

(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
2018 年 10 月 4 日 (04.10.2018)



(10) 国际公布号
WO 2018/176758 A1

- (51) 国际专利分类号:
G06F 17/24 (2006.01)
- (21) 国际申请号: PCT/CN2017/102620
- (22) 国际申请日: 2017 年 9 月 21 日 (21.09.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201710206961.1 2017年3月31日 (31.03.2017) CN
- (71) 申请人: 百度在线网络技术(北京)有限公司 (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) [CN/CN]; 中国北京市海淀区上地十街10号百度大厦三层, Beijing 100085 (CN)。
- (72) 发明人: 王文斌(WANG, Wenbin); 中国北京市海淀区上地十街10号百度大厦三层, Beijing 100085 (CN)。 施鹏(SHI, Peng); 中国北京市海淀区上地十街10号百度大厦三层, Beijing 100085 (CN)。 吴广发(WU, Guangfa); 中国北京市海淀区上地十街10号百度大厦三层, Beijing 100085 (CN)。
- (74) 代理人: 北京英赛嘉华知识产权代理有限公司 (INSIGHT INTELLECTUAL PROPERTY LIMITED); 中国北京市海淀区知春路甲48号盈都大厦A座19A, Beijing 100098 (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, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK,

(54) Title: METHOD AND DEVICE FOR GENERATING ARTICLE

(54) 发明名称: 用于生成文章的方法和装置

100

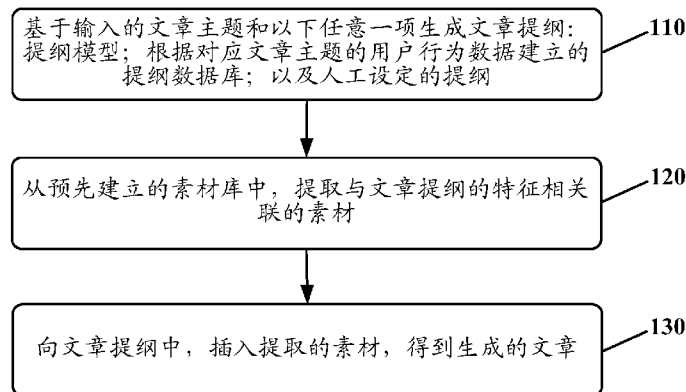


图 1

- 110 Generate an article outline on the basis of an input article topic and any one of an outline model, an outline database established according to user behavior data of a corresponding article topic, and a manually set outline
- 120 Extract, from a pre-established material library, a material associated with the feature of the article outline
- 130 Insert the extracted material into the article outline to obtain a generated article

(57) Abstract: Disclosed are a method and device for generating an article. A specific embodiment of the method comprises: generating an article outline on the basis of an input article topic and any one of an outline model, an outline database established according to user behavior data of a corresponding article topic, and a manually set outline; extracting, from a pre-established material library, a material associated with the feature of the article outline; and inserting the extracted material into the article outline to obtain a generated article. By means of the embodiment, an outline can be generated according to an input article topic; the quality of the article outline is



WO 2018/176758 A1

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))。

improved; reasonable writing logic and rich form of the generated article are ensured; by inserting a material associated with the feature of the article outline according to the article outline, the content of the article is enriched, so that the generated article has reasonable logic and rich form and content.

(57) 摘要: 本申请公开了用于生成文章的方法和装置。方法的一具体实施方式包括: 基于输入的文章主题和以下任意一项生成文章提纲: 提纲模型; 根据对应文章主题的用户行为数据建立的提纲数据库; 以及人工设定的提纲; 从预先建立的素材库中, 提取与文章提纲的特征相关联的素材; 向文章提纲中, 插入提取的素材, 得到生成的文章。该实施方式实现了根据输入的文章主题生成提纲, 提高了文章提纲的质量, 保证了生成的文章的行文逻辑合理、形式丰富, 并根据文章提纲插入与文章提纲的特征相关联的素材, 丰富了文章的内容, 从而使得生成的文章逻辑合理并且形式、内容丰富。

用于生成文章的方法和装置

5 相关申请的交叉引用

本申请要求于 2017 年 3 月 31 日提交的中国专利申请号为“201710206961.1”的优先权，其全部内容作为整体并入本申请中。

技术领域

10 本申请涉及计算机技术领域，具体涉及计算机网络技术领域，尤其涉及用于生成文章的方法和装置。

背景技术

15 目前，通过机器实现自动化写作，从而生成文章的方式，基本停留在特殊领域的特殊题材，多是采用将素材填入规则或模板的技术来生成文章。例如，可以筛选原始文章后直接引用；或者，对原始文章进行简单的变换直接发布；或者，将原始文章以一定的顺序进行组合和摘要提取；又或者，通过模板对数据进行组织并展现。

20 然而，目前的生成文章的方式，由于题材和方法的限制，产出的文章形式和内容比较单调，并且行文可能出现前后逻辑不合理、文法风格不一致等情况，机器写作的痕迹较重。

发明内容

25 本申请的目的在于提出一种改进的用于生成文章的方法和装置，来解决以上背景技术部分提到的技术问题。

第一方面，本申请实施例提供了一种用于生成文章的方法，方法包括：基于输入的文章主题和以下任意一项生成文章提纲：提纲模型，根据对应文章主题的用户行为数据建立的提纲数据库，以及人工设定的提纲；从预先建立的素材库中，提取与文章提纲的特征相关联的素材；向文章提纲中，插入提取的素材，得到生成的文章。

30

在一些实施例中，根据对应文章主题的用户行为数据建立的提纲数据库包括：检索全网围绕文章主题的子主题，建立子主题数据库；根据用户对子主题数据库中的子主题的点击顺序和/或子主题数据库中的子主题的语义递进顺序，排序子主题数据库中的子主题；剔除子主题数据库中不符合预定逻辑规则的子主题，得到符合预定逻辑规则
5 的子主题；将各符合预定逻辑规则的子主题作为提纲，得到提纲数据库。

在一些实施例中，预先建立的素材库通过以下步骤建立：获取素材的特征，素材为将现有的文章的内容根据筛选规则筛选得到和/或变换现有的文章的内容得到；根据素材的特征建立索引结构，得到素材
10 库。

在一些实施例中，方法还包括：对生成的文章进行优化处理，得到优化后的生成的文章，优化处理包括以下一项或多项：润色处理、插入富媒体数据处理以及排版优化处理。

在一些实施例中，润色处理包括以下一项或多项：统一生成的文章的文法风格；删除与前后语句不连贯的语句；以及替换与前后语句不连贯的语句。
15

在一些实施例中，插入富媒体数据处理包括：从预先建立的资源库，提取与生成的文章的特征相关联的富媒体数据；向生成的文章中，
20 插入提取的富媒体数据。

在一些实施例中，从预先建立的资源库，提取与生成的文章的特征相关联的富媒体数据包括：根据以下一项或多项从预先建立的资源库中提取富媒体数据生成候选富媒体列表：文章主题、文章提纲、生成的文章的各段落的摘要以及生成的文章的各段落的关键词；采用质量
25 筛选从候选富媒体列表中提取与生成的文章的特征相关联的富媒体数据。

在一些实施例中，预先建立的资源库通过以下步骤建立：获取富媒体数据的特征；根据富媒体数据的特征建立索引结构，得到资源库。

在一些实施例中，质量筛选根据以下一项或多项进行：图文相关性、图片分辨率、图片长宽比、图片来源权威度、广告过滤策略、反
30

作弊过滤策略、反黄过滤策略和水印过滤策略。

在一些实施例中，方法还包括：将文章主题和文章提纲输入标题模型，得到生成的文章的标题。

在一些实施例中，方法还包括：对标题中的核心词进行属性扩展；

5 对属性扩展后的标题中的核心词进行替换和改写，得到更新后的标题。

第二方面，本申请实施例提供了一种用于生成文章的装置，装置包括：提纲生成单元，用于基于输入的文章主题和以下任意一项生成文章提纲：提纲模型，根据对应文章主题的用户行为数据建立的提纲数据库，以及人工设定的提纲；素材提取单元，用于从预先建立的素
10 材库中，提取与文章提纲的特征相关联的素材；素材插入单元，用于向文章提纲中，插入提取的素材，得到生成的文章。

在一些实施例中，提纲生成单元中的根据对应文章主题的用户行为数据建立的提纲数据库包括：检索全网围绕文章主题的子主题，建立子主题数据库；根据用户对子主题数据库中的子主题的点击顺序和/
15 或子主题数据库中的子主题的语义递进顺序，排序子主题数据库中的子主题；剔除子主题数据库中不符合预定逻辑规则的子主题，得到符合预定逻辑规则的子主题；将各符合预定逻辑规则的子主题作为提纲，得到提纲数据库。

在一些实施例中，素材提取单元中的预先建立的素材库通过以下步骤建立：获取素材的特征，素材为将现有的文章的内容根据筛选规则筛选得到和/或变换现有的文章的内容得到；根据素材的特征建立索引结构，得到素材库。
20

在一些实施例中，装置还包括：文章优化单元，用于对生成的文章进行优化处理，得到优化后的生成的文章，优化处理包括以下一项或多项：润色处理、插入富媒体数据处理以及排版优化处理。
25

在一些实施例中，文章优化单元中的润色处理包括以下一项或多项：统一生成的文章的文法风格；删除与前后语句不连贯的语句；以及替换与前后语句不连贯的语句。

在一些实施例中，文章优化单元中的插入富媒体数据处理包括：
30 从预先建立的资源库，提取与生成的文章的特征相关联的富媒体数据；

向生成的文章中，插入提取的富媒体数据。

在一些实施例中，文章优化单元中的从预先建立的资源库，提取与生成的文章的特征相关联的富媒体数据包括：根据以下一项或多项从预先建立的资源库中提取富媒体数据生成候选富媒体列表：文章主题、文章提纲、生成的文章的各段落的摘要以及生成的文章的各段落的关键词；采用质量筛选从候选富媒体列表中提取与生成的文章的特征相关联的富媒体数据。

在一些实施例中，文章优化单元中的预先建立的资源库通过以下步骤建立：获取富媒体数据的特征；根据富媒体数据的特征建立索引结构，得到资源库。

在一些实施例中，文章优化单元中的质量筛选根据以下一项或多项进行：图文相关性、图片分辨率、图片长宽比、图片来源权威度、广告过滤策略、反作弊过滤策略、反黄过滤策略和水印过滤策略。

在一些实施例中，装置还包括：标题生成单元，用于将文章主题和文章提纲输入标题模型，得到生成的文章的标题。

在一些实施例中，装置还包括：属性扩展单元，用于对标题中的核心词进行属性扩展；标题更新单元，用于对属性扩展后的标题中的核心词进行替换和改写，得到更新后的标题。

第三方面，本申请实施例提供了一种设备，包括：一个或多个处理器；存储装置，用于存储一个或多个程序；当一个或多个程序被一个或多个处理器执行，使得一个或多个处理器实现如上任一所述的用于生成文章的方法。

第四方面，本申请实施例提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现如上任一所述的用于生成文章的方法。

本申请实施例提供的用于生成文章的方法和装置，首先基于输入的文章主题和以下任意一项生成文章提纲：提纲模型；根据对应文章主题的用户行为数据建立的提纲数据库；以及人工设定的提纲；之后，从预先建立的素材库中，提取与文章提纲的特征相关联的素材；之后，向文章提纲中，插入提取的素材，得到生成的文章。本实施例实现了

根据输入的文章主题生成提纲，提高了文章提纲的质量，保证了生成的文章的行文逻辑合理、形式丰富，并根据文章提纲插入与文章提纲的特征相关联的素材，丰富了文章的内容，从而使得生成的文章逻辑合理并且内容丰富。

5

附图说明

通过阅读参照以下附图所作的对非限制性实施例所作的详细描述，本申请的其它特征、目的和优点将会变得更明显：

图 1 是根据本申请的用于生成文章的方法的一个实施例的示意性流程图；

图 2 是根据本申请的用于生成文章的方法的又一个实施例的示意性流程图；

图 3 是应用本申请的用于生成文章的方法的一个实施例的示例性应用场景；

图 4 是根据本申请的用于生成文章的装置的一个实施例的示例性结构图；

图 5 是适于用来实现本申请实施例的终端设备或服务器的计算机系统的结构示意图。

具体实施方式

下面结合附图和实施例对本申请作进一步的详细说明。可以理解的是，此处所描述的具体实施例仅仅用于解释相关发明，而非对该发明的限定。另外还需要说明的是，为了便于描述，附图中仅示出了与有关发明相关的部分。

需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。下面将参考附图并结合实施例来详细说明本申请。

图 1 示出了根据本申请的用于生成文章的方法的一个实施例的流程 100。该用于生成文章的方法包括：

在步骤 110 中，基于输入的文章主题和以下任意一项生成文章提

纲: 提纲模型; 根据对应文章主题的用户行为数据建立的提纲数据库; 以及人工设定的提纲。

在本实施例中, 输入的文章主题可以为机器挖掘或人工输入的文章主题。

5 提纲模型通常是指以文章主题为自变量的函数。首先, 可以设定文章模型 $=f(\text{主题}, \text{提纲}, \text{素材})$, 也即文章模型由函数 f 中的自变量(主题、提纲和素材)得到, 并借由该文章模型, 可以得到一种用于生成文章的方法, 即选定主题, 通过提纲模型挖掘提纲并排序, 通过素材库来挂载素材; 最后通过配图、排版和润色得到文章。

10 根据对应文章主题的用户行为数据建立的提纲数据库是指从文章主题角度确定文章目录, 并根据用户行为数据对文章目录进行合理排序和筛选, 得到提纲数据库。应当理解, 这里的提纲生成策略生成的提纲具有一定的逻辑顺序, 以保障行文的合理性。

15 在本实施例的一些可选实现方式中, 上述的根据对应所述文章主题的用户行为数据建立的提纲数据库包括: 检索全网围绕文章主题的子主题, 建立子主题数据库; 根据用户对子主题数据库中的子主题的点击顺序和/或子主题数据库中的子主题的语义递进顺序, 排序子主题数据库中的子主题; 剔除子主题数据库中不符合预定逻辑规则的子主题, 得到符合预定逻辑规则的子主题; 将各符合预定逻辑规则的子主题
20 题作为提纲, 得到提纲数据库。

在本实现方式中, 根据对应所述文章主题的用户行为数据建立的提纲数据库, 充分考虑了用户的行为数据来建立提纲, 可以提高建立的提纲的针对性, 进而增强了生成的文章与用户的交互能力。

25 在步骤 120 中, 从预先建立的素材库中, 提取与文章提纲的特征相关联的素材。

在本实施例中, 预先建立的素材库, 是指根据素材的特征建立索引结构得到的素材库。当素材的特征与文章提纲的特征相关联时, 可以提取该素材以备后续使用。当多个素材的特征均与文章提纲的特征相关联时, 可以从多个素材中, 提取特征与文章提纲的特征最为相关的
30 的预定数量个素材, 以备后续使用。

在本实施例的一些可选实现方式中，预先建立的素材库通过以下步骤建立：获取素材的特征，素材为将现有的文章的内容根据筛选规则筛选得到和/或变换现有的文章的内容得到；根据素材的特征建立索引结构，得到素材库。

5 在本实现方式中，素材库的生成包括有明确主题的素材和无明确主题的素材，后者需要使用文章摘要技术提取主题。获取素材的特征，可以理解为从文本素材中提取特征，这些特征可以说明文本素材的主题、关键词、核心语义等信息，用于和文章提纲、文章主题进行相关性计算和排序。

10 具体地，上述的根据筛选规则筛选得到可以包括根据以下一项或多项内容进行筛选：文章的内容长度、文章的内容质量评分、文章的内容满意度评分、文章的浏览量、文章的时效性等等。而上述的变换现有的文章内容主要是为了控制素材的粒度，可以采用预定规则来完成变换。例如，将字数大于预定值的段落进行拆解分段。假设一个素
15 材是一篇原始语料，筛选后根据提纲排序组合就可以了；假设一个素材是一段，就需要考虑段落的主题相关性、段落间排序等；同理，还可以假设素材是一句话、一个词，当素材的粒度越小时，拆解和/或变换的难度越大。

20 在步骤 130 中，向文章提纲中，插入提取的素材，得到生成的文章。

在本实施例中，可以向步骤 110 中得到的文章提纲中，插入步骤 120 中提取的素材，从而得到生成的文章。

25 本申请的上述实施例提供的用于生成文章的方法，通过生成文章提纲，提取与文章提纲的特征相关联的素材，插入提取的素材，得到生成的文章，可以根据输入的文章主题生成文章提纲，并且插入文章提纲的素材极为丰富，因此生成的文章的逻辑顺序合理、形式和内容更为丰富，接近于专业人士写的文章，从而摒弃了目前机器写作的局限性。

30 进一步参考图 2，图 2 示出了根据本申请的用于生成文章的方法的又一个实施例的示意性流程图。该用于生成文章的方法 200 包括：

在步骤 210 中，基于输入的文章主题和以下任意一项生成文章提纲：提纲模型；根据对应文章主题的用户行为数据建立的提纲数据库；以及人工设定的提纲。

在本实施例中，在本实施例中，输入的文章主题可以为机器挖掘或人工输入的文章主题。

提纲模型通常是指以文章主题为自变量的函数。首先，可以设定文章模型 $=f(\text{主题}, \text{提纲}, \text{素材})$ ，也即文章模型由函数 f 中的自变量（主题、提纲和素材）得到，并借由该文章模型，可以得到一种用于生成文章的方法，即选定主题，通过提纲模型挖掘提纲并排序，通过素材库来挂载素材；最后通过配图、排版和润色得到文章。

根据对应文章主题的用户行为数据建立的提纲数据库是指从文章主题角度确定文章目录，并根据用户行为数据对文章目录进行合理排序和筛选，得到提纲数据库。应当理解，这里的提纲生成策略生成的提纲具有一定的逻辑顺序，以保障行文的合理性。

在步骤 220 中，从预先建立的素材库中，提取与文章提纲的特征相关联的素材。

在本实施例中，预先建立的素材库，是指根据素材的特征建立索引结构得到的素材库。当素材的特征与文章提纲的特征相关联时，可以提取该素材以备后续使用。当多个素材的特征均与文章提纲的特征相关联时，可以从多个素材中，提取特征与文章提纲的特征最为相关的预定数量个素材，以备后续使用。

在步骤 230 中，向文章提纲中，插入提取的素材，得到生成的文章。

在本实施例中，可以向步骤 210 中得到的文章提纲中，插入步骤 220 中提取的素材，从而得到初具雏形的生成的文章。

在步骤 240 中，对生成的文章进行优化处理，得到优化后的生成的文章。

在本实施例中，优化处理包括以下一项或多项：润色处理、插入富媒体数据处理以及排版优化处理。

对于生成的文章，由于素材库中存在不同的文法风格的素材，并

且前后连接处可能并不连贯，因此可以对生成的文章进行润色处理，也即对文章的文法风格和语句等进行处理。这里的文法，即文章的书写法规，一般用来指以文字、词语、短句、句子的编排而组成的完整语句和文章的合理性组织。这里的风格，是指具有独特于其他文章的表现，带有综合性的总体特点。

在本实施例的一些可选实现方式中，进行润色处理包括以下一项或多项：统一生成的文章的文法风格；删除与前后语句不连贯的语句；以及替换与前后语句不连贯的语句。

在本实现方式中，统一生成的文章的文法风格，可以通过对于特定词汇、特定句式的替换和变换实现，从而使得文章的文法风格一致。而删除与前后语句不连贯的语句，或者替换与前后语句不连贯的语句，均可改善语句的不连贯现象。

在本实施例的一些可选实现方式中，插入富媒体数据处理包括：从预先建立的资源库，提取与生成的文章的特征相关联的富媒体数据，向生成的文章中，插入提取的富媒体数据。

在本实施例中，向生成的文章中，插入提取的富媒体数据包括：首先根据主题、提纲、段落摘要和关键词中的一项或多项查找富媒体数据，之后通过质量筛选挑选出优质富媒体数据库，并根据图片间字数或段落数，保证插入的富媒体数据相对均匀。例如，若文章中有两张图之间 1000 字，而另外两个图间 10 个字，那么插入的富媒体数据不均匀，并不符合用户群体的阅读习惯。富媒体数据为可以包含流媒体、声音、Flash、以及 Java、Javascript、动态的 HTML 等程序设计语言的形式之一或者几种的组合。富媒体数据可以应用于各种网络服务中，如网站设计、电子邮件、网站页面的横幅、按钮、弹出式广告、插播式广告等。应当理解，富媒体数据可以加强信息，而信息更准确的定向会具有更好的交互效果。

在本实施例的一些可选实现方式中，从预先建立的资源库，提取与润色后的文章的特征相关联的富媒体数据包括：根据以下一项或多项从预先建立的资源库中提取富媒体数据生成候选富媒体列表：文章主题、文章提纲、润色后的文章的各段落的摘要以及润色后的文章的

各段落的关键词；采用质量筛选从候选富媒体列表中提取与润色后的文章的特征相关联的富媒体数据。

在本实现方式中，通过根据文章主题、文章提纲、润色后的文章
的各段落的摘要以及润色后的文章各段落的关键词中的一项或多项
5 提取富媒体数据，生成富媒体列表；之后采用质量筛选从富媒体列表
中提取与润色后的文章的特征相关联的富媒体数据，可以提高资源库
中的富媒体数据的质量。

在本实施例的一些可选实现方式中，预先建立的资源库可以通过
以下步骤建立：获取富媒体数据的特征；根据富媒体数据的特征建立
10 索引结构，得到资源库。

在本实施例的一些可选实现方式中，上述的质量筛选可以根据以
下一项或多项进行：图文相关性、图片分辨率、图片长宽比、图片来
源权威度、广告过滤策略、反作弊过滤策略、反黄过滤策略和水印过
滤策略。

15 在本实现方式中，广告过滤策略可以包括广告过滤规则和广告过
滤模型；反作弊过滤策略可以包括反作弊过滤规则和反作弊过滤模型；
反黄过滤策略可以包括反黄过滤规则和反黄过滤模型；水印过滤策略
则可以包括水印过滤策略和水印过滤模型。

在本实施例中，排版优化处理可以采用现有技术或未来发展的技
20 术中的排版优化方法来完成，本申请对此不做限定。例如，排版优化
处理可以为在确定各种需要呈现的文章内容之后，选择需要重点突出
的内容，最后搭配恰当的颜色版式，从而得到优化后的文章。这里的
排版优化处理，也可以根据对文章样本数据和用户针对文章样本数据
的行为数据的分析结果来确定与生成的文章相适应的排版，从而得到
25 优化后的文章。

在步骤 250 中，将文章主题和文章提纲输入标题模型，生成文章
的标题。

在本实施例中，在得到生成的文章之后，可以将文章主题和文章
提纲输入标题模型，以便生成文章的主题。这里的标题模型，是自变
30 量为文章主题和文章提纲的函数，当接收到文章主题和文章提纲时，

根据该函数即可输出文章的主题。例如，可以为机器根据现有的文章样本中包括的文章主题、文章提纲和文章的标题学习得到的标题模型，也可以为人为设定的标题模型。

在本实施例的一些可选实现方式中，方法还包括：对标题中的核心词进行属性扩展；对属性扩展后的标题中的核心词进行替换和改写，得到更新后的标题。

在本实现方式中，可以首先挖掘标题中的核心词，之后对核心词进行属性扩展，再对属性扩展后的标题中的核心词进行替换和改写，得到更新后的标题。例如，对于皇帝 XXX 的介绍，挖掘出标题中的核心词为 XXX，之后可以得到 XXX 的属性是放牛娃出身的皇帝，因此可以将皇帝 XXX 的介绍替换和改写为：放牛娃出身的皇帝是谁？

应当理解，上述图 2 中的描述仅为本申请实施例的用于生成文章的方法的一个示例性描述，并不代表对本申请的限定。例如，本申请实施例中的用于生成文章的方法，也可以不包括上述步骤 240，或者不包括上述步骤 250，从而得到新的用于生成文章的方法。图 2 中的步骤 210、步骤 220 和步骤 230 分别与图 1 中的步骤 110、步骤 120 和步骤 130 相对应，因此，图 1 中针对步骤 110、步骤 120 和步骤 130 描述的操作和特征同样适用于步骤 210、步骤 220 和步骤 230，在此不再赘述。

本申请的上述实施例提供的用于生成文章的方法，通过与图 1 中描述的用于生成文章的方法相比，增加了步骤 240 和步骤 250，根据步骤 240 和步骤 250，可以得到优化后的生成的文章以及得到生成的文章的标题，从而使得生成的文章的内容更为全面，包含的信息更为丰富，文章的标题更具有吸引力，并且文章的内容和标题更为适应用户群体的阅读习惯。

以下结合图 3，描述本申请实施例的用于生成文章的方法的一个示例性应用场景。

如图 3 所示，根据本申请实施例的用于生成文章的方法，首先，根据输入的文章主题 310 的具体实施例 311 “诸葛亮 称帝”，可以生成文章提纲 320 的具体实施例，也即包括提纲 321：刘备托孤时为什

么让诸葛亮称帝；提纲 322：诸葛亮为什么不称帝；以及提纲 323：诸葛亮如果称帝会怎么样。之后，从预先建立的素材库中，提取与文章提纲 321 至 323 的特征相关联的素材 330，包括以下素材：素材 331 “政权问题”、素材 332 “欲擒故纵”、素材 333 “明智决定”、素材 334 “文人是造不了反的”、素材 335 “集团外部的阻力”、素材 336 “集团内部的阻力”、素材 337 “外部方面的阻力”、素材 338 “兵民厌战”以及素材 339 “最关键的一点”。之后，向文章提纲中，插入提取的素材 330（包括素材 331-339），得到生成的文章。之后，对生成的文章进行润色 340，具体包括在步骤 341 中，统一文章的文风，以及在步骤 342 中，连贯语句，得到润色后的文章。然后，从预先建立的资源库，提取与润色后的文章的特征相关联的富媒体 350，包括标号为 351 的图片 1、标号为 352 的图片 2 以及标号为 353 的图片 3。之后，向润色后的文章中，插入提取的富媒体 350（包括富媒体 351-353），得到插入富媒体后的文章；之后，在标题 360 的生成步骤，将文章主题和文章提纲输入标题模型，得到初始标题，并对初始标题中的核心词进行属性扩展，对属性扩展后的初始标题中的核心词进行替换和改写，得到更新后的标题 361 “有颜有实力，集尽万千追捧的男神为何终未加冕？”。之后，在排版 370 的处理步骤中，对插入富媒体后的文章进行排版优化处理，例如进行具体操作 371，突出重点，并进行颜色版式调整，从而得到排版优化后的文章。最后，在输出 380 的处理步骤中，可以具体进行操作 381，输出排版优化后的文章。

本申请的上述应用场景中提供的用于生成文章的方法，提高了文章的生成效率，并丰富了文章的内容，使得生成的文章的行文与现有技术相比，前后逻辑、文法风格一致，形式、内容更为丰富且更为合理。

进一步参考图 4，作为对上述方法的实现，本申请实施例提供了一种用于生成文章的装置的一个实施例，该用于生成文章的方法的实施例与图 1 至图 3 所示的用于生成文章的方法的实施例相对应，由此，上文针对图 1 至图 3 中用于生成文章的方法描述的操作和特征同样适用于用于生成文章的装置 400 及其中包含的单元，在此不再赘述。

如图 4 所示, 该配置用于生成文章的装置 400 包括: 装置包括:
提纲生成单元 410, 用于基于输入的文章主题和以下任意一项生成文
章提纲: 提纲模型; 根据对应文章主题的用户行为数据建立的提纲数
据库; 以及人工设定的提纲; 素材提取单元 420, 用于从预先建立的
5 素材库中, 提取与文章提纲的特征相关联的素材; 素材插入单元 430,
用于向文章提纲中, 插入提取的素材, 得到生成的文章。

在一些实施例中, 提纲生成单元中的根据对应文章主题的用户行
为数据建立的提纲数据库包括: 检索全网围绕文章主题的子主题, 建
立子主题数据库; 根据用户对子主题数据库中的子主题的点击顺序和/
10 或子主题数据库中的子主题的语义递进顺序, 排序子主题数据库中的
子主题; 剔除子主题数据库中不符合预定逻辑规则的子主题, 得到符
合预定逻辑规则的子主题; 将各符合预定逻辑规则的子主题作为提纲,
得到提纲数据库。

在一些实施例中, 素材提取单元中的预先建立的素材库通过以下
15 步骤建立: 获取素材的特征, 素材为将现有的文章的内容根据筛选规
则筛选得到和/或变换现有的文章的内容得到; 根据素材的特征建立索
引结构, 得到素材库。

在一些实施例中, 装置还包括: 文章优化单元 440, 用于对生成
的文章进行优化处理, 得到优化后的生成的文章, 优化处理包括以下
20 一项或多项: 润色处理、插入富媒体数据处理以及排版优化处理。

在一些实施例中, 文章优化单元中的润色处理包括以下一项或多
项: 统一生成的文章的文法风格; 删除与前后语句不连贯的语句; 以
及替换与前后语句不连贯的语句。

在一些实施例中, 文章优化单元中的插入富媒体数据处理包括:
25 从预先建立的资源库, 提取与生成的文章的特征相关联的富媒体数据;
向生成的文章中, 插入提取的富媒体数据。

在一些实施例中, 文章优化单元中的从预先建立的资源库, 提取
与生成的文章的特征相关联的富媒体数据包括: 根据以下一项或多项
从预先建立的资源库中提取富媒体数据生成候选富媒体列表: 文章主
30 题、文章提纲、生成的文章的各段落的摘要以及生成的文章的各段落

的关键词；采用质量筛选从候选富媒体列表中提取与生成的文章的特征相关联的富媒体数据。

在一些实施例中，文章优化单元中的预先建立的资源库通过以下步骤建立：获取富媒体数据的特征；根据富媒体数据的特征建立索引结构，得到资源库。

在一些实施例中，文章优化单元中的质量筛选根据以下一项或多项进行：图文相关性、图片分辨率、图片长宽比、图片来源权威度、广告过滤策略、反作弊过滤策略、反黄过滤策略和水印过滤策略。

在一些实施例中，装置还包括：标题生成单元 450，用于将文章主题和文章提纲输入标题模型，得到生成的文章的标题。

在一些实施例中，装置还包括：属性扩展单元（图中未示出），用于对标题中的核心词进行属性扩展；标题更新单元（图中未示出），用于对属性扩展后的标题中的核心词进行替换和改写，得到更新后的标题。

本申请还提供了一种设备的实施例，包括：一个或多个处理器；存储装置，用于存储一个或多个程序；当一个或多个程序被一个或多个处理器执行，使得一个或多个处理器实现如上任一所述的用于生成文章的方法。

本申请还提供了一种计算机可读存储介质的实施例，其上存储有计算机程序，该程序被处理器执行时实现如上任一所述的用于生成文章的方法。

下面参考图 5，其示出了适于用来实现本申请实施例的终端设备或服务器的计算机系统 500 的结构示意图。图 5 示出的终端设备仅仅是一个示例，不应对本申请实施例的功能和使用范围带来任何限制。

如图 5 所示，计算机系统 500 包括中央处理单元（CPU）501，其可以根据存储在只读存储器（ROM）502 中的程序或者从存储部分 508 加载到随机访问存储器（RAM）503 中的程序而执行各种适当的动作和处理。在 RAM 503 中，还存储有系统 500 操作所需的各种程序和数据。CPU 501、ROM 502 以及 RAM 503 通过总线 504 彼此相连。输入/输出（I/O）接口 505 也连接至总线 504。

以下部件连接至 I/O 接口 505: 包括键盘、鼠标等的输入部分 506; 包括诸如阴极射线管 (CRT)、液晶显示器 (LCD) 等以及扬声器等的输出部分 507; 包括硬盘等的存储部分 508; 以及包括诸如 LAN 卡、调制解调器等的网络接口卡的通信部分 509。通信部分 509 经由诸如
5 因特网的网络执行通信处理。驱动器 510 也根据需要连接至 I/O 接口 505。可拆卸介质 511, 诸如磁盘、光盘、磁光盘、半导体存储器等等, 根据需要安装在驱动器 510 上, 以便于从其上读出的计算机程序根据需要被安装入存储部分 508。

特别地, 根据本公开的实施例, 上文参考流程图描述的过程可以
10 被实现为计算机软件程序。例如, 本公开的实施例包括一种计算机程序产品, 其包括承载在计算机可读介质上的计算机程序, 所述计算机程序包含用于执行流程图所示的方法的程序代码。在这样的实施例中, 该计算机程序可以通过通信部分 509 从网络上被下载和安装, 和/或从可拆卸介质 511 被安装。在该计算机程序被中央处理单元 (CPU) 501
15 执行时, 执行本申请的方法中限定的上述功能。

需要说明的是, 本申请所述的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件, 或者任意以上的组合。计算机
20 可读存储介质的更具体的例子可以包括但不限于: 具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器 (RAM)、只读存储器 (ROM)、可擦式可编程只读存储器 (EPROM 或闪存)、光纤、便携式紧凑磁盘只读存储器 (CD-ROM)、光存储器件、磁存储器件、或者上述的任意合适的组合。在本申请中, 计算机可读存储介
25 质可以是任何包含或存储程序的有形介质, 该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本申请中, 计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号, 其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式, 包括但不限于电磁信号、光信号或上述的任意合适的组合。
30 计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算

机可读介质，该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于：无线、电线、光缆、RF 等等，或者上述的任意合适的组合。

5 附图中的流程图和框图，图示了按照本申请各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个单元、程序段、或代码的一部分，所述单元、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。
10 例如，两个接连地表示的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或操作的专用的基于硬件的系统来实现，
15 或者可以用专用硬件与计算机指令的组合来实现。

描述于本申请实施例中所涉及到的单元可以通过软件的方式实现，也可以通过硬件的方式来实现。所描述的单元也可以设置在处理器中，例如，可以描述为：一种处理器包括提纲生成单元、素材提取单元、素材插入单元。其中，这些单元的名称在某种情况下并不构成对该单元本身的限定，例如，提纲生成单元还可以被描述为“基于输入的文章主题和提纲生成策略，生成文章提纲的单元”。
20

作为另一方面，本申请还提供了一种非易失性计算机存储介质，该非易失性计算机存储介质可以是上述实施例中所包含的非易失性计算机存储介质；也可以是单独存在，未装配入终端中的非易失性计算机存储介质。上述非易失性计算机存储介质存储有一个或者多个程序，当所述一个或者多个程序被一个设备执行时，使得所述设备：基于输入的文章主题和以下任意一项生成文章提纲：提纲模型；根据对应文章主题的用户行为数据建立的提纲数据库；以及人工设定的提纲；从预先建立的素材库中，提取与文章提纲的特征相关联的素材；向文章提纲中，插入提取的素材，得到生成的文章。
25
30

以上描述仅为本申请的较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解，本申请中所涉及的发明范围，并不限于上述技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离上述发明构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本申请中公开的(但不限于)具有类似功能的技术特征进行互相替换而形成的技术方案。

5

权 利 要 求 书

1、一种用于生成文章的方法，其特征在于，所述方法包括：

基于输入的文章主题和以下任意一项生成文章提纲：提纲模型，

5 根据对应所述文章主题的用户行为数据建立的提纲数据库，以及人工设定的提纲；

从预先建立的素材库中，提取与所述文章提纲的特征相关联的素材；

向所述文章提纲中，插入提取的素材，得到生成的文章。

10

2、根据权利要求 1 所述的方法，其特征在于，所述根据对应所述文章主题的用户行为数据建立的提纲数据库包括：

检索全网围绕所述文章主题的子主题，建立子主题数据库；

15 根据用户对所述子主题数据库中的子主题的点击顺序和/或所述子主题数据库中的子主题的语义递进顺序，排序所述子主题数据库中的子主题；

剔除所述子主题数据库中不符合预定逻辑规则的子主题，得到符合预定逻辑规则的子主题；

将各符合预定逻辑规则的子主题作为提纲，得到提纲数据库。

20

3、根据权利要求 1 所述的方法，其特征在于，所述预先建立的素材库通过以下步骤建立：

获取素材的特征，所述素材为将现有的文章的内容根据筛选规则筛选得到和/或变换现有的文章的内容得到；

25 根据所述素材的特征建立索引结构，得到所述素材库。

4、根据权利要求 1 所述的方法，其特征在于，所述方法还包括：对所述生成的文章进行优化处理，得到优化后的所述生成的文章，所述优化处理包括以下一项或多项：润色处理、插入富媒体数据处理以及排版优化处理。

30

5、根据权利要求4所述的方法，其特征在于，所述润色处理包括以下一项或多项：统一所述生成的文章的文法风格；删除与前后语句不连贯的语句；以及替换与前后语句不连贯的语句。

5

6、根据权利要求4所述的方法，其特征在于，所述插入富媒体数据处理包括：

从预先建立的资源库，提取与所述生成的文章的特征相关联的富媒体数据；

10 向所述生成的文章中，插入提取的富媒体数据。

7、根据权利要求6所述的方法，其特征在于，所述从预先建立的资源库，提取与所述生成的文章的特征相关联的富媒体数据包括：

15 根据以下一项或多项从预先建立的资源库中提取富媒体数据生成候选富媒体列表：所述文章主题、所述文章提纲、所述生成的文章的各段落的摘要以及所述生成的文章的各段落的关键词；

采用质量筛选从所述候选富媒体列表中提取与所述生成的文章的特征相关联的富媒体数据。

20 8、根据权利要求6-7任意一项所述的方法，其特征在于，所述预先建立的资源库通过以下步骤建立：

获取富媒体数据的特征；

根据所述富媒体数据的特征建立索引结构，得到所述资源库。

25 9、根据权利要求7所述的方法，其特征在于，所述质量筛选根据以下一项或多项进行：

图文相关性、图片分辨率、图片长宽比、图片来源权威度、广告过滤策略、反作弊过滤策略、反黄过滤策略和水印过滤策略。

30 10、根据权利要求1-9任意一项所述的方法，其特征在于，所述

方法还包括:

将所述文章主题和所述文章提纲输入标题模型, 得到所述生成的文章的标题。

- 5 11、根据权利要求 10 所述的方法, 其特征在于, 所述方法还包括:
对所述标题中的核心词进行属性扩展;

对属性扩展后的标题中的核心词进行替换和改写, 得到更新后的标题。

- 10 12、一种用于生成文章的装置, 其特征在于, 所述装置包括:

提纲生成单元, 用于基于输入的文章主题和以下任意一项生成文章提纲: 提纲模型, 根据对应所述文章主题的用户行为数据建立的提纲数据库, 以及人工设定的提纲;

- 15 素材提取单元, 用于从预先建立的素材库中, 提取与所述文章提纲的特征相关联的素材;

素材插入单元, 用于向所述文章提纲中, 插入提取的素材, 得到生成的文章。

- 20 13、根据权利要求 12 所述的装置, 其特征在于, 所述提纲生成单元中的所述根据对应所述文章主题的用户行为数据建立的提纲数据库包括:

检索全网围绕所述文章主题的子主题, 建立子主题数据库;

- 25 根据用户对所述子主题数据库中的子主题的点击顺序和/或所述子主题数据库中的子主题的语义递进顺序, 排序所述子主题数据库中的子主题;

剔除所述子主题数据库中不符合预定逻辑规则的子主题, 得到符合预定逻辑规则的子主题;

将各符合预定逻辑规则的子主题作为提纲, 得到提纲数据库。

- 30 14、根据权利要求 12 所述的装置, 其特征在于, 所述素材提取单

元中的所述预先建立的素材库通过以下步骤建立：

获取素材的特征，所述素材为将现有的文章的内容根据筛选规则筛选得到和/或变换现有的文章的内容得到；

根据所述素材的特征建立索引结构，得到所述素材库。

5

15、根据权利要求 12 所述的装置，其特征在于，所述装置还包括：

文章优化单元，用于对所述生成的文章进行优化处理，得到优化后的所述生成的文章，所述优化处理包括以下一项或多项：润色处理、插入富媒体数据处理以及排版优化处理。

10

16、根据权利要求 15 所述的装置，其特征在于，所述文章优化单元中的所述润色处理包括以下一项或多项：统一所述生成的文章的文法风格；删除与前后语句不连贯的语句；以及替换与前后语句不连贯的语句。

15

17、根据权利要求 15 所述的装置，其特征在于，所述文章优化单元中的所述插入富媒体数据处理包括：

从预先建立的资源库，提取与所述生成的文章的特征相关联的富媒体数据；

20

向所述生成的文章中，插入提取的富媒体数据。

18、根据权利要求 17 所述的装置，其特征在于，所述文章优化单元中的所述从预先建立的资源库，提取与所述生成的文章的特征相关联的富媒体数据包括：

25

根据以下一项或多项从预先建立的资源库中提取富媒体数据生成候选富媒体列表：所述文章主题、所述文章提纲、所述生成的文章的各段落的摘要以及所述生成的文章的各段落的关键词；

采用质量筛选从所述候选富媒体列表中提取与所述生成的文章的特征相关联的富媒体数据。

30

19、根据权利要求 17-18 任意一项所述的装置，其特征在于，所述文章优化单元中的所述预先建立的资源库通过以下步骤建立：

获取富媒体数据的特征；

根据所述富媒体数据的特征建立索引结构，得到所述资源库。

5

20、根据权利要求 18 所述的装置，其特征在于，所述文章优化单元中的所述质量筛选根据以下一项或多项进行：

图文相关性、图片分辨率、图片长宽比、图片来源权威度、广告过滤策略、反作弊过滤策略、反黄过滤策略和水印过滤策略。

10

21、根据权利要求 12-20 任意一项所述的装置，其特征在于，所述装置还包括：

标题生成单元，用于将所述文章主题和所述文章提纲输入标题模型，得到所述生成的文章的标题。

15

22、根据权利要求 21 所述的装置，其特征在于，所述装置还包括：

属性扩展单元，用于对所述标题中的核心词进行属性扩展；

标题更新单元，用于对属性扩展后的标题中的核心词进行替换和改写，得到更新后的标题。

20

23、一种设备，其特征在于，包括：

一个或多个处理器；

存储装置，用于存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求 1-11 中任一所述的用于生成文章的方法。

25

24、一种计算机可读存储介质，其上存储有计算机程序，其特征在于，该程序被处理器执行时实现如权利要求 1-11 中任一所述的用于生成文章的方法。

30

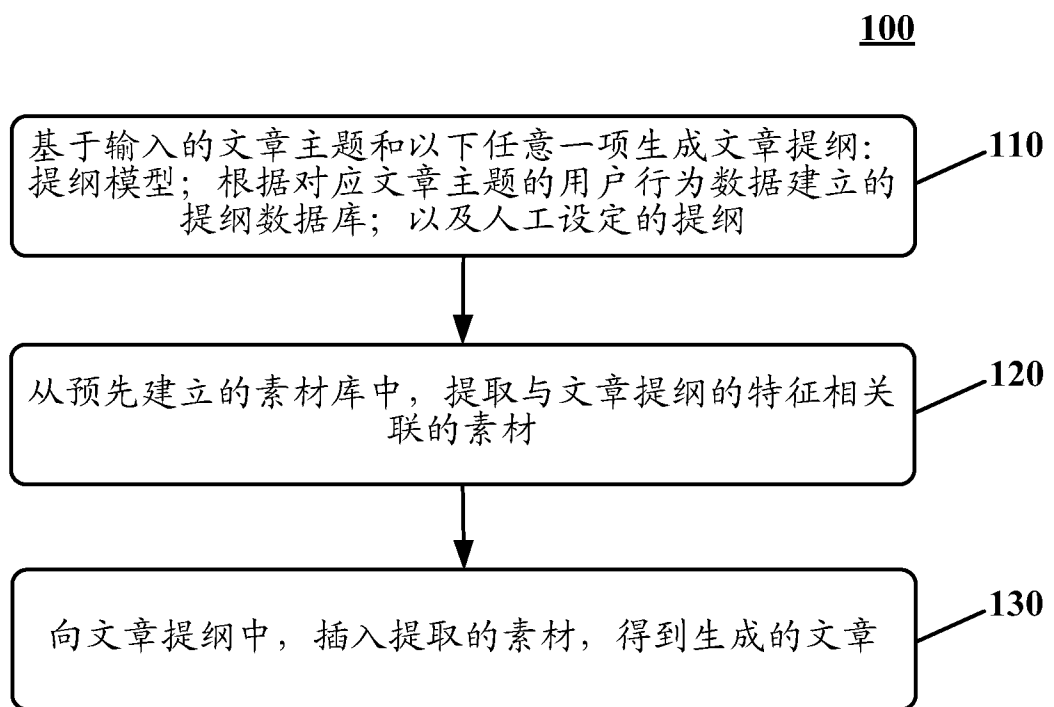


图 1

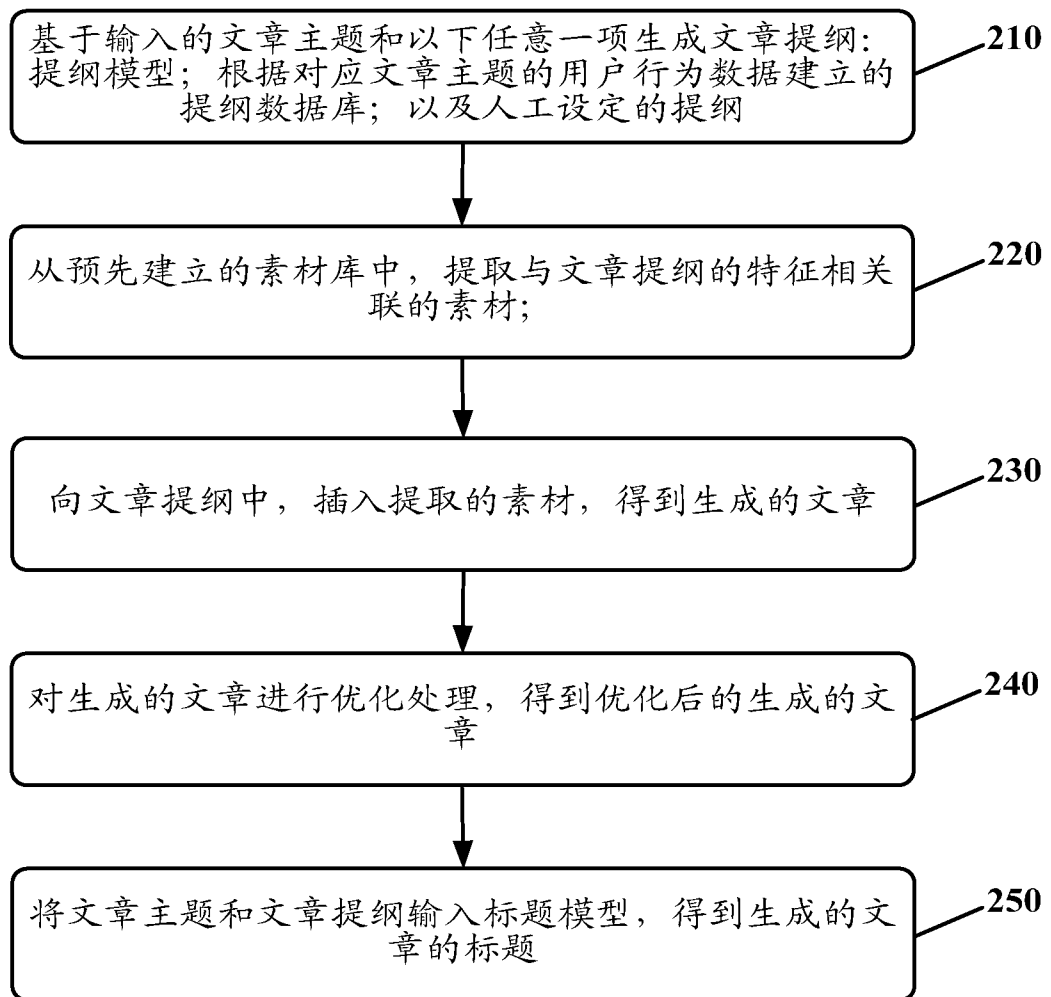
200

图 2

主题310	311: 诸葛亮 称帝
提纲320	321: 刘备托孤时为什么让诸葛亮称帝 322: 诸葛亮为什么不称帝 323: 诸葛亮如果称帝会怎么样
素材330	331: 政权问题 332: 欲擒故纵 333: 明智决定 334: 文人是造不了反的 335: 集团外部的阻力 336: 集团内部的阻力 337: 外部方面的阻力 338: 兵民厌战 339: 最关键的一点
润色340	341: 统一文风 342: 连贯语句
富媒体350	351: 图片1 352: 图片2 353: 图片3
标题360	361: 有颜有实力，集尽万千追捧的男神为何终未加冕？
排版370	371: 突出重点，颜色版式调整
输出380	381: 输出排版优化后的文章

图 3

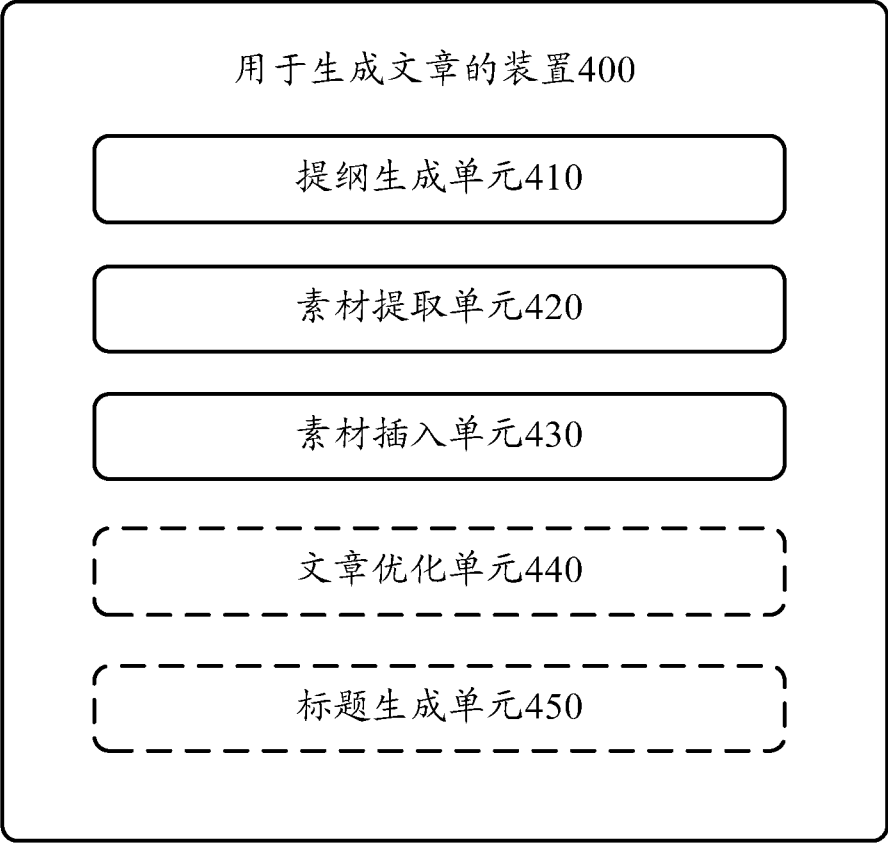


图 4

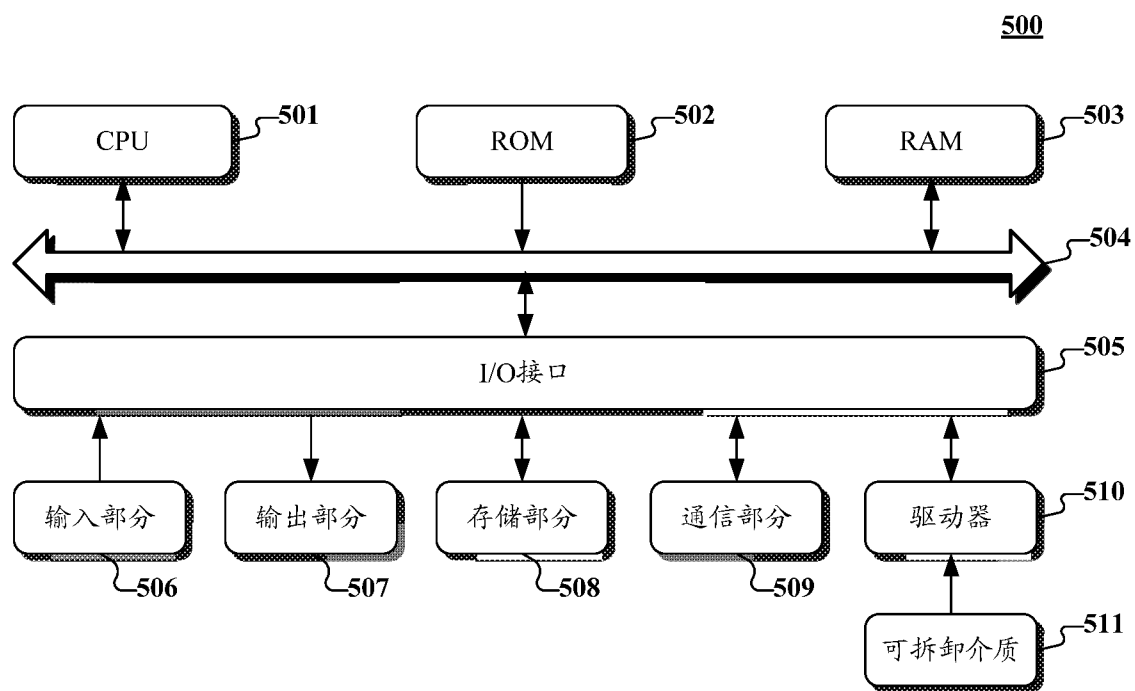


图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/102620

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/24 (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 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)

USTXT; CNTXT; CNABS; DWPI; CNKI: 文章, 生成, 主题, 提纲; article, file, publication, generate, title, theme, outline, synopsis

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 106970898 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.), 21 July 2017 (21.07.2017), claims 1-24	1-24
Y	CN 106503255 A (IFLYTEK CO., LTD.), 15 March 2017 (15.03.2017), description, paragraphs [0089]-[0094]	1-24
Y	CN 102566945 A (IFLYTEK CO., LTD. et al.), 11 July 2012 (11.07.2012), description, paragraphs [0042]-[0055]	1-24
A	CN 106407168 A (CAPITAL NORMAL UNIVERSITY et al.), 15 February 2017 (15.02.2017), entire document	1-24
A	CN 104123269 A (HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY et al.), 29 October 2014 (29.10.2014), entire document	1-24
A	CN 101441621 A (PEKING UNIVERSITY FOUNDER GROUP CORP. et al.), 27 May 2009 (27.05.2009), entire document	1-24

☐ 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 02 November 2017	Date of mailing of the international search report 30 November 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 HAO, Xiaoli Telephone No. (86-10) 62089281
---	---

International application No.
PCT/CN2017/102620

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号

PCT/CN2017/102620

A. 主题的分类

G06F 17/24(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)

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))

USTXT;CNTXT;CNABS;DWPI;CNKI:文章, 生成, 主题, 提纲; article, file, publication, generate, title, theme, outline, synopsis

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
PX	CN 106970898 A (百度在线网络技术北京有限公司) 2017年 7月 21日 (2017 - 07 - 21) 权利要求1-24	1-24
Y	CN 106503255 A (科大讯飞股份有限公司) 2017年 3月 15日 (2017 - 03 - 15) 说明书第[0089]-[0094]段	1-24
Y	CN 102566945 A (北大方正集团有限公司等) 2012年 7月 11日 (2012 - 07 - 11) 说明书第[0042]-[0055]段	1-24
A	CN 106407168 A (首都师范大学等) 2017年 2月 15日 (2017 - 02 - 15) 全文	1-24
A	CN 104123269 A (华中科技大学等) 2014年 10月 29日 (2014 - 10 - 29) 全文	1-24
A	CN 101441621 A (北大方正集团有限公司等) 2009年 5月 27日 (2009 - 05 - 27) 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

2017年 11月 2日

国际检索报告邮寄日期

2017年 11月 30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

郝晓丽

电话号码 (86-10)62089281

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/102620

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	106970898	A	2017年 7月 21日	无			
CN	106503255	A	2017年 3月 15日	无			
CN	102566945	A	2012年 7月 11日	CN	102566945	B	2015年 3月 18日
CN	106407168	A	2017年 2月 15日	无			
CN	104123269	A	2014年 10月 29日	CN	104123269	B	2016年 10月 5日
CN	101441621	A	2009年 5月 27日	CN	101441621	B	2010年 12月 1日

表 PCT/ISA/210 (同族专利附件) (2009年7月)