(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
2008 年10 月30 日 (30.10.2008)

PCT

(10) 国际公布号
WO 2008/128423 A1

(51) 国际专利分类号:
G10L 15/22 (2006.01) *G10L 15/18* (2006.01)

(21) 国际申请号: PCT/CN2008/000764

(22) 国际申请日: 2008 年4 月15 日 (15.04.2008)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
200710074112.1
2007 年4 月19 日 (19.04.2007) CN

(71) 申请人 (对除美国外的所有指定国): **深圳先进技术研究院(SHENZHEN INSTITUTE OF ADVANCED TECHNOLOGY)** [CN/CN]; 中国广东省深圳市南山区蛇口南海大道1019号南山医疗器械产业园A座3层, Guangdong 518067 (CN)。

(72) 发明人; 及

(75) 发明人/申请人 (仅对美国): **徐扬生(XU, Yangsheng)** [CN/CN]; 中国广东省深圳市南山区蛇口南海大道1019号南山医疗器械产业园A座3层, Guangdong 518067 (CN)。 **李崇国(LI, Chong Guo)** [CN/CN]; 中国广东省深圳市南山区蛇口南海大道1019号南山医疗器械产业园A座3层, Guangdong 518067 (CN)。 **阎镜予(YAN, Jingyu)** [CN/CN]; 中国广东省深圳市南山区蛇口南海大道1019号南山医疗器械产业园A座3层, Guangdong 518067 (CN)。 **程俊(CHENG, Jun)** [CN/CN]; 中国广东省深圳市南山区蛇口南海大道1019号南山医疗器械产业园A座3层, Guangdong 518067 (CN)。 **吴新宇(WU, Xinyu)** [CN/CN]; 中国

[见续页]

(54) **Title:** AN INTELLIGENT DIALOG SYSTEM AND A METHOD FOR REALIZATION THEREOF

(54) **发明名称:** 一种智能聊天系统及其实现方法

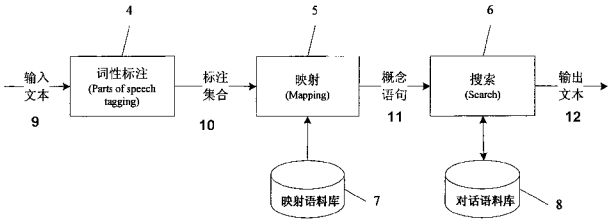


图 3 / Fig. 3

4 PARTS OF SPEECH TAGGING
5 MAPPING
6 SEARCH
7 MAPPING CORPUS
8 DIALOG CORPUS
9 INPUT TEXT
10 TAG SET
11 CONCEPT SENTENCE
12 OUTPUT TEXT

(57) Abstract: An intelligent dialog system includes a text-comprehending answering module (2) which is used to obtain an output text based on an input text. The module includes a word dividing unit, a mapping corpus (7), a mapping unit (5), a dialog corpus (8) and a searching unit (6). The word dividing unit is used to tag parts of speech for said input text (4), and obtain a word set with parts of speech tagging. The mapping corpus (7) is used to set and store a mapping relationship between key words and concept sentences. The mapping unit (5) is used to search said mapping corpus (7) based on said word set and map to obtain a concept sentence (5). The dialog corpus (8) is used to set and store mapping relationship between concept sentences and output text. The searching unit (6) is used to search said dialog corpus (8) based on said concept sentence and obtain the output text.

(57) 摘要:

一种智能聊天系统，包括用于根据输入文本得到输出文本的文本理解回答模块（2）；该模块包括分词单元、映射语料库（7）、映射单元（5）、对话语料库（8）和搜索单元（6）；分词单元用于对所述输入文本进行词性标注，得到具有词性标注的词语集合；映射语料库（7）用于建立和存储关键词到概念语句的映射关系；映射单元（5）用于根据所述词语集合，搜索所述映射语料库（7），映射得到概念语句（5）；对话语料库（8）用于建立和存储概念语句到输出文本的映射关系；搜索单元（6）用于根据所述概念语句，搜索所述对话语料库（8），映射得到输出文本。

WO 2008/128423 A1

广东省深圳市南山区蛇口南海大道1019号南山医疗器械产业园A座3层, Guangdong 518067 (CN)。

(74) 代理人：广东国欣律师事务所(GUANGDONG GUOXIN LAW FIRM); 中国广东省深圳市罗湖区红岭中路1010号国际信托大厦六楼, Guangdong 518008 (CN)。

(81) 指定国（除另有指明，要求每一种可提供的国家保护）：AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), 欧洲 (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)。

本国际公布：

— 包括国际检索报告。