



(12) 发明专利申请

(10) 申请公布号 CN 105760359 A

(43) 申请公布日 2016. 07. 13

(21) 申请号 201410782497. 7

(22) 申请日 2014. 12. 17

(30) 优先权数据

103140400 2014. 11. 21 TW

(71) 申请人 财团法人工业技术研究院

地址 中国台湾新竹县竹东镇中兴路 4 段 195 号

(72) 发明人 沈民新 邱中人 张如莹 张俊盛

(74) 专利代理机构 中科专利商标代理有限责任公司 11021

代理人 宋焰琴

(51) Int. Cl.

G06F 17/27(2006. 01)

权利要求书2页 说明书8页 附图10页

(54) 发明名称

问句处理系统及其方法

(57) 摘要

本发明公开了一种问句处理系统及其方法，问句处理系统包括一具有错字处理单元与误用词处理单元的问句建构模块以及一问句类别处理模块。错字处理单元用于检测并修正自然语言问句的错别字词或火星文，以产生一符合自然语言问句的问句意图的校正问句。误用词处理单元用于分析校正问句的至少二词组的搭配关系，并依据搭配关系修正校正问句的误用字词以产生至少一符合问句意图的候选问句。问句类别处理模块用于分析候选问句以产生问句类别。由此，本发明可具备良好的容错能力。

1. 一种问句处理系统,应用于具有处理器、存储器与作业系统的电子装置中,其特征为,该问句处理系统包括:

问句建构模块,其具有:

错字处理单元,用于检测并修正自然语言问句的错别字词或火星文,以产生一符合该自然语言问句的问句意图的校正问句;以及

误用词处理单元,用于分析该校正问句的至少二词组的搭配关系,并依据该搭配关系修正该校正问句的误用字词以产生至少一符合该问句意图的候选问句;以及

问句类别处理模块,用于分析该候选问句以产生该候选问句的问句类别。

2. 如权利要求 1 所述的问句处理系统,其特征为,该系统更包括使用者界面,用于供使用者输入该自然语言问句。

3. 如权利要求 1 所述的问句处理系统,其特征为,该错字处理单元用于依据错别字检测法或火星文转译法,以检测并修正该自然语言问句的错别字词或火星文。

4. 如权利要求 1 所述的问句处理系统,其特征为,该错字处理单元具有翻译模型与语言模型,该翻译模型用于提供该自然语言问句的错别字词或火星文的修正资料,该语言模型用于依据该修正资料修正该自然语言问句的错别字词或火星文以产生该校正问句。

5. 如权利要求 1 所述的问句处理系统,其特征为,该误用词处理单元用于分析该问句意图与该校正问句的词组的语境,并依据该语境自语料库或同义/近义词库中撷取至少一第一搭配词以修正该校正问句的误用字词而产生该候选问句。

6. 如权利要求 1 所述的问句处理系统,其特征为,该问句建构模块更具有缺漏词处理单元,用于分析该候选问句的缺漏字词,并自语料库或同义/近义词库中撷取至少一第二搭配词以补足该校正问句的缺漏字词而产生该候选问句。

7. 如权利要求 1 所述的问句处理系统,其特征为,该至少一候选问句为多个最优先的候选问句,该问句类别处理模块用于借助问句分类模型与知识库分析这些最优先的候选问句而产生这些最优先的候选问句的问句类别。

8. 如权利要求 7 所述的问句处理系统,其特征为,该问句类别处理模块更依据这些最优先的候选问句的可信度重新排序这些最优先的候选问句,并自这些最优先的候选问句中撷取具有最高可信度者作为第一优先的候选问句。

9. 如权利要求 8 所述的问句处理系统,其特征为,该问句建构模块更具有关键词组撷取单元,用于依据该第一优先的候选问句产生至少一关键词组或一问句建构结果。

10. 如权利要求 9 所述的问句处理系统,其特征为,该系统更包括段落检索模块与答案处理模块,该段落检索模块用于自文件中撷取符合该第一优先的候选问句的关键词组或问句建构结果的段落,该答案处理模块用于自该段落中撷取符合该第一优先的候选问句的问句类别的答案。

11. 一种问句处理方法,应用于具有处理器、存储器与作业系统的电子装置中,该问句处理方法包括:

检测并修正自然语言问句的错别字词或火星文,以产生一符合该自然语言问句的问句意图的校正问句;

分析该校正问句的至少二词组的搭配关系,并依据该搭配关系修正该校正问句的误用字词以产生至少一符合该问句意图的候选问句;以及

分析该候选问句以产生该候选问句的问句类别。

12. 如权利要求 11 所述的问句处理方法,其特征为,该方法更包括依据错别字检测法或火星文转译法以检测并修正该自然语言问句的错别字词或火星文。

13. 如权利要求 11 所述的问句处理方法,其特征为,该方法更包括提供该自然语言问句的错别字词或火星文的修正资料,并依据该修正资料修正该自然语言问句的错别字词或火星文以产生该校正问句。

14. 如权利要求 11 所述的问句处理方法,其特征为,该方法更包括分析该问句意图与该校正问句的词组的语境,并依据该语境自语料库或同义 / 近义词库中撷取至少一第一搭配词以修正该校正问句的误用字词而产生该候选问句。

15. 如权利要求 11 所述的问句处理方法,其特征为,该方法更包括分析该候选问句的缺漏字词,并自语料库或同义 / 近义词库中撷取至少一第二搭配词以补足该校正问句的缺漏字词而产生该候选问句。

16. 如权利要求 11 所述的问句处理方法,其特征为,该至少一候选问句包括多个最优先的候选问句,以藉由问句分类模型与知识库分析这些最优先的候选问句而产生这些最优先的候选问句的问句类别。

17. 如权利要求 16 所述的问句处理方法,其特征为,该方法更包括依据这些最优先的候选问句的可信度重新排序这些最优先的候选问句,并自这些最优先的候选问句中撷取具有最高可信度者作为第一优先的候选问句。

18. 如权利要求 17 所述的问句处理方法,其特征为,该方法更包括依据该第一优先的候选问句产生至少一关键词组或一问句建构结果。

19. 如权利要求 18 所述的问句处理方法,其特征为,该方法更包括自文件中撷取符合该第一优先的候选问句的关键词组或问句建构结果的段落,并自该段落中撷取符合该第一优先的候选问句的问句类别的答案。

问句处理系统及其方法

技术领域

[0001] 本发明涉及一种问句处理系统及其方法,特别是指一种具备容错能力的问句处理系统及其方法。

背景技术

[0002] 现有技术的搜索引擎或问答系统中,由于其不具备问句容错能力,因此当使用者输入含有错别字词、火星文 (Martian language)、误用字词或缺漏字词的自然语言问句时,可能会造成该搜索引擎或问答系统误判该自然语言问句的问句意图,因而回复错误的答案给该使用者。

[0003] 图 1A 至图 1C 分别表示现有技术中含有火星文、音似的错别字词与形似的错别字词的自然语言问句的表单。图中,这些自然语言问句中含有许多不恰当的用语(以底线标示),例如图 1A 所示的火星文(如注音文),或者图 1B 所示音似的错别字词,亦或者图 1C 所示形似的错别字词,这些火星文或错别字词将明显地降低搜索引擎或问答系统对该自然语言问句的回复答案的正确率。

[0004] 图 2A 至图 2G 分别表示现有技术中以搜索引擎或自动问答系统提供含有关键词组 11、错别字词 13、火星文 14 或误用字词 15 的自然语言问句的答案的网页。

[0005] 在图 2A 的搜索引擎(如 Google)中,该搜索引擎对于例如中文的自然语言问句“日本战岭台湾几年”,虽可将关键词组 11 “战岭”修正为关键词组 12 “占领”,并直接以“日本占领台湾几年”进行搜索,但仍无法搜索到有关“几年”的正确答案。

[0006] 在图 2B 的搜索引擎(如 Google)中,该搜索引擎对于例如中文的自然语言问句“杨传广是那—足的”,并无法将错别字词 13 “那—足的”进行修正,以致无法搜索到适当的答案。

[0007] 而在图 2C 的搜索引擎(如 Google)中,该搜索引擎对于例如中文的自然语言问句“郑成功的ㄐㄩㄥ、点?”,也无法将火星文 14(如注音文)“ㄐㄩㄥ、”进行修正,因而无法搜索到正确的答案。

[0008] 又在图 2D 的搜索引擎(如 Google)中,该搜索引擎对于例如英文的自然语言问句“rice plented”,虽可将关键词组 11 “rice plented”修正为“rice planted”以进行搜索,但仍无法搜索到适当的答案。

[0009] 另在图 2E 的搜索引擎(如 Google)中,该搜索引擎对于例如英文的自然语言问句“whom is taiwan president”,虽可直接以“who”代替关键词组 11 “whom”进行搜索,但仍无法搜索到正确的答案。

[0010] 而在图 2F 的自动问答系统(如 WorframAlpha)中,该自动问答系统对于例如英文的自然语言问句“where does rice live”,并无法将误用字词 15 “rice live”进行修正,以致回复错误的答案 16。

[0011] 又在图 2G 的自动问答系统(如 WorframAlpha)中,该自动问答系统对于例如英文的自然语言问句“Where is the Taiwan President”,也无法将误用字词 15 “Where”进行

修正,因而回复错误的答案 16。

[0012] 因此,如何克服上述现有技术的问题,实已成目前亟需解决的课题。

发明内容

[0013] 本发明提供一种问句处理系统及其方法,其可具备良好的容错能力,以提高对自然语言问句的回复答案的正确率。

[0014] 本发明的问句处理系统应用于具有处理器、存储器与作业系统的电子装置中,且该问句处理系统包括一具有错字处理单元与误用词处理单元的问句建构模块以及一问句类别处理模块。该错字处理单元用于检测并修正自然语言问句的错别字词或火星文,以产生一符合该自然语言问句的问句意图的校正问句。该误用词处理单元用于分析该校正问句的至少二词组的搭配关系,并依据该搭配关系修正该校正问句的误用字词以产生至少一符合该问句意图的候选问句。该问句类别处理模块用于分析该候选问句以产生该候选问句的问句类别。

[0015] 本发明的问句处理方法应用于具有处理器、存储器与作业系统的电子装置中,且该问句处理方法包括:检测并修正自然语言问句的错别字词或火星文,以产生一符合该自然语言问句的问句意图的校正问句;分析该校正问句的至少二词组的搭配关系,并依据该搭配关系修正该校正问句的误用字词以产生至少一符合该问句意图的候选问句;以及分析该候选问句以产生该候选问句的问句类别。

[0016] 上述的问句处理系统及其方法中,可以缺漏词处理单元分析该候选问句的缺漏字词,并自语料库或同义/近义词库中撷取至少一搭配词以补足该校正问句的缺漏字词而产生该候选问句。

[0017] 由上述内容可知,本发明的问句处理系统及其方法中,主要是通过问句建构模块的错字处理单元、误用词处理单元与缺漏词处理单元,以分别修正自然语言问句的错别字词、火星文、误用字词及缺漏字词,并通过问句类别处理模块分析该自然语言问句的问句类别。

[0018] 由此,本发明能具备良好的容错能力,以容忍该自然语言问句的错别字词、火星文、误用字词及缺漏字词,并降低对该自然语言问句的问句意图的分析错误所造成的影响,进而提高对该自然语言问句的回复答案的正确率。

附图说明

[0019] 图 1A 至图 1C 分别表示现有技术中含有火星文、音似的错别字词与形似的错别字词的问句的自然语言问句的表单;

[0020] 图 2A 至图 2G 分别表示现有技术中以搜索引擎搜索或自动问答系统提供含有关键词组、错别字词、火星文或误用字词的问句的答案的网页;

[0021] 图 3 表示本发明的问句处理系统的方框结构示意图;

[0022] 图 4 表示本发明的问句处理方法的流程示意图;

[0023] 图 5 表示本发明的问句处理系统及其方法的实施例示意图。

[0024] 符号说明:

[0025] 11、12 关键词组;

[0026]	13	错别字词；
[0027]	14	火星文；
[0028]	15	误用字词；
[0029]	16	答案；
[0030]	2	问句处理系统；
[0031]	20	使用者界面；
[0032]	21	问句建构模块；
[0033]	211	错字处理单元；
[0034]	211a	翻译模型；
[0035]	211b	语言模型；
[0036]	212	误用词处理单元；
[0037]	213	缺漏词处理单元；
[0038]	214	关键词组撷取单元；
[0039]	22	问句类别处理模块；
[0040]	23	语料库；
[0041]	24	同义 / 近义词库；
[0042]	25	知识库；
[0043]	26	段落检索模块；
[0044]	261	文件；
[0045]	262	段落；
[0046]	27	答案处理模块；
[0047]	271	答案；
[0048]	41	自然语言问句；
[0049]	42	候选问句；
[0050]	43、44	关键词组；
[0051]	S31 至 S36	步骤。

具体实施方式

[0052] 为使本发明的目的、技术方案和优点更加清楚明白，以下结合具体实施例，并参照附图，对本发明作进一步的详细说明。

[0053] 图 3 表示本发明的问句处理系统 2 的方框结构示意图。如图所示，问句处理系统 2 可应用于具有处理器、存储器与作业系统的电子装置中，且该问句处理系统 2 主要包括一使用者界面 (User Interface, UI) 20、一具有错字处理单元 211 与误用词处理单元 212 的问句建构模块 21、以及一问句类别处理模块 22。该电子装置可为个人电脑、平板电脑、笔记本电脑、网络服务器、云端服务器、移动电话或智能手机等。

[0054] 该使用者界面 20 用于供使用者输入自然语言问句，且该自然语言问句可为中文、英文或各种的语言。该错字处理单元 211 用于检测并修正该自然语言问句的错别字词或火星文，以产生一符合该自然语言问句的问句意图的校正问句。

[0055] 具体而言，该错字处理单元 211 可依据错别字检测法或火星文转译法，以检测并

修正该自然语言问句的错别字词或火星文,且该错字处理单元 211 可具有翻译模型 211a 与语言模型 (Language Model) 211b。

[0056] 该翻译模型 211a 用于提供该自然语言问句的错别字词或火星文的修正资料,例如:(1) 音似或形似的错误字、(2) 词组的字汇的特征值(如同偏旁、同字首、部首差别、部首笔划差、偏旁笔划差、注音差或调号差)、(3) 易混淆字汇(如躁 vs. 燥)。

[0057] 此外,该语言模型 211b 用于依据该修正资料修正该自然语言问句的错别字词或火星文以产生该校正问句,且该语言模型 211b 可为以 n- 连词为基础的统计式语言模型 (Statistical Language Model, SLM)、或基于神经网络语言模型 (Neural Network-based Language Modeling, NNLM) 等。该语言模型 211b 可具有解码单元 (decoder) 以转换一个中文字、一个注音符号或一串完整的注音符号至原字、音似 / 形似字或注音的对应字。

[0058] 上述的错别字词例如可以为现有技术图 1B 所示音似的错别字词、或图 1C 所示形似的错别字词,且该错别字词可为该自然语言问句的疑问词 (question words) 或功能词 (function words) 等,该功能词可为限定词(如这、一只、我的)、代名词(如你、我、他)、前置词 / 介系词 / 后置词(如上、下、为了)、或连接词(如和、或、如果)等。该火星文可例如为现有技术图 1A 所示的注音文,也可为表情符号(如*、#、!)等。

[0059] 举例来说,在例如中文的自然语言问句中,该错字处理单元 211 可检测使用者所输入的自然语言问句“水稻住在舍么勿”的错别字词“舍”与火星文“勿”,并依据该自然语言问句的问句意图将“舍”与“勿”分别修正为正确字词“什”与“地”,也就是将该自然语言问句“水稻住在舍么勿”修正为校正问句“水稻住在什么地”。

[0060] 该误用词处理单元 212 用于分析该校正问句的至少二词组的搭配关系,并依据该搭配关系修正该校正问句的误用字词以产生至少一符合该问句意图的候选问句。

[0061] 具体地说,该误用词处理单元 212 用于分析该问句意图与该校正问句的至少二词组的语境 (context) 是否冲突,并于发生冲突时依据该语境且自语料库 23 或同义 / 近义词库 24 中撷取至少一第一搭配词,以利用该第一搭配词修正该校正问句的误用字词而产生该候选问句,使得该候选问句的词组的语境不具有冲突且符合该问句意图。

[0062] 例如,该误用词处理单元 212 分析出该校正问句“水稻住在什么地”中,三个词组“水稻”、“住”与“地”的搭配关系较差并只有冲突性,因“水稻”通常不与“住”共用在同一问句中,而且该词组“住”应为误用字词,故该误用词处理单元 212 可依据该三个词组的搭配关系以撷取至少一第一搭配词“种植”或“栽种”来修正该词组“住”,也就是将该校正问句“水稻住在什么地”修正为符合该问句意图的候选问句“水稻种植在什么地”或“水稻栽植在什么地”……,等等。

[0063] 该问句类别处理模块 22 用于分析该候选问句以产生该候选问句的问句类别,且该问句类别可为人、事、时、地、物、数量、质量、速度、高度、尺寸、……、等各种类型或种类。

[0064] 该问句建构模块 21 也可具有缺漏词处理单元 213,用于分析该候选问句的缺漏字词,并自该语料库 23 或同义 / 近义词库 24 中撷取至少一第二搭配词,以利用该第二搭配词补足该校正问句的缺漏字词而产生该候选问句,使得该候选问句的词组的语境完整且符合该问句意图。

[0065] 例如,该缺漏词处理单元 213 分析出该候选问句“水稻种植在什么地”或“水稻栽植在什么地”中,“地”应为“地方”或“地区”之意,则该缺漏词处理单元 213 撷取至少一第

二搭配词“地方”或“地区”以修正“地”并加上问号“? ”,以便补足该校正问句的缺漏字词而产生完整的候选问句,也就是将该候选问句“水稻种植在什么地”或“水稻栽植在什么地”修正为“水稻种植在什么地方?”、“水稻种植在什么地区?”、“水稻栽植在什么地方?”或“水稻栽植在什么地区?”等等。

[0066] 上述至少一候选问句可为多个最优先的候选问句,且该问句类别处理模块 22 可依据问句分类模型与知识库 25,以分析这些最优先的候选问句而产生这些最优先的候选问句的问句类别。

[0067] 该问句分类模型可包括混合分类法 (hybrid approaches)、正规表示规则 (regular expression rule)、机器学习分类器 (classifier for machine learning)、支持向量机 (support vector machine, SVM)、最大熵函数分类器 (Maximum Entropy classifier)、或决策树分类器 (decision tree classifier) 等。

[0068] 该知识库 25 可提供对应这些最优先的候选问句的问句类别的资料,例如:最优先的候选问句为“至圣先师是哪一位?”,则该知识库 25 提供该问句类别为“人”。该知识库 25 也可提供对应这些最优先的候选问句的问句类别的规则,例如:假如“有”字后面或前面接“哪些人”、“哪位”或“哪几位”,则该知识库 25 提供该问句类别为“人”;或者,假如“要”字后面接“多久”,则该知识库 25 提供该问句类别为“时”。

[0069] 该问句类别处理模块 22 可依据这些最优先的候选问句的可信度重新排序这些最优先的候选问句,并自这些最优先的候选问句中撷取超过预定的可信度阈值且具有最高可信度者作为第一优先的候选问句。

[0070] 例如,该问句类别处理模块 22 可重新排序上述的候选问句“水稻种植在什么地方?”、“水稻种植在什么地区?”、“水稻栽植在什么地方?”与“水稻栽植在什么地区?”,并以“水稻栽植在什么地区?”作为该第一优先的候选问句。

[0071] 该问句建构模块 21 可具有关键词组撷取单元 214,用于依据该第一优先的候选问句产生至少一关键词组或一问句建构结果。例如,该关键词组撷取单元 214 可自该第一优先的候选问句“水稻栽植在什么地区?”中产生三个关键词组“水稻”、“栽种”及“地区”,或者产生一个问句建构结果“水稻栽种地区”。

[0072] 该问句处理系统 2 可包括段落检索模块 26 与答案处理模块 27,该段落检索模块 26 用于自至少一文件 261 中撷取符合该第一优先的候选问句的关键词组或问句建构结果的段落 262,而该答案处理模块 27 用于自该段落 262 中撷取符合该第一优先的候选问句的问句类别的答案 271,以将该答案 271 (或包括该段落 262) 显示于该使用者界面 20。

[0073] 图 4 表示本发明的问句处理方法的流程示意图,图 5 表示本发明的问句处理系统 2 及具方法的实施例示意图,请一并参阅上述图 3 的问句处理系统 2。

[0074] 本发明的问句处理方法可应用于具有处理器、存储器与作业系统的电子装置中,且该电子装置可为个人电脑、平板电脑、笔记本电脑、网络服务器、云端服务器、移动电话或智能手机等。同时,本发明的问句处理方法主要包括下列步骤:

[0075] (1) 如图 4 的步骤 S31 与图 5 所示,在小学生知识问答系统中,先由使用者自使用者界面 20 中输入自然语言问句 41 “水稻住在舍么勿”,并由该问句处理系统 2 接收该自然语言问句 41。接着,跳到步骤 S32。

[0076] (2) 如图 4 的步骤 S32 所示,由问句建构模块 21 的错字处理单元 211 检测并修正

该自然语言问句 41 的错别字词或火星文,以产生一符合该自然语言问句 41 的问句意图的校正问句。

[0077] 具体而言,该错字处理单元 211 可依据错别字检测法或火星文转译法,以检测并修正该自然语言问句 41 的错别字词或火星文。同时,该错字处理单元 211 可具有翻译模型 211a 与语言模型 211b,该翻译模型 211a 用于提供该自然语言问句 41 的错别字词或火星文的修正资料,且该语言模型 211b 用于依据该修正资料修正该自然语言问句 41 的错别字词或火星文以产生该校正问句。

[0078] 例如,该错字处理单元 211 可检测该自然语言问句“水稻住在舍么勿”的错别字词“舍”与火星文“勿”,并依据该自然语言问句 41 的问句意图将“舍”与“勿”分别修正为正确字词“什”与“地”,也就是将该自然语言问句“水稻住在舍么勿”修正为校正问句“水稻住在什么地”。接着,跳到步骤 S33。

[0079] (3) 如图 4 的步骤 S33 所示,由该问句建构模块 21 的误用词处理单元 212 分析该校正问句的至少二词组的搭配关系,并依据该搭配关系修正该校正问句的误用字词。同时,可由该问句建构模块 21 的缺漏词处理单元 213 补足该校正问句的缺漏字词而产生一或多个最优先的候选问句。

[0080] 具体来说,该误用词处理单元 212 用于分析该问句意图与该校正问句的至少二词组的语境是否冲突,并于发生冲突时依据该语境自语料库 23 或同义/近义词库 24 中撷取至少一第一搭配词,以利用该第一搭配词修正该校正问句的误用字词而产生该候选问句,使得该候选问句的词组的语境不具有冲突且符合该问句意图。

[0081] 例如,该误用词处理单元 212 分析出该校正问句“水稻住在什么地”中,三个词组“水稻”、“住”与“地”的搭配关系较差并具有冲突性,因“水稻”通常不与“住”共用在同一问句中,而且该词组“住”应为误用字词,故该误用词处理单元 212 可依据该三个词组的搭配关系,以撷取至少一第一搭配词“种植”或“栽种”来修正该词组“住”,也就是将校正问句“水稻住在什么地”修正为符合该问句意图的候选问句“水稻种植在什么地”或“水稻栽植在什么地”等等。

[0082] 而该缺漏词处理单元 213 用于分析该候选问句的缺漏字词,并自该语料库 23 或同义/近义词库 24 中撷取至少一第二搭配词,以利用该第二搭配词补足该校正问句的缺漏字词而产生该候选问句,使得该候选问句的词组的语境完整且符合该问句意图。

[0083] 例如,该缺漏词处理单元 213 分析出该候选问句“水稻种植在什么地”或“水稻栽植在什么地”中,“地”应为“地方”或“地区”之意,则该缺漏词处理单元 213 撷取至少一第二搭配词“地方”或“地区”修正“地”并加上问号“?”,藉以补足该校正问句的缺漏字词而产生完整的候选问句,也就是将该候选问句“水稻种植在什么地”或“水稻栽植在什么地”修正为“水稻种植在什么地方?”、“水稻种植在什么地区?”、“水稻栽植在什么地方?”或“水稻栽植在什么地区?”等等。接着,跳到步骤 S34。

[0084] (4) 如图 4 的步骤 S34 所示,由问句类别处理模块 22 依据问句分类模型与知识库 25 分析这些最优先的候选问句以产生这些最优先的候选问句的问句类别。

[0085] 另外,可由该问句类别处理模块 22 依据这些最优先的候选问句的可信度重新排序这些最优先的候选问句,并自这些最优先的候选问句中撷取超过预定的可信度阈值且具有最高可信度者作为第一优先的候选问句。

[0086] 例如,该问句类别处理模块 22 可重新排序上述的候选问句“水稻种植在什么地方?”、“水稻种植在什么地区?”、“水稻栽植在什么地方?”与“水稻栽植在什么地区?”,并以“水稻栽植在什么地区?”作为该第一优先的候选问句,如图 5 所示“我猜你想问‘水稻栽植在什么地区?’”的候选问句 42 “水稻栽植在什么地区? ”。接着,跳到步骤 S35。

[0087] (5) 如图 4 的步骤 S35 所示,由该问句建构模块 21 的关键词组提取单元 214 依据该第一优先的候选问句产生至少一关键词组或一问句建构结果。例如,自该第一优先的候选问句“水稻栽植在什么地区?”中,产生如图 5 所示的关键词组 43 “稻”及关键词组 44 “栽种”等,或者产生一个问句建构结果“水稻栽植地区”。接着,跳到步骤 S36。

[0088] (6) 如图 4 的步骤 S36 所示,由段落检索模块 26 自至少一文件 261 中提取符合该第一优先的候选问句的关键词组或问句建构结果的段落 262,并由答案处理模块 27 自该段落 262 中提取符合该第一优先的候选问句的问句类别的答案 271,以将该答案 271 (或包括该段落 262) 显示于该使用者界面 20 上。

[0089] 例如,自图 5 所示维基百科的文件中提取符合关键词组 43 “稻”及关键词组 44 “栽种”的段落,并将答案“水稻在中国大陆广为栽种后,逐渐向西传播到印度,中世纪引入欧洲南部,现时全世界有一半的人口食用水稻,主要在亚洲、欧洲南部和中美洲及非洲部分地区”显示于该使用者界面 20 上。而且,该答案可以是上述的一个段落,也可以是一个简单答案,如“中国大陆”。

[0090] 同理,在例如英文的自然语言问句中,一样可以采用上述图 3 的问句处理系统 2 与图 4 的问句处理方法,下面以一个例子简单说明之。

[0091] (1) 如同上述图 3 与图 4 的步骤 S31 所示,由使用者自使用者界面 20 输入自然语言问句 41 “What does rice live?”,并由该问句处理系统 2 接收该自然语言问句 41。

[0092] (2) 如同上述图 3 与图 4 的步骤 S32 所示,由问句建构模块 21 的错字处理单元 211 检测并修正该自然语言问句 41 “What does rice live?”的错别字词或火星文,以产生一符合该自然语言问句 41 的问句意图的校正问句。

[0093] 因该错字处理单元 211 并未检测到该自然语言问句 41 “What does rice live?”中含有错别字词或火星文,也符合该自然语言问句 41 的问句意图,故可直接以该自然语言问句 41 作为该校正问句“What does rice live?”。

[0094] (3) 如同上述图 3 与图 4 的步骤 S33 所示,由该问句建构模块 21 的误用词处理单元 212 分析该校正问句的至少二词组“Where”、“does”及“live”的搭配关系,并依据该搭配关系修正该校正问句的误用字词“live”为正确字词“grown”或“planted”。

[0095] 同时,可由该问句建构模块 21 的缺漏词处理单元 213 补足该校正问句的缺漏字词而产生一或多个最优先的候选问句。因该校正问句“What does rice live?”中并未含有缺漏字词,故该缺漏词处理单元 213 可直接产生一或多个最优先的候选问句,例如该候选问句为“where does rice grown?”与“where is rice planted?”。

[0096] (4) 如同上述图 3 与图 4 的步骤 S34 所示,由问句类别处理模块 22 依据问句分类模型与知识库 25 分析这些最优先的候选问句,以产生这些最优先的候选问句的问句类别,例如该问句类别为“where”。

[0097] 另外,可由该问句类别处理模块 22 依据这些最优先的候选问句的可信度重新排序这些最优先的候选问句,并自这些最优先的候选问句中提取超过预定的可信度阈值且

具有最高可信度者作为第一优先的候选问句,例如该第一优先的候选问句为“where does rice grown?”。

[0098] (5) 如同上述图 3 与图 4 的步骤 S35 所示,由该问句建构模块 21 的关键词组撷取单元 214 依据该第一优先的候选问句产生至少一关键词组或一问句建构结果,例如该关键词组为“where”、“rice”及“grown”,或者该问句建构结果为“where rice grown”。

[0099] (6) 如同上述图 3 与图 4 的步骤 S36 所示,由段落检索模块 26 自至少一文件 261 中撷取符合该第一优先的候选问句的关键词组或问句建构结果的段落 262,并由答案处理模块 27 自该段落 262 中撷取符合第一优先的候选问句的问句类别的答案 271,以将该答案 271(或包括该段落 262)显示于该使用者界面 20 上。

[0100] 由上述内容可知,本发明的问句处理系统及其方法中,主要是通过问句建构模块的错字处理单元、误用词处理单元与缺漏词处理单元,以分别修正自然语言问句的错别字词、火星文、误用字词及缺漏字词,并通过问句类别处理模块分析该自然语言问句的问句类别。

[0101] 由此,本发明能具备良好的容错能力,以容忍该自然语言问句的错别字词、火星文、误用字词及缺漏字词,并降低对该自然语言问句的问句意图的分析错误所造成的影响,进而提高对该自然语言问句的回复答案的正确率。

[0102] 以上所述的具体实施例,对本发明的目的、技术方案和有益效果进行了进一步详细说明,应理解的是,以上所述仅为本发明的只体实施例而已,并不用于限制本发明,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

问句	词性
孔子说的话又 _レ 怎么说？	疑问词
郑成功的 _レ 山口、点在哪里	名词
台湾 _レ 山口、最多神 _レ 山口、的地方？	名词
台湾最 _レ 又 _レ < 一么、的山壁	形容词
以后，我们生活中的水会不会变 _レ 下尤？	形容词
鞭炮是在 _レ 口赶什么？	动词

图 1A

问句	词性
什么祭典在阿美族最驻名？	形容词
可以核融核的原理制造火箭？	名词
核溶合是什么？	名词
元子比较大还是电子比较大	名词
日本战岭台湾几年	动词
吃元宵时像争什么	动词
为什么中秋节相征团圆？	动词
台湾最长被破坏的东西是啥	形容词
台湾最漠落的地区	形容词

图 1B

问句	词性
孔子的第二个名子叫啥	名词
台湾是一个寒岛吗	名词
为何日本人让台【 <u>让变</u> 】人投票	名词
台湾第一座西【 <u>工戈</u> 】炮台上刻了什么字？	形容词
台湾海峡拉于台湾的哪里？	动词
盖亿载金城花子多少银两？	感叹词
舍时武昌起义	疑问词
为盒会下雨	疑问词

图 1C

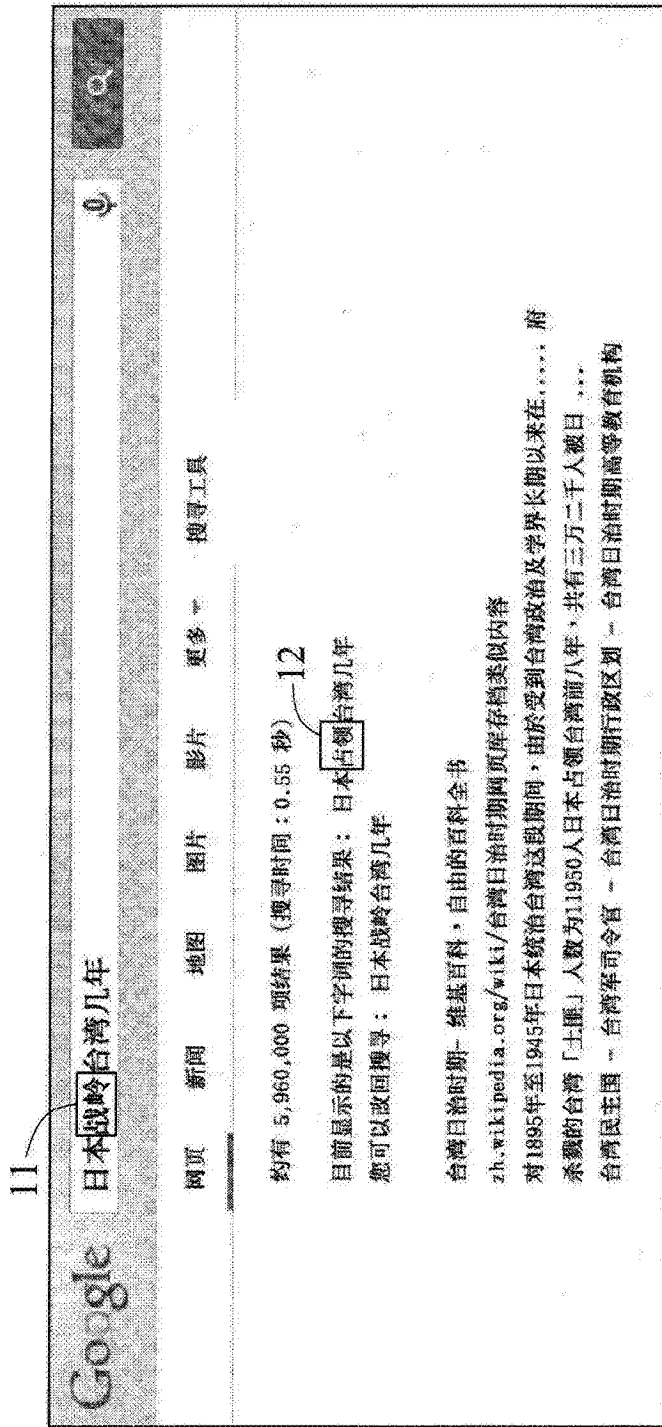


图 2A

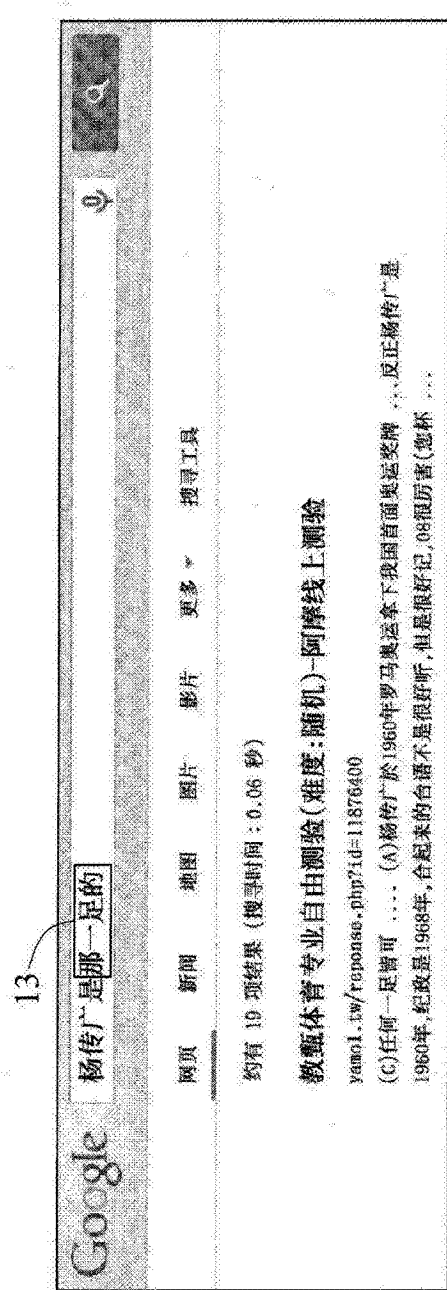


图 2B

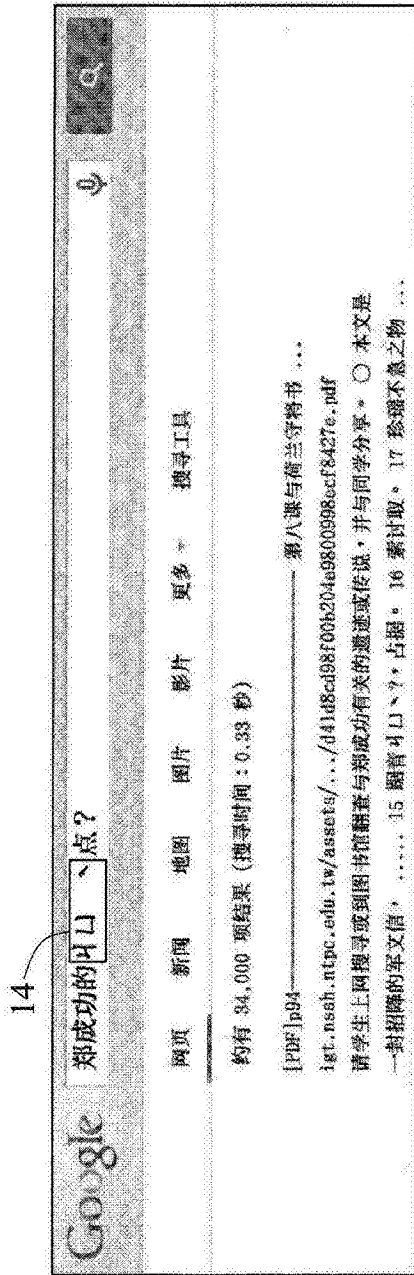


图 2C

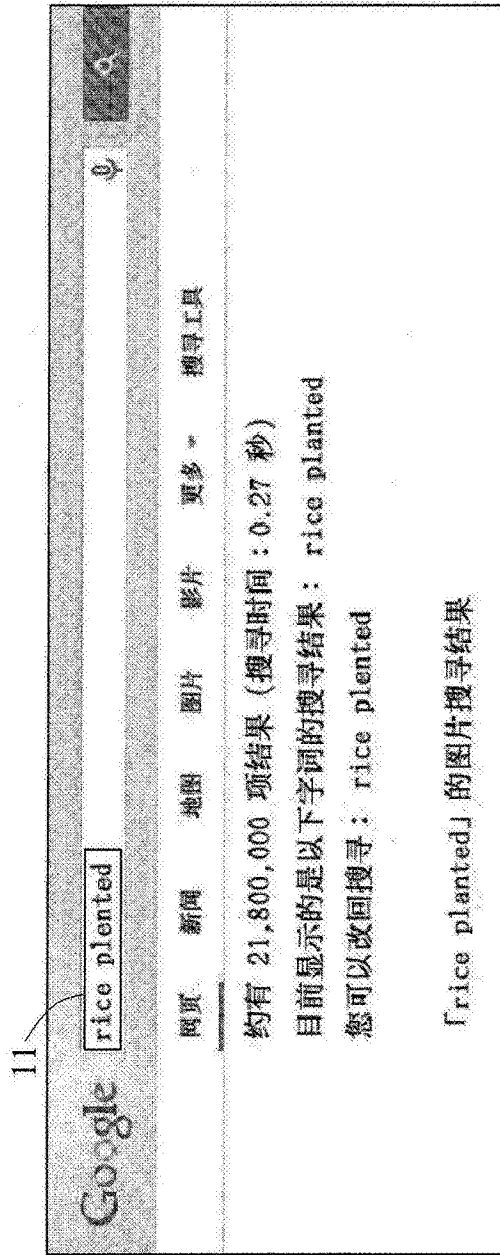


图 2D

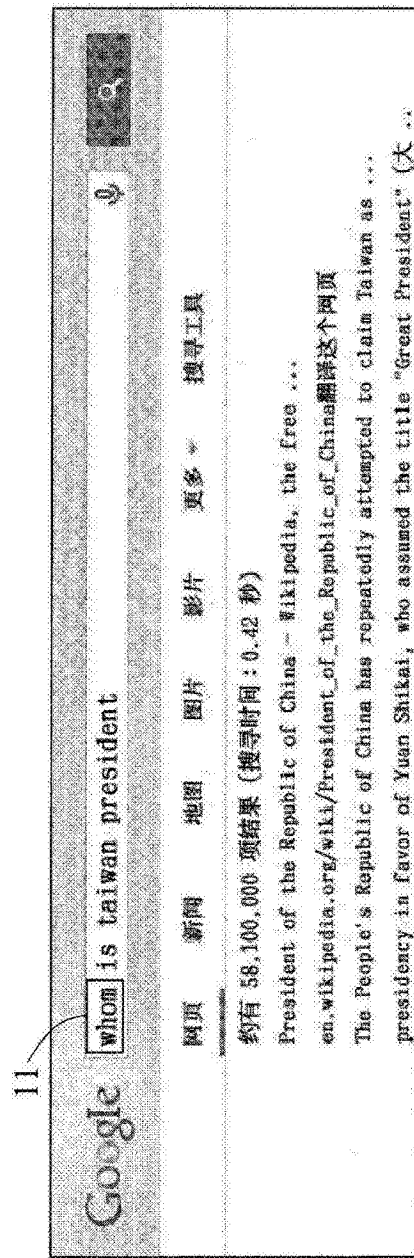


图 2E

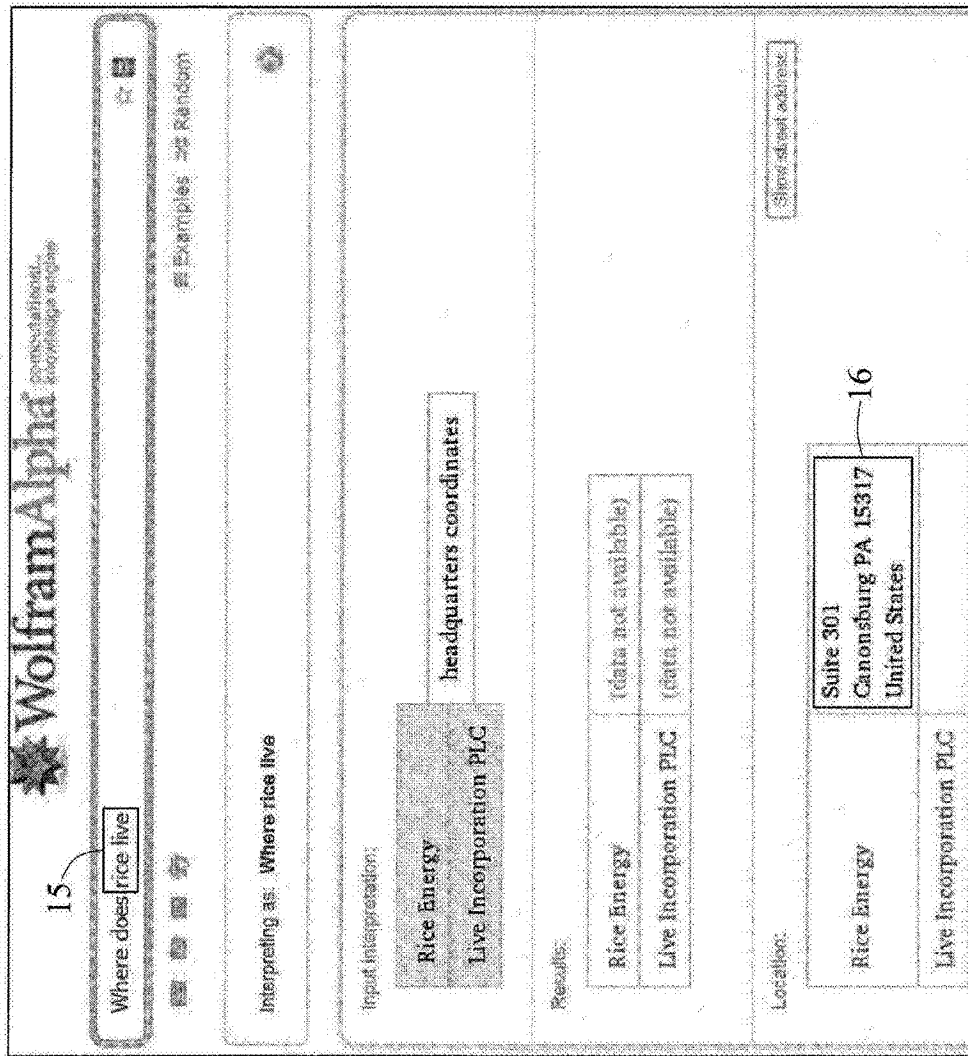


图 2F

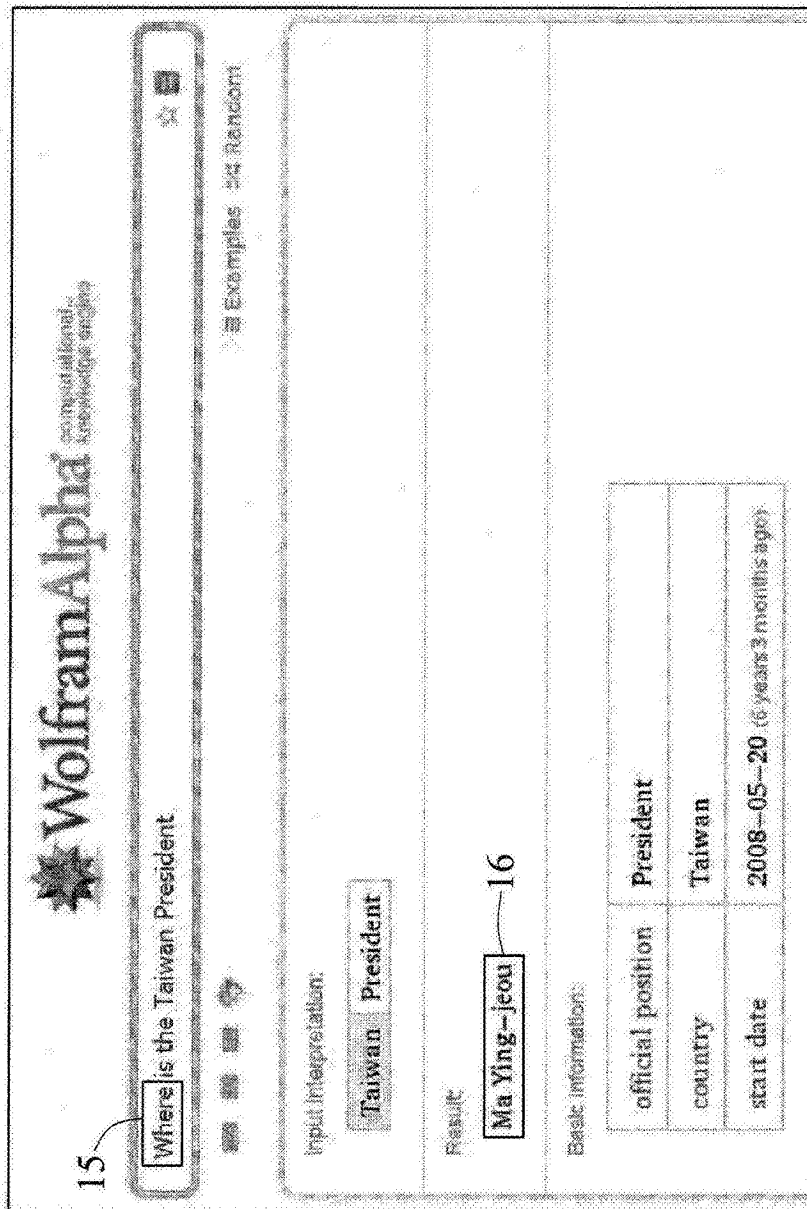


图 2G

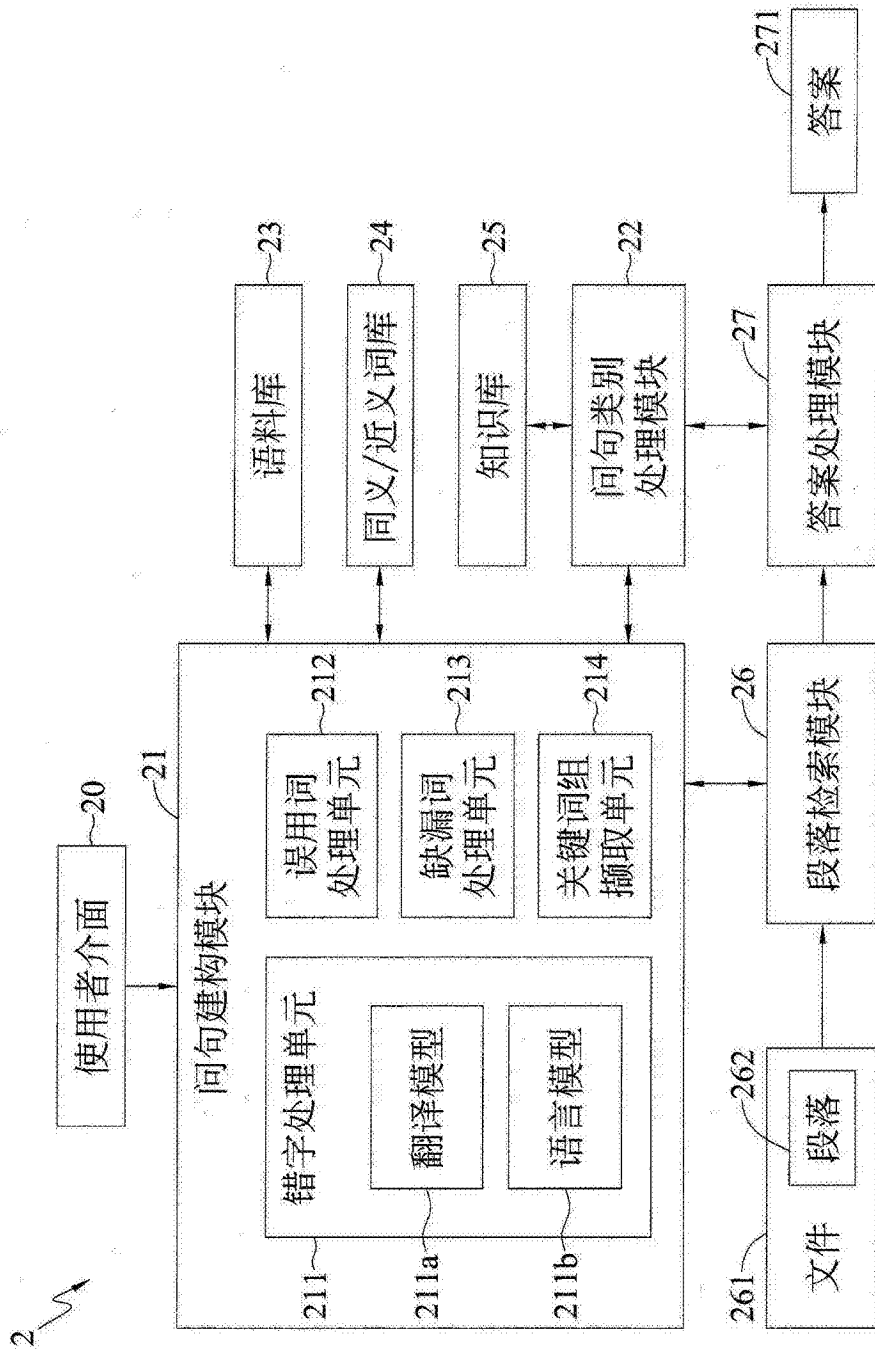


图 3

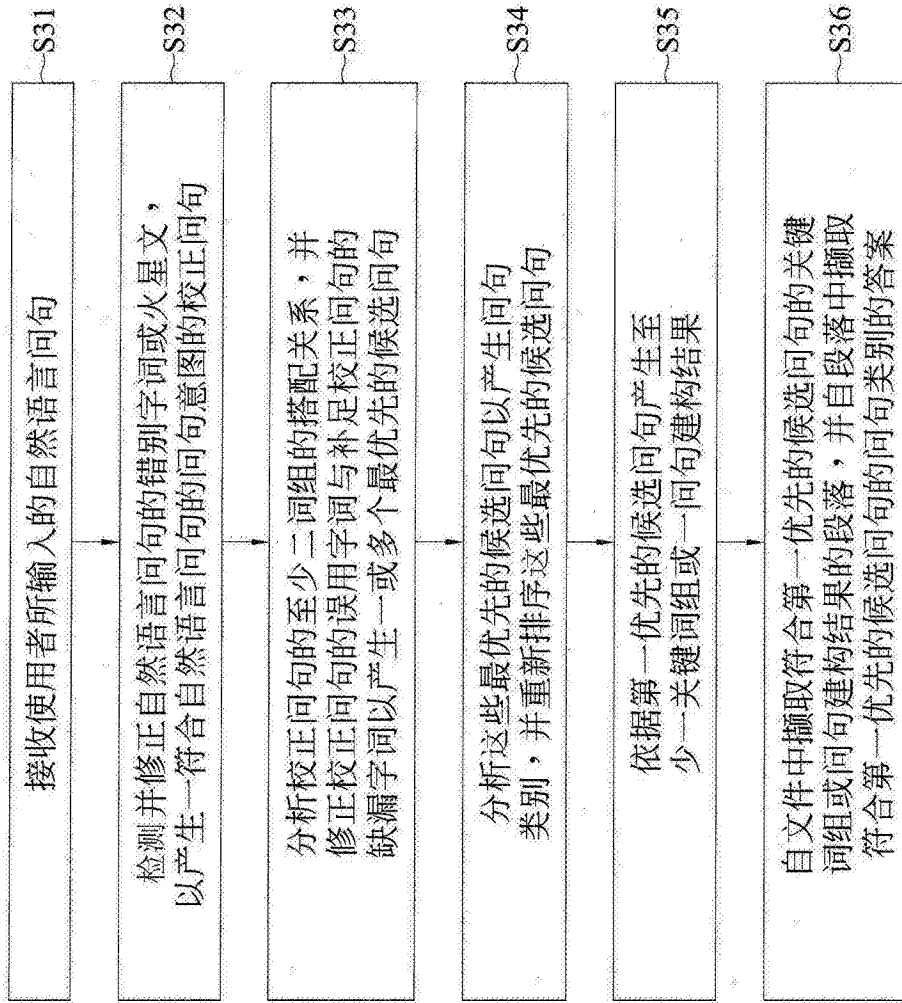


图 4

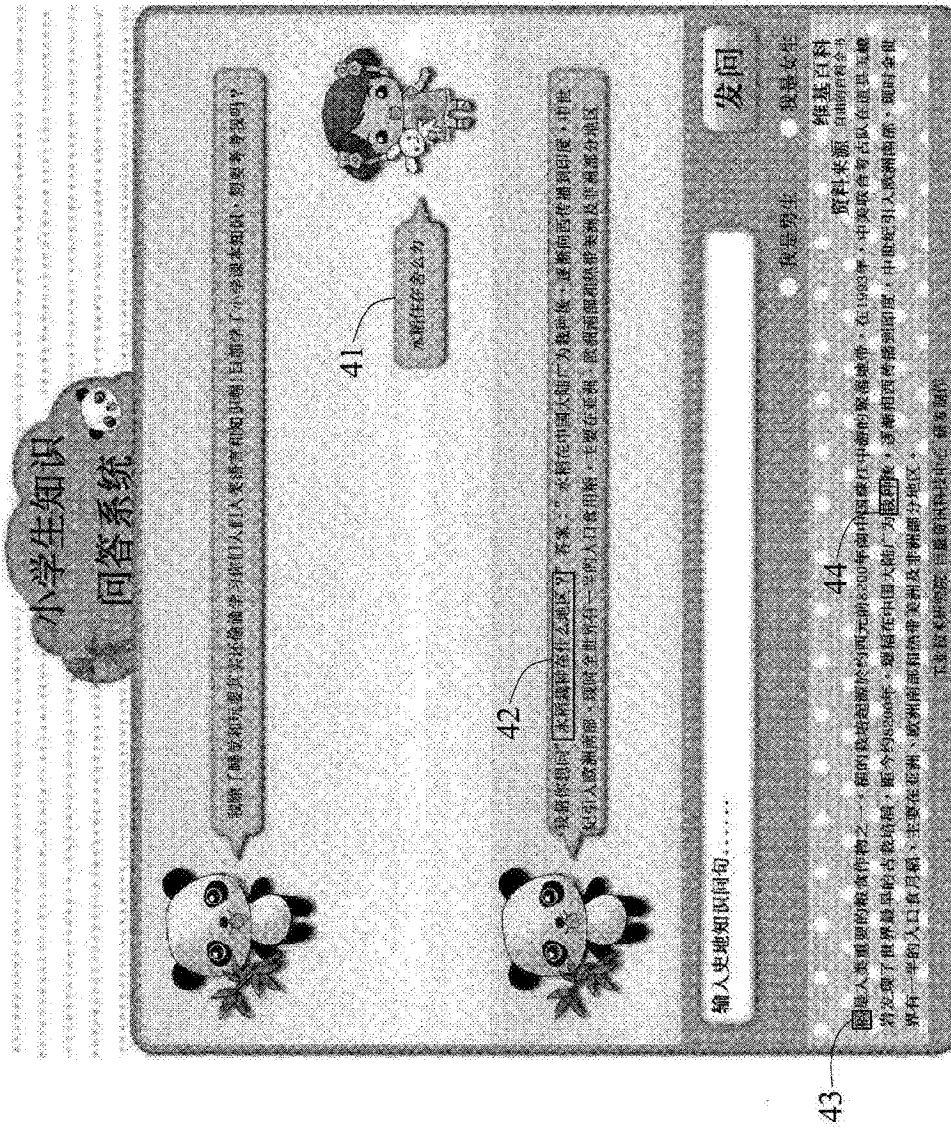


图 5