

(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
2020年5月7日 (07.05.2020)



(10) 国际公布号
WO 2020/087655 A1

- (51) 国际专利分类号:
G06F 17/28 (2006.01)
- (21) 国际申请号: PCT/CN2018/119329
- (22) 国际申请日: 2018年12月5日 (05.12.2018)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201811276866.X 2018年10月30日 (30.10.2018) CN
- (71) 申请人: 科大讯飞股份有限公司 (IFLYTEK CO., LTD.) [CN/CN]; 中国安徽省合肥市高新区望江西路666号A1楼502室, Anhui 230088 (CN)。
- (72) 发明人: 孔常青 (KONG, Changqing); 中国安徽省合肥市高新区望江西路666号A1楼502室, Anhui 230088 (CN)。
- (74) 代理人: 北京集佳知识产权代理有限公司 (UNITALEN ATTORNEYS AT LAW); 中国北京市朝阳区建国门外大街22号赛特广场7层, Beijing 100004 (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,

(54) Title: TRANSLATION METHOD, APPARATUS AND DEVICE, AND READABLE STORAGE MEDIUM

(54) 发明名称: 一种翻译方法、装置、设备及可读存储介质

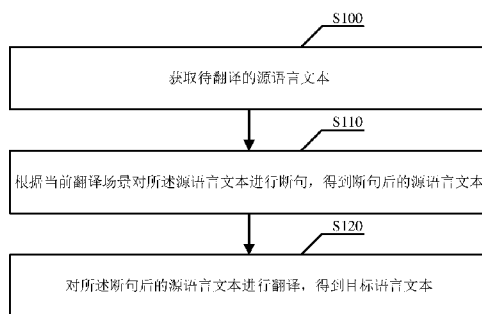


图 1

- S100 Obtain a source language text to be translated
- S110 Perform sentence segmentation on the source language text according to the current translation scene to obtain the source language text after sentence segmentation
- S120 Translate the source language text after sentence segmentation to obtain a target language text

(57) Abstract: Disclosed are a translation method, apparatus and device and a readable storage medium. The method comprises: obtaining a source language text to be translated; and performing sentence segmentation on the source language text further according to the current translation scene, so that the obtained source language text after sentence segmentation better conforms to the current translation scene. Obviously, compared with the existing translation method, the present application adds the sentence segmentation optimization process to the obtained source language text, namely, the sentence segmentation mode of the source language text is more



WO 2020/087655 A1

GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, 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, 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))。

optimized by considering the situation that the sentence segmentation is performed on the source language text again in the current translation scene, and on this basis, the source language text after the sentence segmentation is translated, so that the quality of the obtained target language text is higher.

(57) 摘要: 本申请公开了一种翻译方法、装置、设备及可读存储介质, 方法包括: 在获取到待翻译的源语言文本时, 进一步根据当前的翻译场景对源语言文本进行断句, 得到的断句后的源语言文本更加符合当前的翻译场景, 显然, 相比于现有翻译方法, 本申请对得到的源语言文本增加了断句优化过程, 即考虑了当前翻译场景对源语言文本进行重新断句, 使得源语言文本的断句方式更加优化, 进而基于此对断句后的源语言文本进行翻译, 得到的目标语言文本的质量也会更高。

一种翻译方法、装置、设备及可读存储介质

技术领域

5 本申请要求于 2018 年 10 月 30 日提交中国专利局、申请号为 201811276866.X、发明名称为“一种翻译方法、装置、设备及可读存储介质”的国内申请的优先权，其全部内容通过引用结合在本申请中。

背景技术

10 文本翻译的过程，即将待翻译的源语言文本翻译为目标语音文本的过程。对于待翻译的源语言文本，其断句方式并不规范，受源语言文本的来源的影响，如对于通过语音识别得到的源语言文本，其主要依靠语音的停顿信息进行断句，往往受说话人习惯影响。

现有技术基于此类断句方式并不优化的源语言文本进行机器翻译时，会大大影响机器翻译的质量。

发明内容

15 有鉴于此，本申请提供了一种翻译方法、装置、设备及可读存储介质，用于解决现有待翻译的源语言文本断句不优化，导致机器翻译质量低的问题。

为了实现上述目的，现提出的方案如下：

一种翻译方法，包括：

20 获取待翻译的源语言文本；

根据当前翻译场景对所述源语言文本进行断句，得到断句后的源语言文本；

对所述断句后的源语言文本进行翻译，得到目标语言文本。

25 优选地，所述根据翻译场景对所述源语言文本进行断句，得到断句后的源语言文本，包括：

将所述源语言文本输入预置的文本断句模型，得到文本断句模型输出的断句后的源语言文本；

其中，所述文本断句模型为，以源语言训练文本作为训练数据，以所

述源语言训练文本的符合所述当前翻译场景的断句结果作为训练标签训练得到。

优选地，所述文本断句模型的确定过程包括：

获取源语言训练文本；

- 5 确定所述源语言训练文本的符合所述当前翻译场景的断句结果，作为目标断句结果；

以所述源语言训练文本作为训练数据，以所述目标断句结果作为训练标签，训练文本断句模型。

- 10 优选地，所述确定所述源语言训练文本的符合所述当前翻译场景的断句结果，作为目标断句结果，包括：

获取所述源语言训练文本在所述当前翻译场景下的翻译后的目标语言训练文本；

- 15 参考设定的断句更改方式，对所述源语言训练文本的断句方式进行更改，得到更改后的源语言训练文本，由更改后的源语言训练文本及所述源语言训练文本组成候选源语言训练文本；

利用预置的机器翻译模型，对每一所述候选源语言训练文本进行翻译，得到每一所述候选源语言训练文本的机器翻译结果；

- 20 确定每一所述候选源语言训练文本的机器翻译结果，与所述目标语言训练文本的相似度，将相似度最高的候选源语言训练文本作为所述目标断句结果。

优选地，所述参考设定的断句更改方式，对所述源语言训练文本的断句方式进行更改，得到更改后的源语言训练文本，包括：

确定所述源语言训练文本包含的非终止型标点；

- 25 将所述源语言训练文本包含的每一非终止型标点，使用终止型标点进行替换，得到更改后的源语言训练文本。

优选地，所述利用预置的机器翻译模型，对每一所述候选源语言训练文本进行翻译，得到每一所述候选源语言训练文本的机器翻译结果，包括：

将每一所述候选源语言训练文本按照其包含的终止型标点进行子句划分，得到划分后的子句序列；

利用预置的机器翻译模型，对所述候选源语言训练文本的子句序列中每一子句分别进行翻译，得到每一子句的机器翻译结果；

按照子句序列中各子句的顺序，将各子句的机器翻译结果合并，得到所述候选源语言训练文本的机器翻译结果。

5 优选地，在所述以所述源语言训练文本作为训练数据，以所述目标断句结果作为训练标签，训练文本断句模型之前，该方法还包括：

获取人工对所述源语言训练文本标注标点的结果，得到人工标注后的源语言训练文本；

10 以所述源语言训练文本作为训练数据，以所述人工标注后的源语言训练文本作为训练标签，训练文本断句模型，得到初步文本断句模型；

则，所述以所述源语言训练文本作为训练数据，以所述目标断句结果作为训练标签，训练文本断句模型，包括：

以所述源语言训练文本作为训练数据，以所述目标断句结果作为训练标签，训练所述初步文本断句模型。

15 优选地，所述对所述断句后的源语言文本进行翻译，得到目标语言文本，包括：

将所述断句后的源语言文本按照其包含的终止型标点进行子句划分，得到划分后的子句序列；

20 利用预置的机器翻译模型，对所述断句后的源语言文本的子句序列中每一子句分别进行翻译，得到每一子句的机器翻译结果；

按照子句序列中各子句的顺序，将各子句的机器翻译结果合并，得到所述目标语言文本。

一种翻译装置，包括：

源语言文本获取单元，用于获取待翻译的源语言文本；

25 文本断句单元，用于根据当前翻译场景对所述源语言文本进行断句，得到断句后的源语言文本；

源语言文本翻译单元，用于对所述断句后的源语言文本进行翻译，得到目标语言文本。

优选地，所述文本断句单元包括：

模型参考单元，用于将所述源语言文本输入预置的文本断句模型，得到文本断句模型输出的断句后的源语言文本；

其中，所述文本断句模型为，以源语言训练文本作为训练数据，以所述源语言训练文本的符合所述当前翻译场景的断句结果作为训练标签训练得到。

优选地，还包括：文本断句模型确定单元，用于确定文本断句模型；所述文本断句模型包括：

源语言训练文本获取单元，用于获取源语言训练文本；

断句结果确定单元，用于确定所述源语言训练文本的符合所述当前翻译场景的断句结果，作为目标断句结果；

第一模型训练单元，用于以所述源语言训练文本作为训练数据，以所述目标断句结果作为训练标签，训练文本断句模型。

优选地，所述断句结果确定单元包括：

目标语言训练文本获取单元，用于获取所述源语言训练文本在所述当前翻译场景下的翻译后的目标语言训练文本；

断句更改单元，用于参考设定的断句更改方式，对所述源语言训练文本的断句方式进行更改，得到更改后的源语言训练文本，由更改后的源语言训练文本及所述源语言训练文本组成候选源语言训练文本；

源语言训练文本翻译单元，用于利用预置的机器翻译模型，对每一所述候选源语言训练文本进行翻译，得到每一所述候选源语言训练文本的机器翻译结果；

相似度确定单元，用于确定每一所述候选源语言训练文本的机器翻译结果，与所述目标语言训练文本的相似度，将相似度最高的候选源语言训练文本作为所述目标断句结果。

优选地，所述断句更改单元包括：

非终止型标点确定单元，用于确定所述源语言训练文本包含的非终止型标点；

非终止型标点替换单元，用于将所述源语言训练文本包含的每一非终止型标点，使用终止型标点进行替换，得到更改后的源语言训练文本。

优选地，所述源语言训练文本翻译单元包括：

第一子句划分单元，用于将每一所述候选源语言训练文本按照其包含的终止型标点进行子句划分，得到划分后的子句序列；

5 第一子句翻译单元，用于利用预置的机器翻译模型，对所述候选源语言训练文本的子句序列中每一子句分别进行翻译，得到每一子句的机器翻译结果；

第一翻译结果合并单元，用于按照子句序列中各子句的顺序，将各子句的机器翻译结果合并，得到所述候选源语言训练文本的机器翻译结果。

优选地，所述文本断句模型还包括：

10 人工标注结果获取单元，用于获取人工对所述源语言训练文本标注标点的结果，得到人工标注后的源语言训练文本；

第二模型训练单元，用于以所述源语言训练文本作为训练数据，以所述人工标注后的源语言训练文本作为训练标签，训练文本断句模型，得到初步文本断句模型；

15 则所述第一模型训练单元具体用于：

以所述源语言训练文本作为训练数据，以所述目标断句结果作为训练标签，训练所述初步文本断句模型。

优选地，所述源语言文本翻译单元，包括：

20 第二子句划分单元，用于将所述断句后的源语言文本按照其包含的终止型标点进行子句划分，得到划分后的子句序列；

第二子句翻译单元，用于利用预置的机器翻译模型，对所述断句后的源语言文本的子句序列中每一子句分别进行翻译，得到每一子句的机器翻译结果；

25 第二翻译结果合并单元，用于按照子句序列中各子句的顺序，将各子句的机器翻译结果合并，得到所述目标语言文本。

一种翻译设备，包括存储器和处理器；

所述存储器，用于存储程序；

所述处理器，用于执行所述程序，实现如上所述的翻译方法的各个步骤。

一种可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时，实现如上所述的翻译方法的各个步骤。

从上述的技术方案可以看出，本申请实施例提供的翻译方法，在获取到待翻译的源语言文本时，进一步根据当前的翻译场景对源语言文本进行断句，得到的断句后的源语言文本更加符合当前的翻译场景，显然，相比于现有翻译方法，本申请对得到的源语言文本增加了断句优化过程，即考虑了当前翻译场景对源语言文本进行重新断句，使得源语言文本的断句方式更加优化，进而基于此对断句后的源语言文本进行翻译，得到的目标语言文本的质量也会更高。

10

附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图1 为本申请实施例公开的一种翻译方法流程图；
图2为本申请实施例公开的一种翻译装置结构示意图；
图3为本申请实施例公开的一种翻译设备的硬件结构框图。

20

具体实施方式

下面将结合本申请实施例中的附图，对本申请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申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的

范围。

文本翻译的过程即将待翻译的源语言文本，翻译为目标语言文本。根

据待翻译的源语言文本的不同来源，其断句方式也不唯一，以待翻译的源语言文本为对待翻译语音识别得到为例进行说明。在不同的翻译场景下，源语言文本的不同断句方式，会影响基于源语言文本翻译后的目标语言文本的质量。示例如，源语言文本在不同的上下文环境中，其翻译结果可能会存在差别。再比如，在不同的翻译场合源语言文本的翻译结果也可能存在差别，如在会议场合下源语言文本的翻译结果要求更加严谨、规范，而在聊天场合下源语言文本的翻译结果可能会更加随意、口语化。

现有技术中，对于待翻译的源语言文本，其直接送入机器翻译模型进行翻译，而源语言文本的断句方式不规范，如通过语音识别得到的源语言文本可能受说话人说话习惯影响，其断句方式并不优化，且并未考虑当前翻译场景，基于此翻译后的结果质量也不高。为此，本申请提供了一种优化后的翻译方法。本申请的翻译方法可以应用于具备数据处理能力的电子设备中。

接下来结合附图1对本案的翻译方法进行介绍，该方法可以包括：
步骤 S100、获取待翻译的源语言文本。

具体地，待翻译的源语言文本可以通过多种途径获取，如用户上传的源语言文本或者接收用户的语音数据进行语音识别得到的文本。以语音翻译过程为例，可以利用语音端点检测技术对获取的实时语音进行处理，得到语音片段。进一步对语音片段进行识别，得到识别后的文本作为待翻译的源语言文本。

这里，源语言即为待翻译文本使用的语言。对应的，翻译后的语言定义为目标语言，本申请的目的也即对源语言文本翻译，得到目标语言文本。

步骤 S110、根据当前翻译场景对所述源语言文本进行断句，得到断句后的源语言文本。

可以理解的是，上一步骤中得到的源语言文本中的断句方式（即源语言文本中的标点）可能受说话人说话习惯的影响，其断句方式并不规范，也未考虑当前的翻译场景，若直接对获取的源语言文本进行翻译，其翻译结果质量也不高。

为此，本步骤中增加了对源语言文本进行断句处理的过程，且该断句

处理过程考虑了当前翻译场景，使得断句后的源语言文本的断句方式更加符合当前翻译场景。对于源语言文本断句处理的详细过程，下文将详细介绍。

步骤 S120、对所述断句后的源语言文本进行翻译，得到目标语言文本。

- 5 一般的，可以使用机器翻译模型对上一步骤中得到的断句后的源语言文本进行翻译，得到翻译后的目标语言文本。

在此基础上，本申请实施例还可以根据用户需要，选择将目标语言文本合成为语音，进而进行语音播报，实现从源语言语音到目标语言语音的转换过程。

- 10 本申请实施例提供的翻译方法，在获取到待翻译的源语言文本时，进一步根据当前的翻译场景对源语言文本进行断句，得到的断句后的源语言文本更加符合当前的翻译场景，显然，相比于现有翻译方法，本申请对得到的源语言文本增加了断句优化过程，即考虑了当前翻译场景对源语言文本进行重新断句，使得源语言文本的断句方式更加优化，进而基于此对断句后的源语言文本进行翻译，得到的目标语言文本的质量也会更高。
- 15

在本申请的另一个实施例中，对上述步骤S110，根据当前翻译场景对所述源语言文本进行断句，得到断句后的源语言文本的过程进行介绍。

- 首先，可以理解的是，不同的翻译场景下断句方式会存在一定的特点，因此本申请可以预先设置与每个翻译场景对应的断句方式规则。示例如，针对会议场合下，可能需要尽量多的使用短句，也即尽量多的使用终止型标点。则可以设置会议场合这一翻译场景对应的断句方式规则中，使用的终止型标点的个数大于非终止型标点。
- 20

- 这里，按照是否能够完整的表达句意，将标点划分为终止型标点和非终止型标点两类，其中终止型标点代表能够完整的表达句意，如句号、问号、感叹号等。非终止型标点代表不能够完整的表达句意，如逗号，顿号等。
- 25

基于此，在进行当前的翻译时，可以查询预置的对应关系，确定当前翻译场景对应的断句方式规则。进而在获取到待翻译的源语言文本之后，

按照确定的断句方式规则，对源语言文本进行断句处理，得到断句后的源语言文本。显然，该断句后的源语言文本能够符合当前翻译场景的需要。

5 进一步地，本申请实施例还提供了另一种对源语言文本进行断句的处理方式，即可以使用机器学习模型来执行对源语言文本进行断句的过程，详细过程如下介绍：

10 定义本实施例中进行断句处理的机器学习模型为文本断句模型，其可以使用现有的各种结构的机器学习模型，如序列标注框架下的BLSTM模型、Self-Attention模型等，或者是编解码Encode-Decode框架下的序列生成模型等，当然还可以使用现有多结构模型的组合。

当然，若采用序列标注框架下的模型，则模型的输入为文本序列中的每个词，模型的输出为每个词对应的标点类别，该标点类别可以是空值、逗号、句号、问号等，其中空值代表对于词后不添加任何标点。

15 若采用Encode-Decode框架下的模型，则模型的输入可以是不带标点的文本序列，模型的输出是包含标点信息的文本序列，也即由模型对输入文本序列添加标点后的结果。具体采用哪种形式的机器学习模型可以根据应用需要来选择，本申请不做严格限定。

20 进一步地，在确定的文本断句模型的结构之后，进一步需要获取模型的训练数据以对文本断句模型进行训练。本申请实施例中可以收集大量的源语言训练文本，作为训练数据。定义源语言训练文本组成的集合为T1。进一步地，还需要确定T1中每一源语言训练文本的符合当前翻译场景的断句结果，作为对应源语言训练文本的训练标签，由该训练标签配合训练数据一起训练文本断句模型。可以理解的是，本实施例中获取的训练数据可以是从待翻译的源语言文本中抽取的。除此之外，还可以通过其它途径获取训练数据，例如，从现有的材料文本中选取部分文本，作为训练数据。

25 依据上述训练数据和训练标签训练后的文本断句模型，具备了将输入样本按照符合当前翻译场景的需要进行断句处理，输出符合当前翻译场景的断句结果的能力。基于此，可以将获取的源语言文本输入文本断句模型，得到文本断句模型输出的断句后的源语言文本，也即得到断句优化处理后

的源语言文本。

本申请的另一个实施例中，对上述文本断句模型的确定过程进行展开说明，文本断句模型的确定过程可以包括：

5 A1、获取源语言训练文本。

同上介绍的，定义源语言训练文本组成的集合为 T1。

A2、确定所述源语言训练文本的符合所述当前翻译场景的断句结果，作为目标断句结果。

10 定义源语言训练文本的符合当前翻译场景的目标断句结果组成的集合为 T2。T2 为对 T1 中每一源语言训练文本进行翻译的结果。

A3、以所述源语言训练文本作为训练数据，以所述目标断句结果作为训练标签，训练文本断句模型。

在上述示例的文本断句模型的确定过程的基础上，本申请实施例提供了另一种文本断句模型的确定方式，即在上述 A3 之前，增加如下步骤：

15 A4、获取人工对所述源语言训练文本标注标点的结果，得到人工标注后的源语言训练文本。

具体地，在获取源语言训练文本之后，可以由人工对源语言训练文本进行标点的标注，得到人工标注后的源语言训练文本。

20 A5、以所述源语言训练文本作为训练数据，以所述人工标注后的源语言训练文本作为训练标签，训练文本断句模型，得到初步文本断句模型。

在 A4 步骤的基础上，可以、以所述源语言训练文本作为训练数据，以所述人工标注后的源语言训练文本作为训练标签，训练文本断句模型，得到初步文本断句模型。

在此基础上，上述 A3 步骤具体可以包括：

25 以所述源语言训练文本作为训练数据，以所述目标断句结果作为训练标签，训练所述初步文本断句模型。

具体地，可以采用模型自适应更新方法，采用所述源语言训练文本作为训练数据，所述目标断句结果作为训练标签，对初步文本断句模型进行参数更新。

采用这种模型更新方法，可以提高模型训练数据量，进而使得训练得到的文本断句模型更加优秀。

5 在本申请的另一个实施例中，对上述A2确定所述源语言训练文本的符合所述当前翻译场景的断句结果，作为目标断句结果的过程进行介绍。

可以理解的是，上述已经说明可以预先设置与每个翻译场景对应的断句方式规则。则本实施例中可以按照当前翻译场景对应的断句方式规则，对源语言训练文本进行断句处理，得到每一源语言训练文本的断句结果。

10 此外，本实施例中还提供了另一种可选的实施方式，具体可以包括：
A21、获取所述源语言训练文本在所述当前翻译场景下的翻译后的目标语言训练文本。

具体地，源语言训练文本在当前翻译场景下的翻译后的目标语言训练文本可以通过人工翻译的方式确定。也即，可以由人工根据当前翻译场景，对源语言训练文本进行翻译，得到目标语言训练文本。

15 A22、参考设定的断句更改方式，对所述源语言训练文本的断句方式进行更改，得到更改后的源语言训练文本，由更改后的源语言训练文本及所述源语言训练文本组成候选源语言训练文本。

具体地，本申请实施例可以预先设定断句更改方式，进而可以按照设定的断句更改方式，对源语言训练文本的断句方式进行更改。

20 可以理解的是，通过合理设置断句更改方式，可以扩展出多条更改后的源语言训练文本。源语言训练文本的符合当前翻译场景的断句方式，或者是源语言训练文本自身的断句方式，或者是某一条更改后的源语言训练文本的断句方式。

25 也即，本步骤的处理过程是为了扩展出候选源语言训练文本，该候选源语言训练文本种包含了源语言训练文本的符合当前翻译场景的断句方式。

A23、利用预置的机器翻译模型，对每一所述候选源语言训练文本进行翻译，得到每一所述候选源语言训练文本的机器翻译结果。

A24、确定每一所述候选源语言训练文本的机器翻译结果，与所述目

标语言训练文本的相似度，将相似度最高的候选源语言训练文本作为所述目标断句结果。

具体地，目标语言训练文本为源语言训练文本在所述当前翻译场景下的翻译后结果。基于此，本步骤中以目标语言训练文本为标准，确定每一候选源语言训练文本的机器翻译结果，与目标语言训练文本的相似度。可以理解的是，相似度越高的候选源语言训练文本，说明其与当前翻译场景的符合程度越高。基于此，可以选取相似度最高的候选源语言训练文本，将其作为源语言训练文本的符合当前翻译场景的目标断句结果。

10 可选的，在本步骤中计算相似度时，可以采用BLEU打分的方式，即以目标语言训练文本为标准，分别对每一候选源语言训练文本的机器翻译结果进行打分评价，打分值越高的候选源语言训练文本代表与目标语言训练文本的相似度越高。

进一步地，本实施例介绍了一种可选的对源语言训练文本的断句方式进行更改，得到更改后的源语言训练文本的方式，具体可以包括：

15 A221、确定所述源语言训练文本包含的非终止型标点。

具体地，对于源语言训练文本集合 T1 中的每个源语言训练文本 $T1_j(j=1\dots n)$ ，n 为 T1 中源语言训练文本的条数，确定 $T1_j$ 包含的非终止型标点的个数 M。

20 A222、将所述源语言训练文本包含的每一非终止型标点，使用终止型标点进行替换，得到更改后的源语言训练文本。

可以理解的是， $T1_j$ 中任意一个非终止型标点可以使用终止型标点进行替换。

25 按照本步骤介绍的替换方式，若 $T1_j$ 包含的非终止型标点的个数为 M，则由替换前的源语言训练文本，以及通过替换得到的更改后的源语言训练文本组成候选源语言训练文本集合，该集合中共包含 2^M （2 的幂次方）条候选源语言训练文本。

下面通过一个例子进行说明：

具体地，对于源语言训练文本：“今天天气不错，我想去爬山，你去吗？”，由于其中有两个逗号是非终止型标点，故每个逗号都可以用终止型

标点如句号替换，最终可以得到 $2^2=4$ 条候选源语言训练文本，如下：

- 1、今天天气不错，我想去爬山，你去吗？
- 2、今天天气不错。我想去爬山，你去吗？
- 3、今天天气不错，我想去爬山。你去吗？
- 5 4、今天天气不错。我想去爬山。你去吗？

可以理解的是，得到的 4 条候选源语言训练文本中，第 1 条为源语言训练文本本身，第 2-4 条为通过标点替换后得到的更改后的源语言训练文本。

- 10 基于上述实施例介绍的 A22 的实现方式，本申请实施例进一步介绍了上述 A23，利用预置的机器翻译模型，对每一所述候选源语言训练文本进行翻译的一种可选实施方式，具体可以包括：

A231、将每一所述候选源语言训练文本按照其包含的终止型标点进行子句划分，得到划分后的子句序列。

- 15 具体地，对于每一所述候选源语言训练文本，从头开始遍历其中包含的终止型标点，将每一终止型标点处作为一个划分点，将候选源语言训练文本划分为若干子句，划分后的各子句按照在候选源语言训练文本中的先后顺序，组成子句序列。

- A232、利用预置的机器翻译模型，对所述候选源语言训练文本的子句序列中每一子句分别进行翻译，得到每一子句的机器翻译结果。
- 20

A232、按照子句序列中各子句的顺序，将各子句的机器翻译结果合并，得到所述候选源语言训练文本的机器翻译结果。

- 可以理解的是，候选源语言训练文本的条数为 2^M ，针对每条候选语言训练文本均按照上述方式进行翻译，则最终可以得到 2^M 条机器翻译结果。
- 25

按照本申请介绍的上述处理方式，可以将一部分非终止型标点转化为终止型标点，则终止型标点的出现概率会提高，而在机器翻译过程，其是以终止型标点之前的内容进行一次翻译，因此按照本申请方案会缩短等待终止型标点的时间，从而提高翻译结果的产出速度，降低用户主观上等待

翻译结果的时间，提升了用户的体验。

仍以上述示例的例子来说明 A23 的实现过程：

为了便于表述，将上述示例的 4 条候选源语言训练文本分别定义为候选文本 1-4。

- 5 对于候选文本 1：由于该句中只有最后出现终止型标点，句中并没有终止型标点，因此该句子无法进一步拆分，或者说拆分后的子句即为候选文本 1 本身。因此，可以将候选文本 1 作为一个句子送入机器翻译模型进行翻译。

- 10 对于候选文本 2：该句中“不错”后是一个句号，可以对句子进行拆分，候选文本 2 可以拆分得到两个子句，分别为：

子句 21：今天天气不错。

子句 22：我想去爬山，你去吗？

对于拆分后的两个子句，分别送入机器翻译模型进行翻译，并将机器翻译结果合并，得到候选文本 2 的机器翻译结果。

- 15 对于候选文本 3：该句中“爬山”后是一个句号，可以对句子进行拆分，候选文本 3 可以拆分得到两个子句，分别为：

子句 31：今天天气不错，我想去爬山。

子句 32：你去吗？

- 20 对于拆分后的两个子句，分别送入机器翻译模型进行翻译，并将机器翻译结果合并，得到候选文本 3 的机器翻译结果。

对于候选文本 4：该句中“不错”和“爬山”后各是一个句号，可以对句子进行拆分，候选文本 4 可以拆分得到三个子句，分别为：

子句 41：今天天气不错。

子句 42：我想去爬山。

- 25 子句 43：你去吗？

对于拆分后的三个子句，分别送入机器翻译模型进行翻译，并将机器翻译结果合并，得到候选文本 4 的机器翻译结果。

此外进一步地，假设对于上述候选文本 1-4，采用 BLEU 方法进行打分，分值依次为：0.1，0.2，0.3，0.4。则可以选取分值最高的候选文本 4，作为

源语言训练文本的符合当前翻译场景的目标断句结果。

则，源语言训练文本：“今天天气不错，我想去爬山，你去吗？”

目标断句结果：“今天天气不错。我想去爬山。你去吗？”

5 可以将此源语言训练文本机器目标断句结果作为训练数据和训练标签，训练文本断句模型。

在本申请的又一个实施例中，对上述步骤S120，对所述断句后的源语言文本进行翻译，得到目标语言文本的过程进行介绍。

10 基于上述实施例的介绍可知，可以使用机器翻译模型对断句后的源语言文本进行翻译。在具体翻译过程，可以首先按照断句后的源语言文本包含的终止型标点进行子句划分，得到划分后的子句序列。进一步，利用预置的机器翻译模型，对断句后的源语言文本对应的子句序列中每一子句分别进行翻译，得到每一子句的机器翻译结果。按照子句序列中各子句的顺序，将各子句的机器翻译结果合并，得到所述断句后的源语言文本的机器
15 翻译结果，也即得到目标语言文本。

基于上述各实施例的介绍可知，本申请考虑了当前翻译场景对源语言文本进行断句，得到的断句后的源语言文本更加符合当前的翻译场景，进而基于此对断句后的源语言文本进行翻译，得到的目标语言文本的质量也会更高。

20 进一步地，本申请在断句时通过标点替换，可以将一部分非终止型标点转化为终止型标点，则终止型标点的出现概率会提高，而在机器翻译过程，其是以终止型标点之前的内容进行一次翻译，因此按照本申请方案会缩短等待终止型标点的时间，从而提高翻译结果的产出速度，降低用户主观上等待翻译结果的时间，提升了用户的体验。

25

下面对本申请实施例提供的翻译装置进行描述，下文描述的翻译装置与上文描述的翻译方法可相互对应参照。

参见图2，图2为本申请实施例公开的一种翻译装置结构示意图。

如图2所示，该装置可以包括：

源语言文本获取单元 11，用于获取待翻译的源语言文本；

文本断句单元 12，用于根据当前翻译场景对所述源语言文本进行断句，得到断句后的源语言文本；

源语言文本翻译单元 13，用于对所述断句后的源语言文本进行翻译，

5 得到目标语言文本。

可选的，上述文本断句单元可以包括：

模型参考单元，用于将所述源语言文本输入预置的文本断句模型，得到文本断句模型输出的断句后的源语言文本；

10 其中，所述文本断句模型为，以源语言训练文本作为训练数据，以所述源语言训练文本的符合所述当前翻译场景的断句结果作为训练标签训练得到。

可选的，本申请的翻译装置还可以包括：文本断句模型确定单元，用于确定文本断句模型。该文本断句模型可以包括：

源语言训练文本获取单元，用于获取源语言训练文本；

15 断句结果确定单元，用于确定所述源语言训练文本的符合所述当前翻译场景的断句结果，作为目标断句结果；

第一模型训练单元，用于以所述源语言训练文本作为训练数据，以所述目标断句结果作为训练标签，训练文本断句模型。

可选的，上述断句结果确定单元可以包括：

20 目标语言训练文本获取单元，用于获取所述源语言训练文本在所述当前翻译场景下的翻译后的目标语言训练文本；

断句更改单元，用于参考设定的断句更改方式，对所述源语言训练文本的断句方式进行更改，得到更改后的源语言训练文本，由更改后的源语言训练文本及所述源语言训练文本组成候选源语言训练文本；

25 源语言训练文本翻译单元，用于利用预置的机器翻译模型，对每一所述候选源语言训练文本进行翻译，得到每一所述候选源语言训练文本的机器翻译结果；

相似度确定单元，用于确定每一所述候选源语言训练文本的机器翻译结果，与所述目标语言训练文本的相似度，将相似度最高的候选源语言训

练文本作为所述目标断句结果。

可选的，上述断句更改单元可以包括：

非终止型标点确定单元，用于确定所述源语言训练文本包含的非终止型标点；

- 5 非终止型标点替换单元，用于将所述源语言训练文本包含的每一非终止型标点，使用终止型标点进行替换，得到更改后的源语言训练文本。

可选的，上述源语言训练文本翻译单元可以包括：

第一子句划分单元，用于将每一所述候选源语言训练文本按照其包含的终止型标点进行子句划分，得到划分后的子句序列；

- 10 第一子句翻译单元，用于利用预置的机器翻译模型，对所述候选源语言训练文本的子句序列中每一子句分别进行翻译，得到每一子句的机器翻译结果；

第一翻译结果合并单元，用于按照子句序列中各子句的顺序，将各子句的机器翻译结果合并，得到所述候选源语言训练文本的机器翻译结果。

- 15 可选的，上述文本断句模型还可以包括：

人工标注结果获取单元，用于获取人工对所述源语言训练文本标注标点的结果，得到人工标注后的源语言训练文本；

- 20 第二模型训练单元，用于以所述源语言训练文本作为训练数据，以所述人工标注后的源语言训练文本作为训练标签，训练文本断句模型，得到初步文本断句模型。基于此，上述第一模型训练单元具体可以用于：

以所述源语言训练文本作为训练数据，以所述目标断句结果作为训练标签，训练所述初步文本断句模型。

可选的，上述源语言文本翻译单元，可以包括：

- 25 第二子句划分单元，用于将所述断句后的源语言文本按照其包含的终止型标点进行子句划分，得到划分后的子句序列；

第二子句翻译单元，用于利用预置的机器翻译模型，对所述断句后的源语言文本的子句序列中每一子句分别进行翻译，得到每一子句的机器翻译结果；

第二翻译结果合并单元，用于按照子句序列中各子句的顺序，将各子

句的机器翻译结果合并，得到所述目标语言文本。

本申请实施例提供的翻译装置可应用于翻译设备，如 PC 终端、云平台、服务器及服务器集群等。可选的，图 3 示出了翻译设备的硬件结构框图，参照图 3，翻译设备的硬件结构可以包括：至少一个处理器 1，至少一个通信接口 2，至少一个存储器 3 和至少一个通信总线 4；

在本申请实施例中，处理器 1、通信接口 2、存储器 3、通信总线 4 的数量为至少一个，且处理器 1、通信接口 2、存储器 3 通过通信总线 4 完成相互间的通信；

10 处理器 1 可能是一个中央处理器 CPU，或者是特定集成电路 ASIC (Application Specific Integrated Circuit)，或者是被配置成实施本发明实施例的一个或多个集成电路等；

存储器 3 可能包含高速 RAM 存储器，也可能还包括非易失性存储器 (non-volatile memory) 等，例如至少一个磁盘存储器；

15 其中，存储器存储有程序，处理器可调用存储器存储的程序，所述程序用于：

获取待翻译的源语言文本；

根据当前翻译场景对所述源语言文本进行断句，得到断句后的源语言文本；

20 对所述断句后的源语言文本进行翻译，得到目标语言文本。

可选的，所述程序的细化功能和扩展功能可参照上文描述。

本申请实施例还提供一种可读存储介质，该可读存储介质可存储有适于处理器执行的程序，所述程序用于：

获取待翻译的源语言文本；

25 根据当前翻译场景对所述源语言文本进行断句，得到断句后的源语言文本；

对所述断句后的源语言文本进行翻译，得到目标语言文本。

可选的，所述程序的细化功能和扩展功能可参照上文描述。

最后，还需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，

5 从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个... ..”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

10 本说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似部分互相参见即可。

对所公开的实施例的上述说明，使本领域专业技术人员能够实现或使用本申请。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的，本文中所定义的一般原理可以在不脱离本申请的精神或范围的情况下，在其它实施例中实现。因此，本申请将不会被限制于本文所示的这些实施例，而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

15

权 利 要 求

1、一种翻译方法，其特征在于，包括：

获取待翻译的源语言文本；

5 根据当前翻译场景对所述源语言文本进行断句，得到断句后的源语言文本；

对所述断句后的源语言文本进行翻译，得到目标语言文本。

2、根据权利要求1所述的方法，其特征在于，所述根据翻译场景对所述源语言文本进行断句，得到断句后的源语言文本，包括：

10 将所述源语言文本输入预置的文本断句模型，得到文本断句模型输出的断句后的源语言文本；

其中，所述文本断句模型为，以源语言训练文本作为训练数据，以所述源语言训练文本的符合所述当前翻译场景的断句结果作为训练标签训练得到。

15 3、根据权利要求2所述的方法，其特征在于，所述文本断句模型的确定过程包括：

获取源语言训练文本；

确定所述源语言训练文本的符合所述当前翻译场景的断句结果，作为目标断句结果；

20 以所述源语言训练文本作为训练数据，以所述目标断句结果作为训练标签，训练文本断句模型。

4、根据权利要求3所述的方法，其特征在于，所述确定所述源语言训练文本的符合所述当前翻译场景的断句结果，作为目标断句结果，包括：

获取所述源语言训练文本在所述当前翻译场景下的翻译后的目标语言训练文本；

25 参考设定的断句更改方式，对所述源语言训练文本的断句方式进行更改，得到更改后的源语言训练文本，由更改后的源语言训练文本及所述源语言训练文本组成候选源语言训练文本；

利用预置的机器翻译模型，对每一所述候选源语言训练文本进行翻译，

得到每一所述候选源语言训练文本的机器翻译结果；

确定每一所述候选源语言训练文本的机器翻译结果，与所述目标语言训练文本的相似度，将相似度最高的候选源语言训练文本作为所述目标断句结果。

- 5 5、根据权利要求4所述的方法，其特征在于，所述参考设定的断句更改方式，对所述源语言训练文本的断句方式进行更改，得到更改后的源语言训练文本，包括：

确定所述源语言训练文本包含的非终止型标点；

- 10 将所述源语言训练文本包含的每一非终止型标点，使用终止型标点进行替换，得到更改后的源语言训练文本。

6、根据权利要求5所述的方法，其特征在于，所述利用预置的机器翻译模型，对每一所述候选源语言训练文本进行翻译，得到每一所述候选源语言训练文本的机器翻译结果，包括：

- 15 将每一所述候选源语言训练文本按照其包含的终止型标点进行子句划分，得到划分后的子句序列；

利用预置的机器翻译模型，对所述候选源语言训练文本的子句序列中每一子句分别进行翻译，得到每一子句的机器翻译结果；

按照子句序列中各子句的顺序，将各子句的机器翻译结果合并，得到所述候选源语言训练文本的机器翻译结果。

- 20 7、根据权利要求3所述的方法，其特征在于，在所述以所述源语言训练文本作为训练数据，以所述目标断句结果作为训练标签，训练文本断句模型之前，该方法还包括：

获取人工对所述源语言训练文本标注标点的结果，得到人工标注后的源语言训练文本；

- 25 以所述源语言训练文本作为训练数据，以所述人工标注后的源语言训练文本作为训练标签，训练文本断句模型，得到初步文本断句模型；

则，所述以所述源语言训练文本作为训练数据，以所述目标断句结果作为训练标签，训练文本断句模型，包括：

以所述源语言训练文本作为训练数据，以所述目标断句结果作为训练

标签，训练所述初步文本断句模型。

8、根据权利要求1所述的方法，其特征在于，所述对所述断句后的源语言文本进行翻译，得到目标语言文本，包括：

将所述断句后的源语言文本按照其包含的终止型标点进行子句划分，

5 得到划分后的子句序列；

利用预置的机器翻译模型，对所述断句后的源语言文本的子句序列中每一子句分别进行翻译，得到每一子句的机器翻译结果；

按照子句序列中各子句的顺序，将各子句的机器翻译结果合并，得到所述目标语言文本。

10 9、一种翻译装置，其特征在于，包括：

源语言文本获取单元，用于获取待翻译的源语言文本；

文本断句单元，用于根据当前翻译场景对所述源语言文本进行断句，得到断句后的源语言文本；

15 源语言文本翻译单元，用于对所述断句后的源语言文本进行翻译，得到目标语言文本。

10、根据权利要求9所述的装置，其特征在于，所述文本断句单元包括：

模型参考单元，用于将所述源语言文本输入预置的文本断句模型，得到文本断句模型输出的断句后的源语言文本；

20 其中，所述文本断句模型为，以源语言训练文本作为训练数据，以所述源语言训练文本的符合所述当前翻译场景的断句结果作为训练标签训练得到。

11、根据权利要求10所述的装置，其特征在于，还包括：文本断句模型确定单元，用于确定文本断句模型；所述文本断句模型包括：

25 源语言训练文本获取单元，用于获取源语言训练文本；

断句结果确定单元，用于确定所述源语言训练文本的符合所述当前翻译场景的断句结果，作为目标断句结果；

第一模型训练单元，用于以所述源语言训练文本作为训练数据，以所述目标断句结果作为训练标签，训练文本断句模型。

12、根据权利要求 11 所述的装置，其特征在于，所述断句结果确定单元包括：

目标语言训练文本获取单元，用于获取所述源语言训练文本在所述当前翻译场景下的翻译后的目标语言训练文本；

5 断句更改单元，用于参考设定的断句更改方式，对所述源语言训练文本的断句方式进行更改，得到更改后的源语言训练文本，由更改后的源语言训练文本及所述源语言训练文本组成候选源语言训练文本；

源语言训练文本翻译单元，用于利用预置的机器翻译模型，对每一所述候选源语言训练文本进行翻译，得到每一所述候选源语言训练文本的机器翻译结果；

10 相似度确定单元，用于确定每一所述候选源语言训练文本的机器翻译结果，与所述目标语言训练文本的相似度，将相似度最高的候选源语言训练文本作为所述目标断句结果。

13、根据权利要求 12 所述的装置，其特征在于，所述断句更改单元包括：

非终止型标点确定单元，用于确定所述源语言训练文本包含的非终止型标点；

非终止型标点替换单元，用于将所述源语言训练文本包含的每一非终止型标点，使用终止型标点进行替换，得到更改后的源语言训练文本。

20 14、根据权利要求 13 所述的装置，其特征在于，所述源语言训练文本翻译单元包括：

第一子句划分单元，用于将每一所述候选源语言训练文本按照其包含的终止型标点进行子句划分，得到划分后的子句序列；

25 第一子句翻译单元，用于利用预置的机器翻译模型，对所述候选源语言训练文本的子句序列中每一子句分别进行翻译，得到每一子句的机器翻译结果；

第一翻译结果合并单元，用于按照子句序列中各子句的顺序，将各子句的机器翻译结果合并，得到所述候选源语言训练文本的机器翻译结果。

15、根据权利要求 11 所述的装置，其特征在于，所述文本断句模型还

包括:

人工标注结果获取单元, 用于获取人工对所述源语言训练文本标注标点
的结果, 得到人工标注后的源语言训练文本;

5 第二模型训练单元, 用于以所述源语言训练文本作为训练数据, 以所
述人工标注后的源语言训练文本作为训练标签, 训练文本断句模型, 得到
初步文本断句模型;

则所述第一模型训练单元具体用于:

以所述源语言训练文本作为训练数据, 以所述目标断句结果作为训练
标签, 训练所述初步文本断句模型。

10 16、根据权利要求 9 所述的装置, 其特征在于, 所述源语言文本翻译
单元, 包括:

第二子句划分单元, 用于将所述断句后的源语言文本按照其包含的终
止型标点进行子句划分, 得到划分后的子句序列;

15 第二子句翻译单元, 用于利用预置的机器翻译模型, 对所述断句后的
源语言文本的子句序列中每一子句分别进行翻译, 得到每一子句的机器翻
译结果;

第二翻译结果合并单元, 用于按照子句序列中各子句的顺序, 将各子
句的机器翻译结果合并, 得到所述目标语言文本。

20 17、一种翻译设备, 其特征在于, 包括存储器和处理器;
所述存储器, 用于存储程序;

所述处理器, 用于执行所述程序, 实现如权利要求 1-8 中任一项所述
的翻译方法的各个步骤。

25 18、一种可读存储介质, 其上存储有计算机程序, 其特征在于, 所述
计算机程序被处理器执行时, 实现如权利要求 1-8 中任一项所述的翻译方
法的各个步骤。

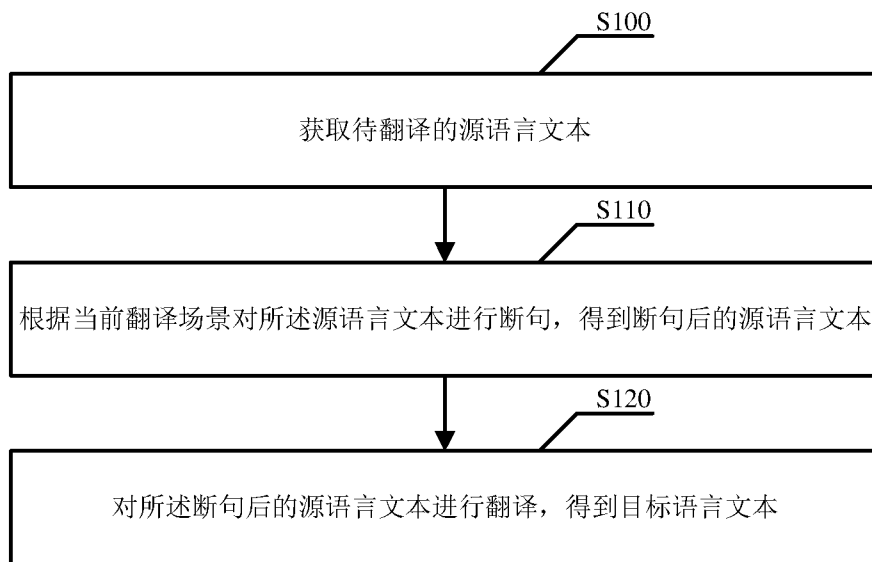


图 1

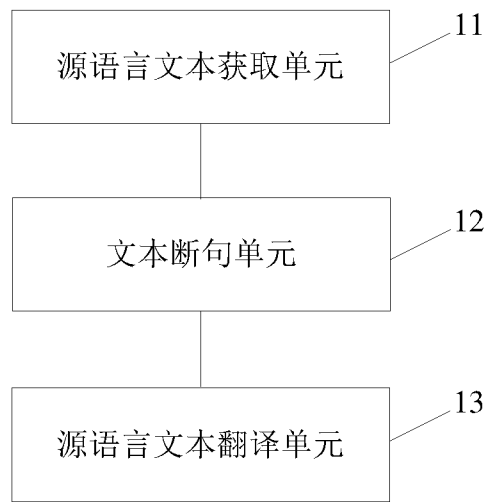


图 2

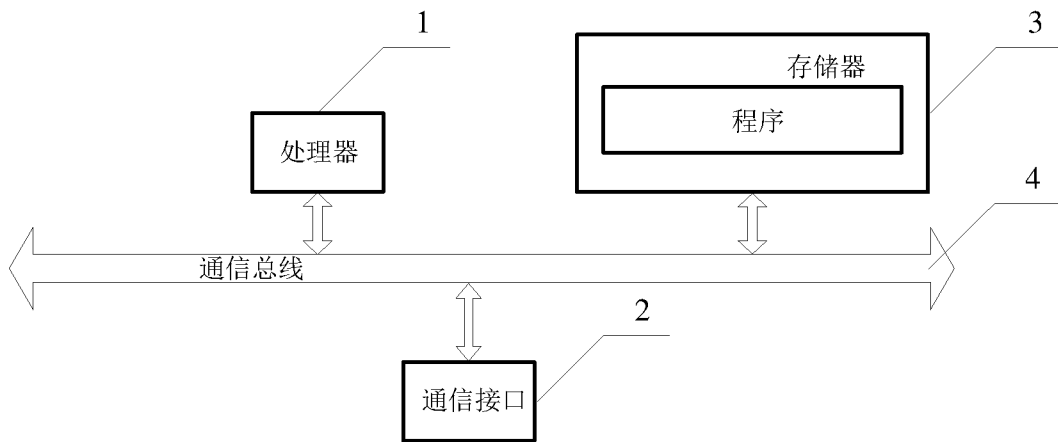


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/119329

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/28(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)

CNABS, CNTXT, VEN, USTXT, EPTXT, WOTXT, CNKI; 翻译, 断句, 句子, 文本, 分割, 划分, 标点, 替换, 代替, 模型, translation, sentence break, sentence, text, segment, divide, punctuation, replace, model

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 103530284 A (CHINA PATENT INFORMATION CENTER) 22 January 2014 (2014-01-22) description, paragraphs 44-80, and figures 1, 3 and 5-6	1-18
Y	CN 107247706 A (CHINA ELECTRONICS STANDARDIZATION INSTITUTE ET AL.) 13 October 2017 (2017-10-13) description, paragraphs 14-21, and figure 1	1-18
Y	CN 108628819 A (BEIJING SOGOU TECHNOLOGY DEVELOPMENT CO., LTD.) 09 October 2018 (2018-10-09) description, paragraphs 74-92 and 103, and figures 1-3	4-6, 12-14, 17-18
A	CN 101458681 A (TOSHIBA CORPORATION) 17 June 2009 (2009-06-17) entire document	1-18
A	US 2003055626 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 20 March 2003 (2003-03-20) entire document	1-18

 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

04 July 2019

Date of mailing of the international search report

31 July 2019

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

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/119329

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	103530284	A	22 January 2014	CN	103530284	B	06 July 2016
CN	107247706	A	13 October 2017	None			
CN	108628819	A	09 October 2018	None			
CN	101458681	A	17 June 2009	US	2009150139	A1	11 June 2009
				JP	2009140503	A	25 June 2009
US	2003055626	A1	20 March 2003	JP	3557605	B2	25 August 2004
				JP	2003108550	A	11 April 2003
				US	7136806	B2	14 November 2006

国际检索报告

国际申请号

PCT/CN2018/119329

<p>A. 主题的分类</p> <p>G06F 17/28 (2006.01) i</p> <p>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</p>																				
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>G06F</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>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))</p> <p>CNABS, CNTXT, VEN, USTXT, EPTXT, WOTXT, CNKI; 翻译, 断句, 句子, 文本, 分割, 划分, 标点, 替换, 代替, 模型, translation, sentence break, sentence, text, segment, divide, punctuation, replace, model</p>																				
<p>C. 相关文件</p> <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>Y</td> <td>CN 103530284 A (中国专利信息中心) 2014年 1月 22日 (2014 - 01 - 22) 说明书第44-80段, 图1、3、5-6</td> <td>1-18</td> </tr> <tr> <td>Y</td> <td>CN 107247706 A (中国电子技术标准化研究院等) 2017年 10月 13日 (2017 - 10 - 13) 说明书第14-21段, 图1</td> <td>1-18</td> </tr> <tr> <td>Y</td> <td>CN 108628819 A (北京搜狗科技发展有限公司) 2018年 10月 9日 (2018 - 10 - 09) 说明书74-92、103段, 图1-3</td> <td>4-6、12-14、17-18</td> </tr> <tr> <td>A</td> <td>CN 101458681 A (株式会社东芝) 2009年 6月 17日 (2009 - 06 - 17) 全文</td> <td>1-18</td> </tr> <tr> <td>A</td> <td>US 2003055626 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2003年 3月 20日 (2003 - 03 - 20) 全文</td> <td>1-18</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	Y	CN 103530284 A (中国专利信息中心) 2014年 1月 22日 (2014 - 01 - 22) 说明书第44-80段, 图1、3、5-6	1-18	Y	CN 107247706 A (中国电子技术标准化研究院等) 2017年 10月 13日 (2017 - 10 - 13) 说明书第14-21段, 图1	1-18	Y	CN 108628819 A (北京搜狗科技发展有限公司) 2018年 10月 9日 (2018 - 10 - 09) 说明书74-92、103段, 图1-3	4-6、12-14、17-18	A	CN 101458681 A (株式会社东芝) 2009年 6月 17日 (2009 - 06 - 17) 全文	1-18	A	US 2003055626 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2003年 3月 20日 (2003 - 03 - 20) 全文	1-18
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
Y	CN 103530284 A (中国专利信息中心) 2014年 1月 22日 (2014 - 01 - 22) 说明书第44-80段, 图1、3、5-6	1-18																		
Y	CN 107247706 A (中国电子技术标准化研究院等) 2017年 10月 13日 (2017 - 10 - 13) 说明书第14-21段, 图1	1-18																		
Y	CN 108628819 A (北京搜狗科技发展有限公司) 2018年 10月 9日 (2018 - 10 - 09) 说明书74-92、103段, 图1-3	4-6、12-14、17-18																		
A	CN 101458681 A (株式会社东芝) 2009年 6月 17日 (2009 - 06 - 17) 全文	1-18																		
A	US 2003055626 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2003年 3月 20日 (2003 - 03 - 20) 全文	1-18																		
<p><input type="checkbox"/> 其余文件在C栏的续页中列出。</p> <p><input checked="" type="checkbox"/> 见同族专利附件。</p>																				
<p>* 引用文件的具体类型:</p> <p>“A” 认为不特别相关的表示了现有技术一般状态的文件</p> <p>“E” 在国际申请日的当天或之后公布的在先申请或专利</p> <p>“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)</p> <p>“O” 涉及口头公开、使用、展览或其他方式公开的文件</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“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</p> <p>“&” 同族专利的文件</p>																				
<p>国际检索实际完成的日期</p> <p>2019年 7月 4日</p>		<p>国际检索报告邮寄日期</p> <p>2019年 7月 31日</p>																		
<p>ISA/CN的名称和邮寄地址</p> <p>中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088</p> <p>传真号 (86-10)62019451</p>		<p>受权官员</p> <p>马婷婷</p> <p>电话号码 86-(20)-28958507</p>																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/119329

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	103530284	A	2014年 1月 22日	CN	103530284	B	2016年 7月 6日
CN	107247706	A	2017年 10月 13日	无			
CN	108628819	A	2018年 10月 9日	无			
CN	101458681	A	2009年 6月 17日	US	2009150139	A1	2009年 6月 11日
				JP	2009140503	A	2009年 6月 25日
US	2003055626	A1	2003年 3月 20日	JP	3557605	B2	2004年 8月 25日
				JP	2003108550	A	2003年 4月 11日
				US	7136806	B2	2006年 11月 14日