

(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
2017 年 12 月 21 日 (21.12.2017)



(10) 国际公布号
WO 2017/215245 A1

(51) 国际专利分类号:
G06F 17/30 (2006.01)

(21) 国际申请号: PCT/CN2016/113202

(22) 国际申请日: 2016 年 12 月 29 日 (29.12.2016)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201610436857.7 2016年6月17日 (17.06.2016) CN

(71) 申请人: 广州视源电子科技有限公司 (GUANGZHOU SHIYUAN ELECTRONICS CO., LTD.) [CN/CN]; 中国广东省广州市黄埔区云埔四路6号, Guangdong 510530 (CN)。

(72) 发明人: 李贤 (LI, Xian); 中国广东省广州市黄埔区云埔四路6号, Guangdong 510530 (CN)。

(74) 代理人: 广州三环专利代理有限公司 (SCIHEAD PATENT AGENT CO., LTD.); 中国广东省广州

市越秀区先烈中路80号汇华商贸大厦1508室, Guangdong 510070 (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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, 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, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, 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,

(54) Title: METHOD AND DEVICE FOR SEARCHING RESUMES

(54) 发明名称: 简历搜索方法和装置

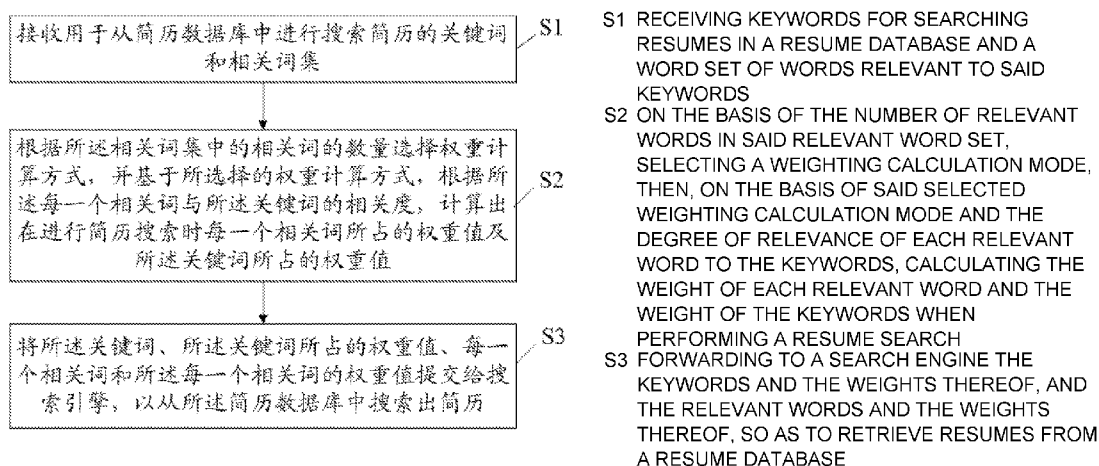


图 1

(57) Abstract: A method and device for searching resumes, the method comprising: receiving keywords for searching resumes in a resume database and a word set of words relevant to said keywords (S1); on the basis of the number of relevant words in said relevant word set, selecting a weighting calculation mode, then, on the basis of said selected weighting calculation mode and the degree of relevance of each relevant word to the keywords, calculating the weight of each relevant word and the weight of the keywords when performing a resume search (S2); forwarding to a search engine the keywords and the weights thereof, and the relevant words and the weights thereof, so as to retrieve resumes from a resume database (S3). Balancing the weights of keywords and relevant words in the process of searching resumes enhances the efficiency of resume searches.

RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, 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))。

(57) 摘要: 一种简历搜索方法及装置, 该方法包括: 接收用于从简历数据库中进行简历搜索的关键词和所述关键词的相关词集(S1); 据所述相关词集中的相关词的数量选择权重计算方式, 并基于所选择的权重计算方式, 根据所述每一个相关词与所述关键词的相关度, 计算出在进行简历搜索时每一个相关词所占的权重值及所述关键词所占的权重值(S2); 将所述关键词、所述关键词所占的权重值、每一个相关词和所述每一个相关词的权重值提交给搜索引擎, 以从所述简历数据库中搜索出简历(S3)。平衡关键词和相关词在简历搜索过程中所占的权重, 提高简历搜索的效率。

简历搜索方法和装置

技术领域

本发明涉及计算机信息检索领域，尤其涉及一种简历搜索方法和装置。

背景技术

通常在对简历进行搜索匹配过程中，一般是通过关键词+相关词检索进行检索的，但是发明人在实施过程中发现该方案存在以下问题：在进行简历搜索时搜索引擎会将关键词和相关词取相同的权重值进行搜索，那么命中的频次决定了搜索出来的简历的排序，则搜索结果极有可能偏向某个频次较高的相关词，另一方面，当相关词的数量大大多于关键词时，则在检索过程中关键词可能会被相关词淹没，难以突出关键词相对于相关词在搜索出简历过程中作为主要检索关键词的作用。

发明内容

本发明实施例提出一种简历搜索方法，平衡关键词和相关词在简历搜索过程中所占的权重，提高简历搜索的效率。

本发明实施例提出一种简历搜索方法，包括：

接收用于从简历数据库中进行简历搜索的关键词和相关词集；其中，所述相关词集记载有多个相关词，以及每一个相关词与所述关键词的相关度；

根据所述相关词集中的相关词的数量选择权重计算方式，并基于所选择的权重计算方式，根据所述每一个相关词与所述关键词的相关度，计算出在进行简历搜索时每一个相关词所占的权重值及所述关键词所占的权重值；

将所述关键词、所述关键词所占的权重值、每一个相关词和所述每一个相关词的权重值提交给搜索引擎，以从所述简历数据库中搜索出简历。

进一步地，当所述相关词集中的相关词的数量处于上限词量和下限词量之间时，选择第一方式作为权重计算方式，则基于所选择的权重计算方式，根据所述每一个相关词与所述关键词的相关度，计算出在进行简历搜索时每一个相关词所占的权重值及所述关键词所占的权重值，具体为：

根据第一权重公式 $W_i = \frac{r_i}{r_{\max}}$ ，计算出所述相关词集中第 i 个相关词在进行简历搜索时所

占的权重值 W_i ；其中， $r_{\max}$ 为在所述相关词集中所记载的所有相关词的相关度的最大值；

根据第二权重公式 $W_A = m \times \text{sum}W$ ，计算出所述关键词在进行简历搜索时所占的权重值 W_A ；其中， A 为所述关键词， m 为第一权重系数， $\text{sum}W$ 为所述相关词集中所有相关词的权重值的总和。

进一步地，当所述相关词集中的相关词的数量不处于所述上限词量和所述下限词量之间时，选择第二方式作为权重计算方式，则基于所选择的权重计算方式，根据所述每一个相关词与所述关键词的相关度，计算出在进行简历搜索时每一个相关词所占的权重值及所述关键词所占的权重值，具体为：

根据基准分值公式 $S_i = \frac{r_i - r_{\min} + 1}{r_{\max} - r_{\min} + 1}$ ，计算出所述相关词集中第 i 个相关词的基准分值 S_i ；

其中， r_i 为所述相关词集中的第 i 个相关词的相关度； $r_{\min}$ 为在所述相关词集中所记载的所有相关词的相关度的最小值； $r_{\max}$ 为在所述相关词集中所记载的所有相关词的相关度的最大值；

对所述所有相关词的基准分值进行求和，获得基准分值总值 sum ；

根据第三权重公式 $W_A = 5 + \log_{1.5}(\text{sum} + 1)$ ，计算出所述关键词在进行简历搜索时所占的权重值 W_A ；其中， A 为所述关键词；

判断所述第 i 个相关词的基准分值是否大于 $\frac{W_A}{n}$ ；若是，则根据第四权重公式

$W_i = \frac{W_A}{n \times \log_{10}(k + 10)}$ ，计算出所述第 i 个相关词在进行简历搜索时所占的权重值 W_i ；若否，

则根据第五权重公式 $W_i = \frac{S_i}{\log_{10}(k + 10)}$ ，计算出所述第 i 个相关词在进行简历搜索时所占的权重值 W_i ；其中， n 为第二权重系数； k 为所述相关词集中的相关词的数量。

再进一步地，在所述判断所述第 i 个相关词的基准分值是否大于 $\frac{W_A}{n}$ 之前，还包括：

根据均值公式 $\bar{S} = \frac{\bar{r}}{r_{\max} - r_{\min} + 1}$ ，计算基准分值均值 $\bar{S}$ ；其中， $\bar{r}$ 为所述所有相关词的相关

度的平均值；

对于所述相关词集中的第 i 个相关词，判断所述第 i 个相关词的基准分值 S_i 是否大于所述基准分值均值 $\bar{S}$ ；

若是，通过更新公式 $S_i = 10 \times (S_i - \bar{S})$ ，更新所述第 i 个相关词的基准分值 S_i 。

更进一步地, 所述将所述关键词、所述关键词所占的权重值、每一个相关词和所述每一个相关词所占的权重值提交给搜索引擎, 以从所述简历数据库中搜索出简历, 具体为:

根据搜索引擎的权重关联格式, 将所述关键词所占的权重值和所述关键词相关联作为第一组合, 以及将所述每一个相关词所占的权重值与其对应的相关词相关联作为第二组合;

将所述第一组合和所述第二组合提交给搜索引擎, 以供所述搜索引擎从所述简历数据库中搜索出简历, 并根据所述搜索引擎内置的排序算法显示搜索出的简历; 其中, 所述简历中包含的所述关键词和所述每一个相关词以高亮形式显示。

相应地, 本发明实施例还提供一种简历搜索装置, 包括:

接收模块, 用于接收用于从简历数据库中进行简历搜索的关键词和相关词集; 其中, 所述相关词集记载有多个相关词, 以及每一个相关词与所述关键词的相关度;

权重计算模块, 用于根据所述相关词集中的相关词的数量选择权重计算方式, 并基于所选择的权重计算方式, 根据所述每一个相关词与所述关键词的相关度, 计算出在进行简历搜索时每一个相关词所占的权重值及所述关键词所占的权重值;

搜索模块, 用于将所述关键词、所述关键词所占的权重值、每一个相关词和所述每一个相关词的权重值提交给搜索引擎, 以从所述简历数据库中搜索出简历。

进一步地, 所述权重计算模块包括:

第一计算单元, 用于当所述相关词集中的相关词的数量处于上限词量和下限词量之间时, 选择第一方式作为权重计算方式, 根据第一权重公式 $W_i = \frac{r_i}{r_{\max}}$, 计算出所述相关词集中第 i 个相关词在进行简历搜索时所占的权重值 W_i ; 其中, $r_{\max}$ 为在所述相关词集中所记载的所有相关词的相关度的最大值;

第二计算单元, 用于根据第二权重公式 $W_A = m \times \text{sum}W$, 计算出所述关键词在进行简历搜索时所占的权重值 W_A ; 其中, A 为所述关键词, m 为第一权重系数, $\text{sum}W$ 为所述相关词集中所有相关词的权重值的总和。

进一步地, 所述权重计算模块还包括:

基准分值计算单元, 用于当所述相关词集中的相关词的数量不处于所述上限词量和所述下限词量之间时, 选择第二方式作为权重计算方式, 根据基准分值公式 $S_i = \frac{r_i - r_{\min} + 1}{r_{\max} - r_{\min} + 1}$, 计算出所述相关词集中第 i 个相关词的基准分值 S_i ; 其中, r_i 为所述相关词集中的第 i 个相关词的相关度; $r_{\min}$ 为在所述相关词集中所记载的所有相关词的相关度的最小值; $r_{\max}$ 为在所述相

关键词集中所记载的所有相关词的相关度的最大值;

求和计算单元, 用于对所述所有相关词的基准分值进行求和, 获得基准分值总值 sum ;

第三计算单元, 用于根据第三权重公式 $W_A = 5 + \log_{1.5}(sum + 1)$, 计算出所述关键词在进行简历搜索时所占的权重值 W_A ; 其中, A 为所述关键词;

第四计算单元, 用于判断所述第 i 个相关词的基准分值是否大于 W_A/n ; 若是, 则根据第四权重公式 $W_i = \frac{W_A}{n \times \log_{10}(k + 10)}$, 计算出所述第 i 个相关词在进行简历搜索时所占的权重值

W_i ; 若否, 则根据第五权重公式 $W_i = \frac{S_i}{\log_{10}(k + 10)}$, 计算出所述第 i 个相关词在进行简历搜索时所占的权重值 W_i ; 其中, n 为第二权重系数; k 为所述相关词集中的相关词的数量。

再进一步地, 所述权重计算模块还包括:

均值计算单元, 用于在所述判断所述第 i 个相关词的基准分值是否大于 W_A/n 之前, 根据均值公式 $\bar{S} = \frac{\bar{r}}{r_{\max} - r_{\min} + 1}$, 计算基准分值均值 $\bar{S}$; 其中, $\bar{r}$ 为所述所有相关词的相关度的平均值;

判断单元, 用于对于所述相关词集中的第 i 个相关词, 判断所述第 i 个相关词的基准分值 S_i 是否大于所述基准分值均值 $\bar{S}$;

更新单元, 用于当判断所述第 i 个相关词的基准分值 S_i 是大于所述平均分 $\bar{S}$ 时, 通过更新公式 $S_i = 10 \times (S_i - \bar{S})$, 更新所述第 i 个相关词的基准分值 S_i 。

更进一步地, 所述搜索模块具体包括:

权重关联单元, 用于根据搜索引擎的权重关联格式, 将所述关键词所占的权重值和所述关键词相关联作为第一组合, 以及将所述每一个相关词所占的权重值与其对应的相关词相关联作为第二组合;

搜索显示单元, 用于将所述第一组合和所述第二组合提交给搜索引擎, 以供所述搜索引擎从所述简历数据库中搜索出简历, 并根据所述搜索引擎内置的排序算法显示搜索出的简历; 其中, 所述简历中包含的所述关键词和所述每一个相关词以高亮形式显示。

实施本发明实施例, 具有如下有益效果:

本发明实施例提供的简历搜索方法和装置, 能根据相关词的数量选取不同的权重计算算法, 来进行计算关键词和相关词在搜索过程所占的权重, 充分考虑相关词的数量对关键词在检索过程的影响, 平衡关键词和相关词在简历搜索过程中所占的权重, 避免在权重计算过程

中出现当相关词过多时导致关键词权重过大或过小，当相关词过少时关键词的权重过大或过小的权重失衡情况，使得简历检索的过程更为合理，减少再次进行简历搜索的发生，提高简历搜索的效率。

附图说明

图 1 是本发明提供的简历搜索方法的一个实施例的流程示意图；

图 2 是本发明提供的简历搜索装置的一个实施例的结构示意图；

图 3 是本发明提供的简历搜索装置的权重计算模块的一个实施例的结构示意图；

图 4 是本发明提供的简历搜索装置的搜索模块的一个实施例的结构示意图。

具体实施方式

下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

参见图 1，是本发明提供的简历搜索方法的一个实施例的流程示意图，该方法包括以下步骤：

S1，接收用于从简历数据库中进行简历搜索的关键词和相关词集；其中，所述相关词集记载有多个相关词，以及每一个相关词与所述关键词的相关度；

S2，根据所述相关词集中的相关词的数量选择权重计算方式，并基于所选择的权重计算方式，根据所述每一个相关词与所述关键词的相关度，计算出在进行简历搜索时每一个相关词所占的权重值及所述关键词所占的权重值；

S3，将所述关键词、所述关键词所占的权重值、每一个相关词和所述每一个相关词的权重值提交给搜索引擎，以从所述简历数据库中搜索出简历。

进一步地，上述步骤 S2 中，当所述相关词集中的相关词的数量处于上限词量和下限词量之间时，选择第一方式作为权重计算方式，则基于所选择的权重计算方式，根据所述每一个相关词与所述关键词的相关度，计算出在进行简历搜索时每一个相关词所占的权重值及所述关键词所占的权重值，具体的实施过程为：

根据第一权重公式 $W_i = \frac{r_i}{r_{\max}}$ ，计算出所述相关词集中第 i 个相关词在进行简历搜索时所

占的权重值 W_i ；其中， $r_{\max}$ 为在所述相关词集中所记载的所有相关词的相关度的最大值；

根据第二权重公式 $W_A = m \times \text{sum}W$ ，计算出所述关键词在进行简历搜索时所占的权重值 W_A ；其中， A 为所述关键词， m 为第一权重系数， $\text{sum}W$ 为所述相关词集中所有相关词的权重值的总和。

需要说明的是，采用上述权重计算算法，进行计算关键词和相关词的权重值，计算过程简单，能快速计算出关键词和相关词的权重值，但本权重计算算法仅适合相关词的数量适当时，才能够即能快速计算出权重值，并能确保关键词所占的权重值处于恰当的区间内。而当相关词的数量过多或过小时，即不在上述上限词量和下限词量的范围之内，采用第二权重计算算法进行计算，具体为步骤 S4 的实施过程。

进一步地，上述步骤 S2 中，当所述相关词集中的相关词的数量不处于所述上限词量和所述下限词量之间时，选择第二方式作为权重计算方式，则基于所选择的权重计算方式，根据所述每一个相关词与所述关键词的相关度，计算出在进行简历搜索时每一个相关词所占的权重值及所述关键词所占的权重值，具体的实施过程为：

根据基准分值公式 $S_i = \frac{r_i - r_{\min} + 1}{r_{\max} - r_{\min} + 1}$ ，计算出所述相关词集中第 i 个相关词的基准分值 S_i ；

其中， r_i 为所述相关词集中的第 i 个相关词的相关度； $r_{\min}$ 为在所述相关词集中所记载的所有相关词的相关度的最小值； $r_{\max}$ 为在所述相关词集中所记载的所有相关词的相关度的最大值；

对所述所有相关词的基准分值进行求和，获得基准分值总值 sum ；

根据第三权重公式 $W_A = 5 + \log_{1.5}(\text{sum} + 1)$ ，计算出所述关键词在进行简历搜索时所占的权重值 W_A ；其中， A 为所述关键词；

判断所述第 i 个相关词的基准分值是否大于 $\frac{W_A}{n}$ ；若是，则根据第四权重公式

$W_i = \frac{W_A}{n \times \log_{10}(k + 10)}$ ，计算出所述第 i 个相关词在进行简历搜索时所占的权重值 W_i ；若否，

则根据第五权重公式 $W_i = \frac{S_i}{\log_{10}(k + 10)}$ ，计算出所述第 i 个相关词在进行简历搜索时所占的权重值 W_i ；其中， n 为第二权重系数； k 为所述相关词集中的相关词的数量。

需要说明的是，计算相关词集中每个相关词的基准分值的目的是将相关词的相关度基于同一个基准，以一个分值的形式，表示该相关词与关键词的相关程度，该基准值优选为 $r_{\max} - r_{\min} + 1$ ；第一权重公式采用对数形式的公式，在基准分值总值增长过快时，关键词的权重值仍保持缓和增长，因而，可以在相关词相对较多的情况下，避免关键词所占权重过大这种情况的出现；在进行相关词的权重计算时，分别采用第二权重公式和第三权重公式进行计

算,是考虑到当相关词的基准分值过大和过小这两种情况出现时,避免计算出该相关词的权重值大于关键词的权重值,或者是大大小于关键词的权重值,也就是说在检索简历过程中,能避免相关词过多把关键词淹没、以及出现关键词过于突出的情况;对于第二权重系数的设置,可由实际情况来决定,一般取值为 2 或 4。

再进一步地,上述步骤 S2 的具体实施过程还包括以下步骤:

在所述判断所述第 i 个相关词的基准分值是否大于 $\frac{W_4}{n}$ 之前,根据均值公式

$$\bar{S} = \frac{\bar{r}}{r_{\max} - r_{\min} + 1}, \text{ 计算基准分值均值 } \bar{S}; \text{ 其中, } \bar{r} \text{ 为所述所有相关词的相关度的平均值;}$$

对于所述相关词集中的第 i 个相关词,判断所述第 i 个相关词的基准分值 S_i 是否大于所述基准分值均值 $\bar{S}$;

若是,通过更新公式 $S_i = 10 \times (S_i - \bar{S})$,更新所述第 i 个相关词的基准分值 S_i 。

需要说明的是,将该相关词集中的相关词的相关度均值基于同一个基准,且该基准与上述基准相同,目的是以一个分值的形式,表示该相关词集与关键词的平均相关程度;当一个相关词与关键词的相关程度(即基准分值)大于该相关词集与关键词的平均相关程度(平均分)时,可将该相关词与关键词的相关程度,与该相关词集与关键词的平均相关程度这两个相关程度的差异值放大 10 倍,作为该相关词与关键词的相关程度的基准分值;反之,则保持该相关词与关键词的相关程度的原基准分值不变,这种方式更能确切地表达该相关词与关键词的相关程度,使得相关词与关键词之间的关系更为合理。

更进一步地,上述步骤 S3 的具体实施过程为:

根据搜索引擎的权重关联格式,将所述关键词所占的权重值和所述关键词相关联作为第一组合,以及将所述每一个相关词所占的权重值与其对应的相关词相关联作为第二组合;

将所述第一组合和所述第二组合提交给搜索引擎,以供所述搜索引擎从所述简历数据库中搜索出简历,并根据所述搜索引擎内置的排序算法显示搜索出的简历;其中,所述简历中包含的所述关键词和所述每一个相关词以高亮形式显示。

需要说明的是,一般采用的搜索引擎为 solr 搜索引擎,则上述权重关联格式,可依据如下格式进行关联:关键词^关键词权重、相关词 1^相关词 1 权重、相关词 2^相关词 2 权重.....

实施本发明实施例的简历搜索方法,能根据相关词的数量选取不同的权重计算算法,来进行计算关键词和相关词在搜索过程所占的权重,充分考虑相关词的数量对关键词在检索过程的影响,平衡关键词和相关词在简历搜索过程中所占的权重,避免在权重计算过程中出现当相关词过多时导致关键词权重过大或过小,当相关词过少时关键词的权重过大或过小的权

重失衡情况，使简历检索的过程更为合理，减少再次进行简历搜索的发生，提高简历搜索的效率。

参见图 2，是本发明提供的简历搜索装置的一个实施例的结构示意图，该简历搜索装置能实施上述简历搜索方法的全部流程，其具体结构如下：

接收模块 10，用于接收用于从简历数据库中进行简历搜索的关键词和相关词集；其中，所述相关词集记载有多个相关词，以及每一个相关词与所述关键词的相关度；

权重计算模块 20，用于根据所述相关词集中的相关词的数量选择权重计算方式，并基于所选择的权重计算方式，根据所述每一个相关词与所述关键词的相关度，计算出在进行简历搜索时每一个相关词所占的权重值及所述关键词所占的权重值；

搜索模块 30，用于将所述关键词、所述关键词所占的权重值、每一个相关词和所述每一个相关词的权重值提交给搜索引擎，以从所述简历数据库中搜索出简历。

进一步地，参见图 3，是本发明提供的简历搜索装置的权重计算模块的一个实施例的结构示意图；该权重计算模块 20 包括：

第一计算单元 21，用于当所述相关词集中的相关词的数量处于上限词量和下限词量之间时，选择第一方式作为权重计算方式，根据第一权重公式 $W_i = \frac{r_i}{r_{\max}}$ ，计算出所述相关词集中第 i 个相关词在进行简历搜索时所占的权重值 W_i ；其中， $r_{\max}$ 为在所述相关词集中的所有相关词的相关度的最大值；

第二计算单元 22，用于根据第二权重公式 $W_A = m \times \text{sum}W$ ，计算出所述关键词在进行简历搜索时所占的权重值 W_A ；其中， A 为所述关键词， m 为第一权重系数， $\text{sum}W$ 为所述相关词集中所有相关词的权重值的总和。

进一步地，如图 3 所示，上述权重计算模块 20 还包括：

基准分值计算单元 23，用于当所述相关词集中的相关词的数量不处于所述上限词量和所述下限词量之间时，选择第二方式作为权重计算方式，根据基准分值公式 $S_i = \frac{r_i - r_{\min} + 1}{r_{\max} - r_{\min} + 1}$ ，计算出所述相关词集中第 i 个相关词的基准分值 S_i ；其中， r_i 为所述相关词集中的第 i 个相关词的相关度； $r_{\min}$ 为在所述相关词集中所记载的所有相关词的相关度的最小值； $r_{\max}$ 为在所述相关词集中所记载的所有相关词的相关度的最大值；

求和计算单元 24，用于对所述所有相关词的基准分值进行求和，获得所述相关词集的基准分值总值 sum ；

第三计算单元 25, 用于根据第三权重公式 $W_A = 5 + \log_{1.5}(sum + 1)$, 计算出所述关键词在进行简历搜索时所占的权重值 W_A ; 其中, A 为所述关键词;

第四计算单元 26, 用于判断所述第 i 个相关词的基准分值是否大于 W_A/n ; 若是, 则根据第四权重公式 $W_i = \frac{W_A}{n \times \log_{10}(k+10)}$, 计算出所述第 i 个相关词在进行简历搜索时所占的权重值 W_i ; 若否, 则根据第五权重公式 $W_i = \frac{S_i}{\log_{10}(k+10)}$, 计算出所述第 i 个相关词在进行简历搜索时所占的权重值 W_i ; 其中, n 为第二权重系数; k 为所述相关词集中的相关词的数量。

再进一步地, 如图 3 所示, 上述权重计算模块 20 还包括:

均值计算单元 27, 用于在所述判断所述第 i 个相关词的基准分值是否大于 W_A/n 之前, 根据平均分公式 $\bar{S} = \frac{\bar{r}}{r_{\max} - r_{\min} + 1}$, 计算基准分值均值 $\bar{S}$; 其中, $\bar{r}$ 为所述相关词集中的所有相关词的相关度的平均值;

判断单元 28, 用于对于所述相关词集中的第 i 个相关词, 判断所述第 i 个相关词的基准分值 S_i 是否大于所述基准分值均值 $\bar{S}$;

更新单元 29, 用于当判断所述第 i 个相关词的基准分值 S_i 是大于所述基准分值均值 $\bar{S}$ 时, 通过更新公式 $S_i = 10 \times (S_i - \bar{S})$, 更新所述第 i 个相关词的基准分值 S_i 。

更进一步地, 参见图 4, 是本发明提供的简历搜索装置的搜索模块的一个实施例的结构示意图; 该搜索模块 30 具体包括:

权重相联单元 31, 用于根据搜索引擎的权重关联格式, 将所述关键词所占的权重值和所述关键词相关联作为第一组合, 以及将所述每一个相关词所占的权重值与其对应的相关词相关联作为第二组合;

搜索显示单元 32, 用于将所述第一组合和所述第二组合提交给搜索引擎, 以供所述搜索引擎从所述简历数据库中搜索出简历, 并根据所述搜索引擎内置的排序算法显示搜索出的简历; 其中, 所述简历中包含的所述关键词和所述每一个相关词以高亮形式显示。

本发明实施例提供的简历搜索装置, 能根据相关词的数量选取不同的权重计算算法, 来进行计算关键词和相关词在搜索过程所占的权重, 充分考虑相关词的数量对关键词在检索过程的影响, 平衡关键词和相关词在简历搜索过程中所占的权重, 避免在权重计算过程中出现当相关词过多时导致关键词权重过大或过小, 当相关词过少时导致关键词的权重过大或过小的权重失衡情况, 使简历检索的过程更为合理, 减少再次进行简历搜索的发生, 提高简历搜

索的效率。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的程序可存储于一计算机可读取存储介质中，该程序在执行时，可包括如上述各方法的实施例的流程。其中，所述的存储介质可为磁碟、光盘、只读存储记忆体（Read-Only Memory, ROM）或随机存储记忆体（Random Access Memory, RAM）等。

以上所述是本发明的优选实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也视为本发明的保护范围。

权利要求书

1、一种简历搜索方法，其特征在于，包括：

接收用于从简历数据库中进行简历搜索的关键词和相关词集；其中，所述相关词集记载有多个相关词，以及每一个相关词与所述关键词的相关度；

根据所述相关词集中的相关词的数量选择权重计算方式，并基于所选择的权重计算方式，根据所述每一个相关词与所述关键词的相关度，计算出在进行简历搜索时每一个相关词所占的权重值及所述关键词所占的权重值；

将所述关键词、所述关键词所占的权重值、每一个相关词和所述每一个相关词的权重值提交给搜索引擎，以从所述简历数据库中搜索出简历。

2、如权利要求1所述的简历搜索方法，其特征在于，当所述相关词集中的相关词的数量处于上限词量和下限词量之间时，选择第一方式作为权重计算方式，则基于所选择的权重计算方式，根据所述每一个相关词与所述关键词的相关度，计算出在进行简历搜索时每一个相关词所占的权重值及所述关键词所占的权重值，具体为：

根据第一权重公式 $W_i = \frac{r_i}{r_{\max}}$ ，计算出所述相关词集中第 i 个相关词在进行简历搜索时所

占的权重值 W_i ；其中， $r_{\max}$ 为在所述相关词集中所记载的所有相关词的相关度的最大值；

根据第二权重公式 $W_A = m \times \sum W$ ，计算出所述关键词在进行简历搜索时所占的权重值 W_A ；其中， A 为所述关键词， m 为第一权重系数， $\sum W$ 为所述相关词集中所有相关词的权重值的总和。

3、如权利要求1或2所述简历搜索方法，其特征在于，当所述相关词集中的相关词的数量不处于所述上限词量和所述下限词量之间时，选择第二方式作为权重计算方式，则基于所选择的权重计算方式，根据所述每一个相关词与所述关键词的相关度，计算出在进行简历搜索时每一个相关词所占的权重值及所述关键词所占的权重值，具体为：

根据基准分值公式 $S_i = \frac{r_i - r_{\min} + 1}{r_{\max} - r_{\min} + 1}$ ，计算出所述相关词集中第 i 个相关词的基准分值 S_i ；

其中， r_i 为所述相关词集中的第 i 个相关词的相关度； $r_{\min}$ 为在所述相关词集中所记载的所有相关词的相关度的最小值； $r_{\max}$ 为在所述相关词集中所记载的所有相关词的相关度的最大值；

对所述所有相关词的基准分值进行求和，获得基准分值总值 $\sum$ ；

根据第三权重公式 $W_A = 5 + \log_{1.5}(sum + 1)$, 计算出所述关键词在进行简历搜索时所占的权重值 W_A ; 其中, A 为所述关键词;

判断所述第 i 个相关词的基准分值是否大于 $\frac{W_A}{n}$; 若是, 则根据第四权重公式 $W_i = \frac{W_A}{n \times \log_{10}(k + 10)}$, 计算出所述第 i 个相关词在进行简历搜索时所占的权重值 W_i ; 若否, 则根据第五权重公式 $W_i = \frac{S_i}{\log_{10}(k + 10)}$, 计算出所述第 i 个相关词在进行简历搜索时所占的权重值 W_i ; 其中, n 为第二权重系数; k 为所述相关词集中的相关词的数量。

4、如权利要求 3 所述的简历搜索方法, 其特征在于, 在所述判断所述第 i 个相关词的基准分值是否大于 $\frac{W_A}{n}$ 之前, 还包括:

根据均值公式 $\bar{S} = \frac{\bar{r}}{r_{\max} - r_{\min} + 1}$, 计算基准分值均值 $\bar{S}$; 其中, $\bar{r}$ 为所述所有相关词的相关度的平均值;

对于所述相关词集中的第 i 个相关词, 判断所述第 i 个相关词的基准分值 S_i 是否大于所述基准分值均值 $\bar{S}$;

若是, 通过更新公式 $S_i = 10 \times (S_i - \bar{S})$, 更新所述第 i 个相关词的基准分值 S_i 。

5、如权利要求 1 所述的简历搜索的方法, 其特征在于, 所述将所述关键词、所述关键词所占的权重值、每一个相关词和所述每一个相关词所占的权重值提交给搜索引擎, 以从所述简历数据库中搜索出简历, 具体为:

根据搜索引擎的权重关联格式, 将所述关键词所占的权重值和所述关键词相关联作为第一组合, 以及将所述每一个相关词所占的权重值与其对应的相关词相关联作为第二组合;

将所述第一组合和所述第二组合提交给搜索引擎, 以供所述搜索引擎从所述简历数据库中搜索出简历, 并根据所述搜索引擎内置的排序算法显示搜索出的简历; 其中, 所述简历中包含的所述关键词和所述每一个相关词以高亮形式显示。

6、一种简历搜索装置, 其特征在于, 包括:

接收模块, 用于接收用于从简历数据库中进行简历搜索的关键词和相关词集; 其中, 所述相关词集记载有多个相关词, 以及每一个相关词与所述关键词的相关度;

权重计算模块, 用于根据所述相关词集中的相关词的数量选择权重计算方式, 并基于所选择的权重计算方式, 根据所述每一个相关词与所述关键词的相关度, 计算出在进行简历搜索时每一个相关词所占的权重值及所述关键词所占的权重值;

搜索模块, 用于将所述关键词、所述关键词所占的权重值、每一个相关词和所述每一个相关词的权重值提交给搜索引擎, 以从所述简历数据库中搜索出简历。

7、如权利要求 6 所述的简历搜索装置, 其特征在于, 所述权重计算模块包括:

第一计算单元, 用于当所述相关词集中的相关词的数量处于上限词量和下限词量之间时, 选择第一方式作为权重计算方式, 根据第一权重公式 $W_i = \frac{r_i}{r_{\max}}$, 计算出所述相关词集中第 i 个相关词在进行简历搜索时所占的权重值 W_i ; 其中, $r_{\max}$ 为在所述相关词集中所记载的所有相关词的相关度的最大值;

第二计算单元, 用于根据第二权重公式 $W_A = m \times \text{sum}W$, 计算出所述关键词在进行简历搜索时所占的权重值 W_A ; 其中, A 为所述关键词, m 为第一权重系数, $\text{sum}W$ 为所述相关词集中所有相关词的权重值的总和。

8、如权利要求 6 或 7 所述简历搜索装置, 其特征在于, 所述权重计算模块还包括:

基准分值计算单元, 用于当所述相关词集中的相关词的数量不处于所述上限词量和所述下限词量之间时, 选择第二方式作为权重计算方式, 根据基准分值公式 $S_i = \frac{r_i - r_{\min} + 1}{r_{\max} - r_{\min} + 1}$, 计算出所述相关词集中第 i 个相关词的基准分值 S_i ; 其中, r_i 为所述相关词集中的第 i 个相关词的相关度; $r_{\min}$ 为在所述相关词集中所记载的所有相关词的相关度的最小值; $r_{\max}$ 为在所述相关词集中所记载的所有相关词的相关度的最大值;

求和计算单元, 用于对所述所有相关词的基准分值进行求和, 获得基准分值总值 sum ;

第三计算单元, 用于根据第三权重公式 $W_A = 5 + \log_{1.5}(\text{sum} + 1)$, 计算出所述关键词在进行简历搜索时所占的权重值 W_A ; 其中, A 为所述关键词;

第四计算单元, 用于判断所述第 i 个相关词的基准分值是否大于 $\frac{W_A}{n}$; 若是, 则根据第四权重公式 $W_i = \frac{W_A}{n \times \log_{10}(k + 10)}$, 计算出所述第 i 个相关词在进行简历搜索时所占的权重值

W_i ; 若否, 则根据第五权重公式 $W_i = \frac{S_i}{\log_{10}(k + 10)}$, 计算出所述第 i 个相关词在进行简历搜索

时所占的权重值 W_i ；其中， n 为第二权重系数； k 为所述相关词集中的相关词的数量。

9、如权利要求 8 所述的简历搜索装置，其特征在于，所述权重计算模块还包括：

均值计算单元，用于在所述判断所述第 i 个相关词的基准分值是否大于 W_i/n 之前，根据

均值公式 $\bar{S} = \frac{\bar{r}}{r_{\max} - r_{\min} + 1}$ ，计算基准分值均值 $\bar{S}$ ；其中， $\bar{r}$ 为所述所有相关词的相关度的平均

值；

判断单元，用于对于所述相关词集中的第 i 个相关词，判断所述第 i 个相关词的基准分值 S_i 是否大于所述基准分值均值 $\bar{S}$ ；

更新单元，用于当判断所述第 i 个相关词的基准分值 S_i 是大于所述平均分 $\bar{S}$ 时，通过更新公式 $S_i = 10 \times (S_i - \bar{S})$ ，更新所述第 i 个相关词的基准分值 S_i 。

10、如权利要求 6 所述的简历搜索的装置，其特征在于，所述搜索模块具体包括：

权重相联单元，用于根据搜索引擎的权重关联格式，将所述关键词所占的权重值和所述关键词相关联作为第一组合，以及将所述每一个相关词所占的权重值与其对应的相关词相关联作为第二组合；

搜索显示单元，用于将所述第一组合和所述第二组合提交给搜索引擎，以供所述搜索引擎从所述简历数据库中搜索出简历，并根据所述搜索引擎内置的排序算法显示搜索出的简历；其中，所述简历中包含的所述关键词和所述每一个相关词以高亮形式显示。

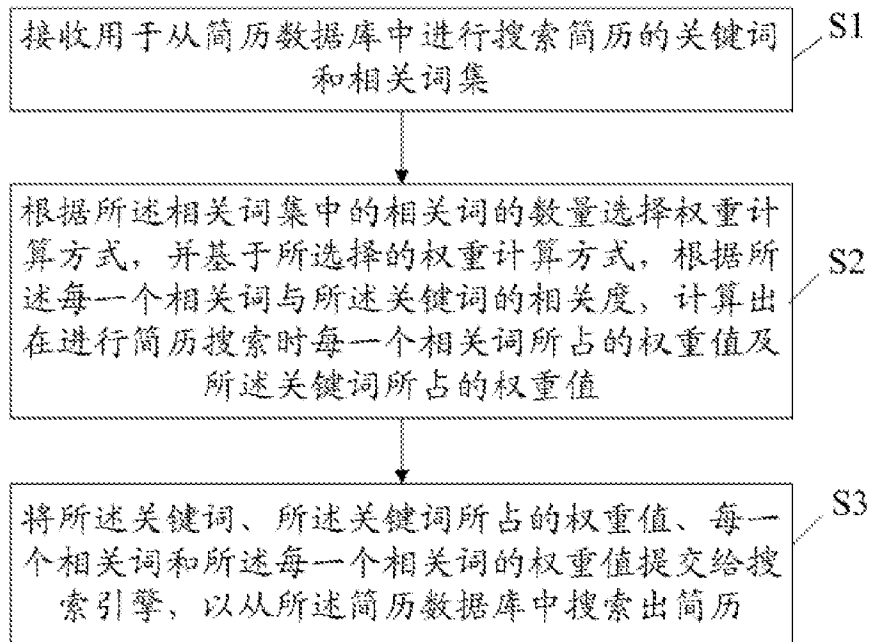


图 1

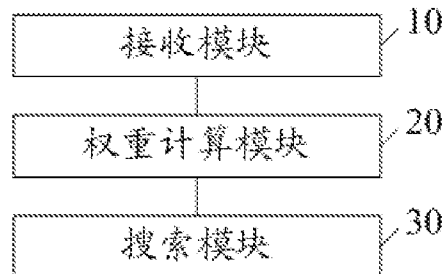


图 2

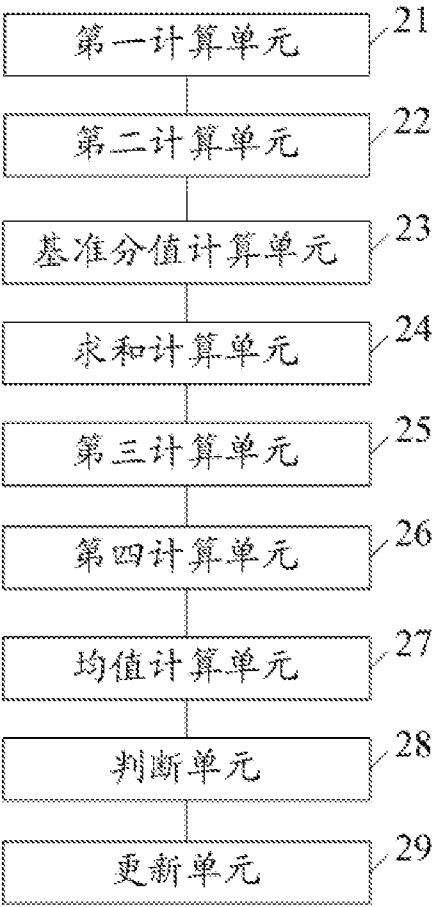


图 3

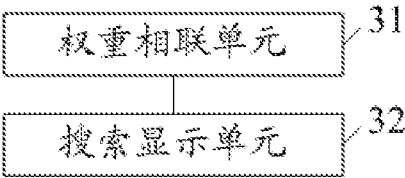


图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/113202

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)

G06F

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)

CNPAT, WPI, EPODOC, GOOGLE: search, key, word, relative, weight, value, score, directory, number, synonyms, choose, expand, set, similar

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 106095982 A (GUANGZHOU CVTE ELECTRONIC TECHNOLOGY CO., LTD.) 09 November 2016 (09.11.2016) claims 1, 2, 5 and 6, and description, paragraphs [0045]-[0073]	1, 6
PX	CN 105956195 A (GUANGZHOU CVTE ELECTRONIC TECHNOLOGY CO., LTD.) 21 September 2016 (21.09.2016) claims 1-10, and description, paragraphs [0045]-[0091]	1-10
A	CN 101140587 A (SHENZHEN THUNDER NETWORK TECHNOLOGY CO., LTD.) 12 March 2008 (12.03.2008) claims 16-34, description, page 8, paragraph 1 to page 14, paragraph 3, and figures 9 and 10	1-10
A	CN 103425687 A (ALIBABA GROUP HOLDING LTD.) 04 December 2013 (04.12.2013) the whole document	1-10
A	CN 103226618 A (FOCUS TECHNOLOGY CO., LTD.) 31 July 2013 (31.07.2013) the whole document	1-10

☒ 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 24 March 2017	Date of mailing of the international search report 12 April 2017
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 LIN, Fang Telephone No. (86-10) 62414085

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/113202

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 105653553 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 08 June 2016 (08.06.2016) the whole document	1-10
A	US 2013226936 A1 (HON HAI PRECISION INDUSTRY CO., LTD.) 29 August 2013 (29.08.2013) the whole document	1-10

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/113202

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 106095982 A	09 November 2016	None	
CN 105956195 A	21 September 2016	None	
CN 101140587 A	12 March 2008	None	
CN 103425687 A	04 December 2013	HK 1188841 A0	16 May 2014
CN 103226618 A	31 July 2013	None	
CN 105653553 A	08 June 2016	None	
US 2013226936 A1	29 August 2013	JP 2013175176 A	05 September 2013
		TW 201335770 A	01 September 2013
		CN 103294684 A	11 September 2013

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) G06F 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)) CNPAT, WPI, EPDOC, GOOGLE: 选择, 扩展词, 检索, 集合, 相似, 搜索, 权重, 词典, 词, 同义词, 相关词, 近义词, 词集, 相关, 数目, 数量, search, key, word, relative, weight, value, score, directory, number, synonyms																										
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 106095982 A (广州视源电子科技有限公司) 2016年 11月 9日 (2016 - 11 - 09) 权利要求1、2、5、6, 说明书[0045]-[0073]段</td> <td>1, 6</td> </tr> <tr> <td>PX</td> <td>CN 105956195 A (广州视源电子科技有限公司) 2016年 9月 21日 (2016 - 09 - 21) 权利要求1-10, 说明书[0045]-[0091]段</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 101140587 A (深圳市迅雷网络技术有限公司) 2008年 3月 12日 (2008 - 03 - 12) 权利要求16-34、说明书第8页第1段-第14页第3段, 图9、10</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 103425687 A (阿里巴巴集团控股有限公司) 2013年 12月 4日 (2013 - 12 - 04) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 103226618 A (焦点科技股份有限公司) 2013年 7月 31日 (2013 - 07 - 31) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 105653553 A (腾讯科技深圳有限公司) 2016年 6月 8日 (2016 - 06 - 08) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>US 2013226936 A1 (HON HAI PRECISION INDUSTRY CO., LTD.) 2013年 8月 29日 (2013 - 08 - 29) 全文</td> <td>1-10</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 106095982 A (广州视源电子科技有限公司) 2016年 11月 9日 (2016 - 11 - 09) 权利要求1、2、5、6, 说明书[0045]-[0073]段	1, 6	PX	CN 105956195 A (广州视源电子科技有限公司) 2016年 9月 21日 (2016 - 09 - 21) 权利要求1-10, 说明书[0045]-[0091]段	1-10	A	CN 101140587 A (深圳市迅雷网络技术有限公司) 2008年 3月 12日 (2008 - 03 - 12) 权利要求16-34、说明书第8页第1段-第14页第3段, 图9、10	1-10	A	CN 103425687 A (阿里巴巴集团控股有限公司) 2013年 12月 4日 (2013 - 12 - 04) 全文	1-10	A	CN 103226618 A (焦点科技股份有限公司) 2013年 7月 31日 (2013 - 07 - 31) 全文	1-10	A	CN 105653553 A (腾讯科技深圳有限公司) 2016年 6月 8日 (2016 - 06 - 08) 全文	1-10	A	US 2013226936 A1 (HON HAI PRECISION INDUSTRY CO., LTD.) 2013年 8月 29日 (2013 - 08 - 29) 全文	1-10
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																								
PX	CN 106095982 A (广州视源电子科技有限公司) 2016年 11月 9日 (2016 - 11 - 09) 权利要求1、2、5、6, 说明书[0045]-[0073]段	1, 6																								
PX	CN 105956195 A (广州视源电子科技有限公司) 2016年 9月 21日 (2016 - 09 - 21) 权利要求1-10, 说明书[0045]-[0091]段	1-10																								
A	CN 101140587 A (深圳市迅雷网络技术有限公司) 2008年 3月 12日 (2008 - 03 - 12) 权利要求16-34、说明书第8页第1段-第14页第3段, 图9、10	1-10																								
A	CN 103425687 A (阿里巴巴集团控股有限公司) 2013年 12月 4日 (2013 - 12 - 04) 全文	1-10																								
A	CN 103226618 A (焦点科技股份有限公司) 2013年 7月 31日 (2013 - 07 - 31) 全文	1-10																								
A	CN 105653553 A (腾讯科技深圳有限公司) 2016年 6月 8日 (2016 - 06 - 08) 全文	1-10																								
A	US 2013226936 A1 (HON HAI PRECISION INDUSTRY CO., LTD.) 2013年 8月 29日 (2013 - 08 - 29) 全文	1-10																								
☐ 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 2017年 3月 24日		国际检索报告邮寄日期 2017年 4月 12日																								
ISA/CN的名称和邮寄地址 中华人民共和国国家知识产权局 (ISA/CN) 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 林芳 电话号码 (86-10) 62414085																								

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/113202

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	106095982	A	2016年 11月 9日	无			
CN	105956195	A	2016年 9月 21日	无			
CN	101140587	A	2008年 3月 12日	无			
CN	103425687	A	2013年 12月 4日	HK	1188841	A0	2014年 5月 16日
CN	103226618	A	2013年 7月 31日	无			
CN	105653553	A	2016年 6月 8日	无			
US	2013226936	A1	2013年 8月 29日	JP	2013175176	A	2013年 9月 5日
				TW	201335770	A	2013年 9月 1日
				CN	103294684	A	2013年 9月 11日