (p(t_1 | d_s), p(t_2 | d_s), \dots, p(t_n | d_s), \dots, p(t_N | d_s))$$

其中, d_s 为第二语言网页, $p(t_n | d_s)$ 为第二语言网页 d_s 属于主题 t_n 的概率。

随后, 计算多个第一向量与第二向量的相似度, 并选择相似度中最大的作为最大相似度。具体相似度计算公式可以是:

$$H = \max_{m=1}^M \frac{\sum_{n=1}^N p(t_n | d_s) \times p(t_n | d_m)}{\sqrt{\sum_{n=1}^N (p(t_n | d_s))^2} \sqrt{\sum_{n=1}^N (p(t_n | d_m))^2}}$$

其中, H 为最大相似度。最大相似度越高, 代表翻译质量越高, 表示翻译置信度就越高。

在步骤 S404 中, 根据最大相似度 H 对多个第一语言网页进行排序。其中, 在排序后的多个第一语言网页中, 最大相似度越高的第一语言网页的位置越靠前。

请参阅图 5, 图 5 是图 1 所示的跨语言搜索中的网页排序方法的排

序过程的第四实施方式的流程示意图。本实施方式主要包括以下几个步骤：

在步骤 S501 中，统计第二语言网页在翻译过程中包含的未登录词的个数。未登录词是指没有被收录在源语言语料中的词，包括各类专有名词（人名、地名、机构名等）、缩写词、新增词汇等等。在机器翻译过程中，未登录词越多，代表翻译质量越差，其翻译置信度就越低。

在步骤 S502 中，根据未登录词的个数对多个第一语言网页进行排序。其中，在排序后的多个第一语言网页中，未登录词的个数越多的第一语言网页的位置越靠后。

请参阅图 6，图 6 是图 1 所示的跨语言搜索中的网页排序方法的排序过程的第五实施方式的流程示意图。本实施方式主要包括以下几个步骤：

在步骤 S601 中，计算第二语言网页在翻译过程中的平均翻译评分。具体来说，根据如下公式计算第二语言网页的平均翻译评分：

$$A = \sum_{k=1}^K score_k / K$$

其中， A 为第二语言网页的平均翻译评分， $score_k$ 为第二语言网页中的第 k 个句子的翻译评分， $1 \leq k \leq K$ ， K 为第二语言网页中的句子数量。在本步骤中，可通过本领域公知的翻译评价方法来确定每个句子的翻译评分，例如归一化的句子翻译概率等自动评价方法。平均翻译评分越高，代表翻译质量越高，表示翻译置信度就越高。

在步骤 S602 中，根据平均翻译评分对多个第一语言网页进行排序。其中，在排序后的多个第一语言网页中，平均翻译评分越高的第一语言网页的位置越靠前。

请参阅图 7，图 7 是图 1 所示的跨语言搜索中的网页排序方法的排序过程的第六实施方式的流程示意图。本实施方式主要包括以下几个步骤：

在步骤 S701 中，统计第二语言网页在翻译过程中的规则使用次数。在机器翻译领域中往往会制定一定的翻译规则，例如针对特定短语的翻译规则。在机器翻译过程中，使用规则的次数越多，代表翻译质量越差，其翻译置信度就越低。

在步骤 S702 中，根据规则使用次数对多个第一语言网页进行排序。其中，在排序后的多个第一语言网页中，规则使用次数越多的第一语言网页的位置越靠后。

上述第一至第四实施方式是从第二语言网页的源语言端获取表示翻译置信度的特征，第五至第六实施例是从第二语言网页的翻译模型或翻译结果获取表示翻译置信度的特征。当然，本领域技术人员完全可以通过其他方式获取表示翻译置信度的其他特征。

进一步，本领域技术人员阅读上述内容后完全可以想到将上文描述的表示翻译置信度的各种特征进行结合，例如使用回归学习（regression learning）方法将包括上述多个特征的特征向量映射成一个实数，进而形成一个综合上述特征的翻译置信度。上述过程可使用公知的工具实现，例如，SVM-light 工具。

此外，在获得翻译置信度后，还可以把翻译置信度作为一个特征与本领域公知的其他排序方法进行结合，例如 learning to rank 或 PageRank 方法。

请参阅图 8，图 8 是本发明实施例的跨语言搜索中的网页排序系统

的示意框图。在本实施例中，该跨语言搜索中的网页排序系统主要包括搜索请求获取单元 801、第一翻译单元 802、搜索单元 803、第二翻译单元 804 以及排序单元 805。

搜索请求获取单元 801 用于获取第一语言搜索请求。用户可通过在浏览器的搜索框内输入想要搜索的第一语言搜索请求 (Query)，例如中文的搜索请求，并点击搜索按钮。该第一语言搜索请求经互联网传输至搜索请求获取单元 801，并被搜索请求获取单元 801 所获取。

第一翻译单元 802 用于将第一语言搜索请求翻译成第二语言搜索请求。第一翻译单元 802 可通过本领域公知的各种机器翻译手段将第一语言搜索请求翻译成第二语言搜索请求，例如，在利用中文搜索英文网页时，将中文的搜索请求翻译成英文的搜索请求。具体的机器翻译手段可包括基于单词、基于短语或句法的统计机器翻译等等。

搜索单元 803 用于利用第二语言搜索请求搜索多个第二语言网页。搜索单元 803 通过本领域公知的各种搜索引擎技术搜索与第二语言搜索请求相关的多个第二语言网页，例如英文网页。

第二翻译单元 804 用于将多个第二语言网页翻译成多个第一语言网页。第二翻译单元 804 可通过上文提到的各种机器翻译手段将第二语言网页中的网页内容翻译成第一语言，例如中文，进而实现跨语言搜索。在本实施方式中，第一翻译单元 802 和第二翻译单元 804 可由同一翻译单元或不同的翻译单元实现。

排序单元 805 用于根据多个第二语言网页的翻译置信度对多个第一语言网页进行排序。在经排序单元 805 排序后的多个第一语言网页中，翻译置信度越高的第一语言网页的位置越靠前，以将翻译质量好的网页

结果优先提供给用户，进而提高用户体验。在下文中将详细描述获得第二语言网页的翻译置信度的多种实施方式，本领域技术人员完全可以想到将本领域公知的其他翻译置信度获取方法应用到排序单元 805。

请参阅图 9，图 9 是图 8 所示的跨语言搜索中的网页排序系统的排序单元 805 的第一实施方式的示意框图。本实施方式的排序单元 805 主要包括源语言语料获取模块 901、语言模型生成模块 902、困惑度计算模块 903 以及排序模块 904。

源语言语料获取模块 901 用于获取翻译第二语言网页时使用的双语语料库中的源语言语料。在机器翻译过程中，一般都利用双语语料库来训练翻译模型。该双语语料库包括多个双语例句对，每一双语例句对包括源语言例句以及与源语言例句对应的目标语言例句。在第二语言网页的翻译过程中，源语言为第二语言，目标语言为第一语言。双语语料库在机器翻译领域中被普遍使用，并可通过各种方式获得，在此不再赘述。

语言模型生成模块 902 用于利用源语言语料生成语言模型，例如 n-gram 语言模型。

困惑度计算模块 903 用于利用语言模型计算第二语言网页的翻译困惑度。具体来说，对于由 L 个词 $w_1, w_2, \dots, w_L$ 组成一个句子 x_i ，通过语言模型可以计算出该句子的出现概率：

$$p(x_i) = p(w_1, w_2, \dots, w_L) = \prod_{l=1}^L p(w_l | w_{l-n}, \dots, w_{l-1})$$

其中， $p(w_l | w_{l-n}, \dots, w_{l-1})$ 表示词 w_l 与前面的 n 个词 $w_{l-n}, \dots, w_{l-1}$ 搭配的出现概率， n 为一正整数。例如，在 2-gram 语言模型中， $n = 2$ ，在 3-gram 语言模型中 $n = 3$ 。

对于包含有 I 个句子的第二语言网页来说，第二语言网页的翻译困

惑度可通过如下公式计算:

$$P = 2^{-\sum_{i=1}^I p(x_i) \log p(x_i)}$$

其中, P 为第二语言网页的翻译困惑度, x_i 为第二语言网页中的第 i 个句子, $1 \leq i \leq I$, I 为第二语言网页中的句子数量, $p(x_i)$ 为通过上述语言模型计算获得的句子 x_i 的出现概率。在翻译过程中, 如果翻译困惑度越高, 表示翻译复杂度越高, 其翻译置信度越低。

排序模块 904 用于根据翻译困惑度对多个第一语言网页进行排序。其中, 在经排序模块 904 排序后的多个第一语言网页中, 翻译困惑度越高的第一语言网页的位置越靠后。

请参阅图 10, 图 10 是图 8 所示的跨语言搜索中的网页排序系统的排序单元 805 的第二实施方式的示意框图。本实施方式的排序单元 805 主要包括调序次数统计模块 1001 以及排序模块 1002。

调序次数统计模块 1001 用于统计第二语言网页在翻译过程中的调序次数。在翻译过程中, 需要对源语言句子中的词或短语的翻译顺序进行调整, 这种调整即为调序。调序次数越多, 表示翻译复杂度越高, 其翻译置信度就越低。

排序模块 1002 用于根据调序次数对多个第一语言网页进行排序。其中, 在经排序模块 1002 排序后的多个第一语言网页中, 调序次数越多的第一语言网页的位置越靠后。

请参阅图 11, 图 11 是图 8 所示的跨语言搜索中的网页排序系统的排序单元 805 的第三实施方式的示意框图。本实施方式的排序单元 805 主要包括源语言语料获取模块 1101、聚类模块 1102、相似度计算模块 1103 以及排序模块 1104。

源语言语料获取模块 1101 用于获取翻译第二语言网页时使用的双语语料库中的源语言语料。

聚类模块 1102 用于将源语言语料聚类成多个文档。具体来说，聚类模块 1102 利用聚类算法对源语言语料中的句子进行聚类，然后把每个类的句子集合到一个文档，进而形成多个文档。随后，聚类模块 1102 利用概率潜在语义分析（Probabilistic Latent Semantic Analysis, PLSA）或其他算法从该多个文档中获取多个主题，并计算每一文档属于每一主题的概率，以形成多个第一向量：

$$Vec(d_m) = (p(t_1 | d_m), p(t_2 | d_m), \dots, p(t_n | d_m), \dots, p(t_N | d_m)),$$

其中， t_n 为第 n 个主题， $1 \leq n \leq N$ ， N 为主题的数量， d_m 为第 m 个文档， $1 \leq m \leq M$ ， M 为文档的数量， $p(t_n | d_m)$ 为文档 d_m 属于主题 t_n 的概率。

相似度计算模块 1103 计算第二语言网页与多个文档的最大相似度。具体来说，相似度计算模块 1103 计算第二语言网页属于每一主题的概率，以形成第二向量：

$$Vec(d_s) = (p(t_1 | d_s), p(t_2 | d_s), \dots, p(t_n | d_s), \dots, p(t_N | d_s))$$

其中， d_s 为第二语言网页， $p(t_n | d_s)$ 为第二语言网页 d_s 属于主题 t_n 的概率。

随后，相似度计算模块 1103 计算多个第一向量与第二向量的相似度，并选择相似度中最大的作为最大相似度。具体相似度计算公式可以是：

$$H = \max_{m=1}^M \frac{\sum_{n=1}^N p(t_n | d_s) \times p(t_n | d_m)}{\sqrt{\sum_{n=1}^N (p(t_n | d_s))^2} \sqrt{\sum_{n=1}^N (p(t_n | d_m))^2}}$$

其中， H 为最大相似度。最大相似度越高，代表翻译质量越高，表示翻译置信度就越高。

排序模块 1104 用于根据最大相似度 H 对多个第一语言网页进行排序。其中，在经排序模块 1104 排序后的多个第一语言网页中，最大相似度越高的第一语言网页的位置越靠前。

请参阅图 12，图 12 是图 8 所示的跨语言搜索中的网页排序系统的排序单元 805 的第四实施方式的示意框图。本实施方式的排序单元 805 主要包括未登录词统计模块 1201 以及排序模块 1202。

未登录词统计模块 1201 用于统计第二语言网页在翻译过程中包含的未登录词的个数。未登录词是指没有被收录在源语言语料中的词，包括各类专有名词（人名、地名、机构名等）、缩写词、新增词汇等等。在机器翻译过程中，未登录词越多，代表翻译质量越差，其翻译置信度就越低。

排序模块 1202 用于根据未登录词的个数对多个第一语言网页进行排序。其中，在经排序模块 1202 排序后的多个第一语言网页中，未登录词的个数越多的第一语言网页的位置越靠后。

请参阅图 13，图 13 是图 8 所示的跨语言搜索中的网页排序系统的排序单元 805 的第五实施方式的示意框图。本实施方式的排序单元 805 主要包括翻译评分计算模块 1301 以及排序模块 1302。

翻译评分计算模块 1301 用于计算第二语言网页在翻译过程中的平均翻译评分。具体来说，翻译评分计算模块 1301 根据如下公式计算第二语言网页的平均翻译评分：

$$A = \sum_{k=1}^K score_k / K$$

其中， A 为第二语言网页的平均翻译评分， $score_k$ 为第二语言网页中的第 k 个句子的翻译评分， $1 \leq k \leq K$ ， K 为第二语言网页中的句子数量。

翻译评分计算模块 1301 可通过本领域公知的翻译评价方法来确定每个句子的翻译评分，例如归一化的句子翻译概率等自动评价方法。平均翻译评分越高，代表翻译质量越高，表示翻译置信度就越高。

排序模块 1302 用于根据平均翻译评分对多个第一语言网页进行排序。其中，在经排序模块 1302 排序后的多个第一语言网页中，平均翻译评分越高的第一语言网页的位置越靠前。

请参阅图 14，图 14 是图 8 所示的跨语言搜索中的网页排序系统的排序单元 805 的第六实施方式的示意框图。本实施方式的排序单元 805 主要包括规则使用次数统计模块 1401 以及排序模块 1402。

规则使用次数统计模块 1401 用于统计第二语言网页在翻译过程中的规则使用次数。在机器翻译领域中往往会制定一定的翻译规则，例如针对特定短语的翻译规则。在机器翻译过程中，使用规则的次数越多，代表翻译质量越差，其翻译置信度就越低。

排序模块 1402 用于根据规则使用次数对多个第一语言网页进行排序。其中，在经排序模块 1402 排序后的多个第一语言网页中，规则使用次数越多的第一语言网页的位置越靠后。

上述第一至第四实施方式是从第二语言网页的源语言端获取表示翻译置信度的特征，第五至第六实施例是从第二语言网页的翻译模型或翻译结果获取表示翻译置信度的特征。当然，本领域技术人员完全可以通过其他方式获取表示翻译置信度的其他特征。

进一步，本领域技术人员阅读上述内容后完全可以想到将上文描述的表示翻译置信度的各种特征进行结合，例如使用回归学习（regression learning）方法将包括上述多个特征的特征向量映射成一个实数，进而形

成一个综合上述特征的翻译置信度。上述过程可使用公知的工具实现，例如，SVM-light 工具。

此外，在获得翻译置信度后，还可以把翻译置信度作为一个特征与本领域公知的其他排序方法进行结合，例如 learning to rank 或 PageRank 方法。

由以上技术方案可以看出，本发明提供的跨语言搜索中的网页排序方法及系统根据翻译置信度对翻译后的搜索结果进行排序，进而提高了用户体验。

在上述实施例中，仅对本发明进行了示范性描述，但是本领域技术人员在阅读本专利申请后可以在不脱离本发明的精神和范围的情况下对本发明进行各种修改。

权利要求书

1、一种跨语言搜索中的网页排序方法，其特征在于，所述跨语言搜索中的网页排序方法包括：

- a. 获取第一语言搜索请求；
- b. 将所述第一语言搜索请求翻译成第二语言搜索请求；
- c. 利用所述第二语言搜索请求搜索得到多个第二语言网页；
- d. 将所述多个第二语言网页翻译成多个第一语言网页；
- e. 根据所述多个第二语言网页的翻译置信度对所述多个第一语言网页进行排序。

2、如权利要求 1 所述的跨语言搜索中的网页排序方法，其特征在于，在所述步骤 e 中，在排序后的所述多个第一语言网页中，所述翻译置信度越高的所述第一语言网页的位置越靠前。

3、如权利要求 1 所述的跨语言搜索中的网页排序方法，其特征在于，所述步骤 e 包括：

- e11. 利用语言模型计算所述第二语言网页的翻译困惑度；
- e12. 根据第一语言网页所对应第二语言网页的所述翻译困惑度对所述多个第一语言网页进行排序；

其中所述语言模型是利用翻译所述第二语言网页时使用的双语语料库中的源语言语料生成的。

4、如权利要求 3 所述的跨语言搜索中的网页排序方法，其特征在于，在所述步骤 e11 中，通过以下公式计算所述翻译困惑度：

$$P = 2^{-\sum_{i=1}^I p(x_i) \log p(x_i)}$$

其中， P 为所述翻译困惑度， x_i 为所述第二语言网页中的第 i 个句子，

$1 \leq i \leq I$, I 为所述第二语言网页中的句子数量, $p(x_i)$ 为通过所述语言模型计算获得的 x_i 的出现概率。

5、如权利要求 4 所述的跨语言搜索中的网页排序方法,其特征在于,所述语言模型为 n-gram 语言模型。

6、如权利要求 1 所述的跨语言搜索中的网页排序方法,其特征在于,所述步骤 e 包括:

e21. 统计所述第二语言网页在翻译过程中的调序次数;

e22. 根据第一语言网页所对应第二语言网页的所述调序次数对所述多个第一语言网页进行排序。

7、如权利要求 1 所述的跨语言搜索方法,其特征在于,所述步骤 e 包括:

e31. 分别计算各第二语言网页与利用翻译所述第二语言网页时使用的双语语料库中的源语言语料进行聚类得到的多个文档的最大相似度;

e32. 根据第一语言网页所对应第二语言网页的所述最大相似度对所述多个第一语言网页进行排序。

8、如权利要求 7 所述的跨语言搜索中的网页排序方法,其特征在于,在步骤 e31 中,从所述多个文档中获取多个主题,并计算每一所述文档属于每一所述主题的概率,以形成多个第一向量;计算所述第二语言网页属于每一所述主题的概率,以形成第二向量,计算所述多个第一向量与所述第二向量的相似度,并选择所述相似度中最大的作为所述最大相似度。

9、如权利要求 8 所述的跨语言搜索中的网页排序方法,其特征在于,在步骤 e31 中,根据如下公式计算所述最大相似度:

$$H = \max_{m=1}^M \frac{\sum_{n=1}^N p(t_n | d_s) \times p(t_n | d_m)}{\sqrt{\sum_{n=1}^N (p(t_n | d_s))^2} \sqrt{\sum_{n=1}^N (p(t_n | d_m))^2}}$$

其中, H 为所述最大相似度, t_n 为第 n 个主题, $1 \leq n \leq N$, N 为所述主题的数量, d_m 为第 m 个文档, $1 \leq m \leq M$, M 为所述文档的数量, $p(t_n | d_m)$ 为 d_m 属于 t_n 的概率, d_s 为所述第二语言网页, $p(t_n | d_s)$ 为 d_s 属于 t_n 的概率。

10、如权利要求 1 所述的跨语言搜索中的网页排序方法, 其特征在于, 所述步骤 e 包括:

e41. 统计所述第二语言网页在翻译过程中包含的未登录词的个数;

e42. 根据第一语言网页所对应第二语言网页的所述未登录词的个数对所述多个第一语言网页进行排序。

11、如权利要求 1 所述的跨语言搜索方法, 其特征在于, 所述步骤 e 包括:

e51. 计算所述第二语言网页所包含句子在翻译过程中的平均翻译评分;

e52. 根据第一语言网页所对应第二语言网页的所述平均翻译评分对所述多个第一语言网页进行排序。

12、如权利要求 1 所述的跨语言搜索中的网页排序方法, 其特征在于, 所述步骤 e 包括:

e61. 统计所述第二语言网页在翻译过程中的翻译规则使用次数;

e62. 根据第一语言网页所对应第二语言网页的所述翻译规则使用次数对所述多个第一语言网页进行排序。

13、一种跨语言搜索中的网页排序系统, 其特征在于, 所述跨语言

搜索中的网页排序系统包括:

搜索请求获取单元, 用于获取第一语言搜索请求;

第一翻译单元, 用于将所述第一语言搜索请求翻译成第二语言搜索请求;

搜索单元, 用于利用所述第二语言搜索请求搜索得到多个第二语言网页;

第二翻译单元, 用于将所述多个第二语言网页翻译成多个第一语言网页;

排序单元, 用于根据所述多个第二语言网页的翻译置信度对所述多个第一语言网页进行排序。

14、如权利要求 13 所述的跨语言搜索中的网页排序系统, 其特征在于, 在经所述排序单元排序后的所述多个第一语言网页中, 所述翻译置信度越高的所述第一语言网页的位置越靠前。

15、如权利要求 13 所述的跨语言搜索中的网页排序系统, 其特征在于, 所述排序单元包括:

源语言语料获取模块, 用于获取翻译所述第二语言网页时使用的双语语料库中的源语言语料;

语言模型生成模块, 用于利用所述源语言语料生成语言模型;

困惑度计算模块, 用于利用所述语言模型计算所述第二语言网页的翻译困惑度;

第一排序模块, 用于根据第一语言网页所对应第二语言网页的所述翻译困惑度对所述多个第一语言网页进行排序。

16、如权利要求 15 所述的跨语言搜索中的网页排序系统, 其特征在

于, 所述困惑度计算模块通过以下公式计算所述翻译困惑度:

$$P = 2^{-\sum_{i=1}^I p(x_i) \log p(x_i)}$$

其中, P 为所述翻译困惑度, x_i 为所述第二语言网页中的第 i 个句子, $1 \leq i \leq I$, I 为所述第二语言网页中的句子数量, $p(x_i)$ 为通过所述语言模型计算获得的 x_i 的出现概率。

17、如权利要求 16 所述的跨语言搜索中的网页排序系统, 其特征在于, 所述语言模型为 n-gram 语言模型。

18、如权利要求 13 所述的跨语言搜索中的网页排序系统, 其特征在于, 所述排序单元包括:

调序次数统计模块, 用于统计所述第二语言网页在翻译过程中的调序次数;

第二排序模块, 用于根据第一语言网页所对应第二语言网页的所述调序次数对所述多个第一语言网页进行排序。

19、如权利要求 13 所述的跨语言搜索中的网页排序系统, 其特征在于, 所述排序单元包括:

源语言语料获取模块, 用于获取翻译所述第二语言网页时使用的双语语料库中的源语言语料;

聚类模块, 用于将所述源语言语料聚类成多个文档;

相似度计算模块, 用于计算所述第二语言网页与所述多个文档的最大相似度;

第三排序模块, 用于根据第一语言网页所对应第二语言网页的所述最大相似度对所述多个第一语言网页进行排序。

20、如权利要求 19 所述的跨语言搜索中的网页排序系统, 其特征在

于, 所述聚类模块从所述多个文档中获取多个主题, 并计算每一所述文档属于每一所述主题的概率, 以形成多个第一向量, 所述相似度计算模块计算所述第二语言网页属于每一所述主题的概率, 以形成第二向量, 计算所述多个第一向量与所述第二向量的相似度, 并选择所述相似度中最大的作为所述最大相似度。

21、如权利要求 20 所述的跨语言搜索中的网页排序系统, 其特征在于, 所述相似度计算模块根据如下公式计算所述最大相似度:

$$H = \max_{m=1}^M \frac{\sum_{n=1}^N p(t_n | d_s) \times p(t_n | d_m)}{\sqrt{\sum_{n=1}^N (p(t_n | d_s))^2} \sqrt{\sum_{n=1}^N (p(t_n | d_m))^2}}$$

其中, H 为所述最大相似度, t_n 为第 n 个主题, $1 \leq n \leq N$, N 为所述主题的数量, d_m 为第 m 个文档, $1 \leq m \leq M$, M 为所述文档的数量, $p(t_n | d_m)$ 为 d_m 属于 t_n 的概率, d_s 为所述第二语言网页, $p(t_n | d_s)$ 为 d_s 属于 t_n 的概率。

22、如权利要求 13 所述的跨语言搜索中的网页排序系统, 其特征在于, 所述排序单元包括:

未登录词统计模块, 用于统计所述第二语言网页在翻译过程中包含的未登录词的个数;

第四排序模块, 用于根据第一语言网页所对应第二语言网页的所述未登录词的个数对所述多个第一语言网页进行排序。

23、如权利要求 13 所述的跨语言搜索中的网页排序系统, 其特征在于, 所述排序单元包括:

翻译评分计算模块, 用于计算所述第二语言网页所包含句子在翻译过程中的平均翻译评分;

第五排序模块，用于根据第一语言网页所对应第二语言网页的所述平均翻译评分对所述多个第一语言网页进行排序。

24、如权利要求 13 所述的跨语言搜索中的网页排序系统，其特征在于，所述排序单元包括：

规则使用次数统计模块，用于统计所述第二语言网页在翻译过程中的翻译规则使用次数；

第六排序模块，用于根据第一语言网页所对应第二语言网页的所述翻译规则使用次数对所述多个第一语言网页进行排序。

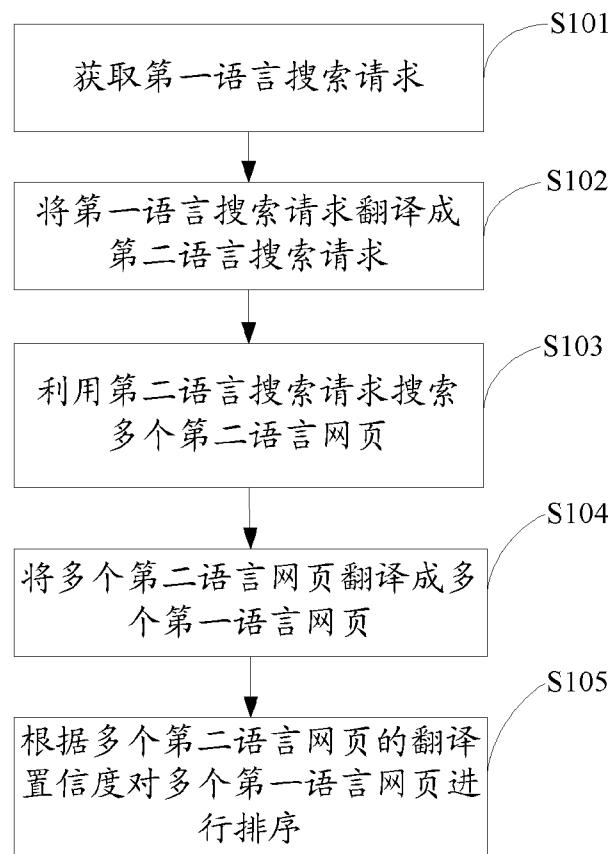


图 1

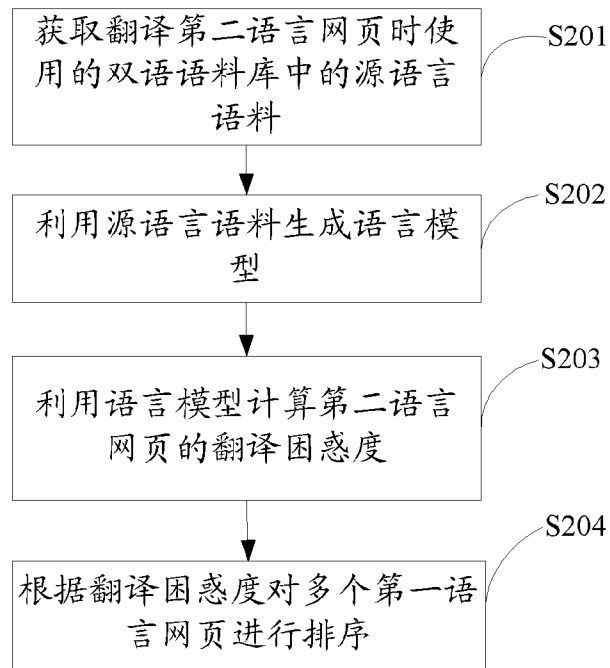


图 2

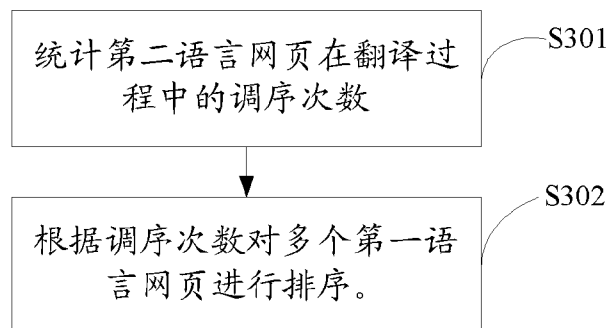


图 3

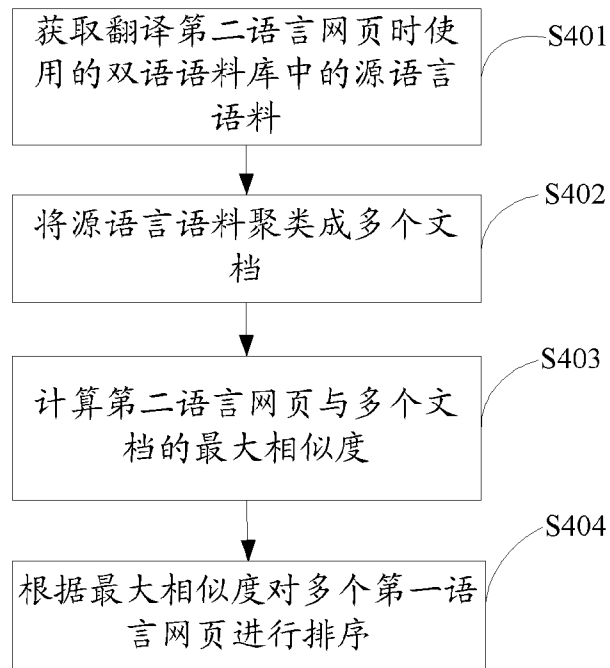


图 4

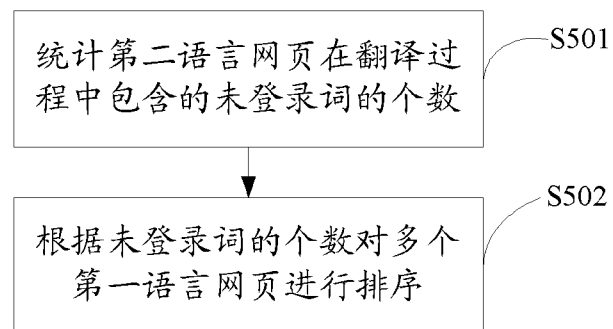


图 5

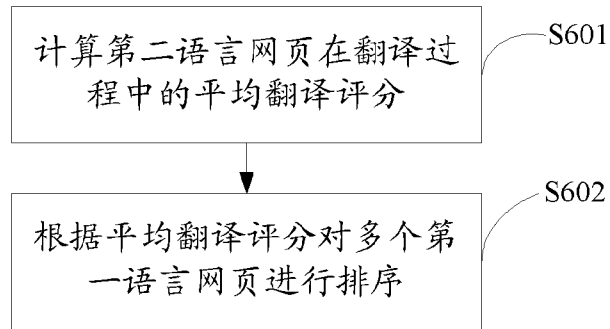


图 6

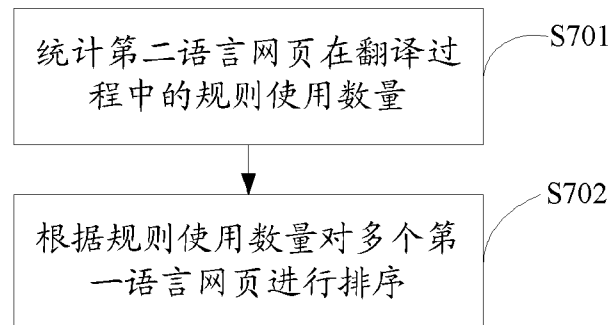


图 7

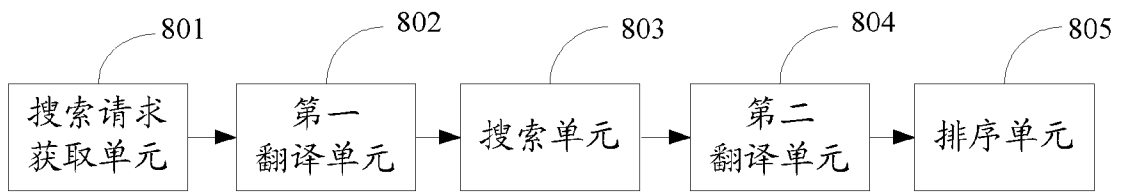


图 8

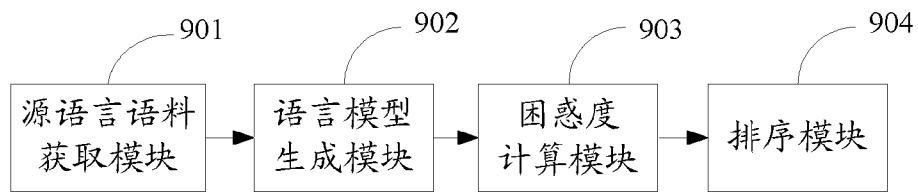


图 9

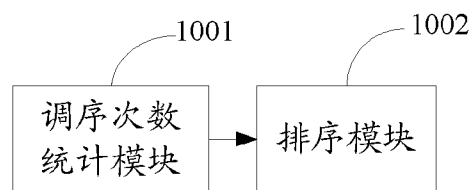


图 10

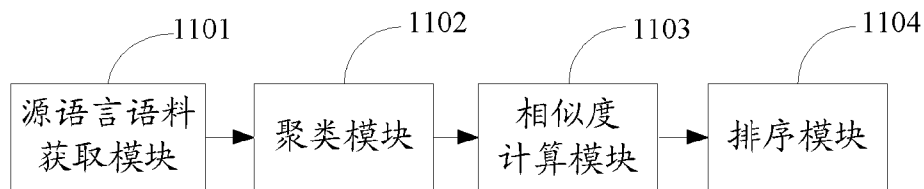


图 11

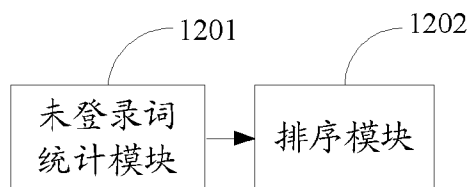


图 12

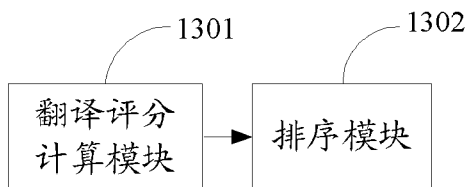


图 13

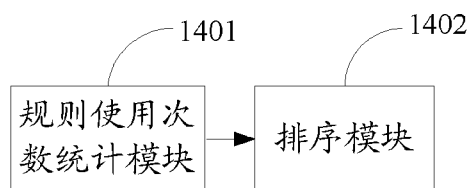


图 14

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2011/083411

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)

CNKI, CNPAT, VEN: (multi or cross) w langu+, search, query, retriev+, translat+, order, precedence, confidence, likelihood, probability,
page, engine, browser

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 101868797 A (GOOGLE INC) 20 Oct. 2010(20.10.2010) description paragraphs 18 – 49, figures 1A-10	1, 2, 6, 10-14, 18, 22-24
Y		3, 7, 15, 19
A		4-5, 8-9, 16-17, 20-21
X	CN 101743544 A (GOOGLE INC) 16 Jun. 2010(16.06.2010) description paragraphs 17-49	1, 2, 6, 10-14, 18, 22-24
Y		3, 7, 15, 19
A		4-5, 8-9, 16-17, 20-21
A	CN 101271461 A (TOSHIBA KK) 24 Sep. 2008(24.09.2008) the whole document	1-24
A	CN 101763402 A (HARBIN INST TECHNOLOGY) 30 Jun. 2010(30.06.2010) the whole document	1-24

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“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)

“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

“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

“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

“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

“&” document member of the same patent family

Date of the actual completion of the international search
01 Mar. 2012(01.03.2012)

Date of mailing of the international search report
15 Mar. 2012(15.03.2012)

Name and mailing address of the ISA
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
LIU, Li
Telephone No. (86-10)62411687

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2011/083411

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
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
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 101271461 A	24.09.2008	US 2008235202 A1	25.09.2008
		JP 2008234656 A	02.10.2008
		CN 101271461 B	13.07.2011
CN 101763402 A	30.06.2010	NONE	

A. 主题的分类		
G06F 17/30 (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: G06F 17/-		
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)) CNKI, CNPAT, VEN: 多语言, 跨语言, 多语种, 检索, 搜索, 查找, 查询, 翻译, 排序, 网页, 概率, 置信, 最大似然, 浏览器, 引擎, (multi or cross) w langu+, search, query, retriev+, translat+, order, precedence, confidence, likelihood, probability, page, engine, browser		
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	CN 101868797 A (谷歌公司) 20.10 月 2010(20.10.2010) 说明书第 18 段-第 49 段, 图 1A-10	1, 2, 6, 10-14, 18, 22-24
Y		3, 7, 15, 19
A		4-5, 8-9, 16-17, 20-21
X	CN 101743544 A (谷歌公司) 16.6 月 2010(16.06.2010) 说明书第 17 段-第 49 段	1, 2, 6, 10-14, 18, 22-24
Y		3, 7, 15, 19
A		4-5, 8-9, 16-17, 20-21
A	CN 101271461 A (株式会社东芝) 24.9 月 2008(24.09.2008) 全文	1-24
A	CN 101763402 A (哈尔滨工业大学) 30.6 月 2010(30.06.2010) 全文	1-24
☐ 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 01.3 月 2012(01.03.2012)		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 刘力 电话号码: (86-10) 62411687

国际检索报告

关于同族专利的信息

国际申请号

PCT/CN2011/083411

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 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
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 101271461 A	24.09.2008	US 2008235202 A1	25.09.2008
		JP 2008234656 A	02.10.2008
		CN 101271461 B	13.07.2011
CN 101763402 A	30.06.2010	无	