(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification:
G06F 17/27 (2006.01)

(21) International Application Number:
PCT/US2017/017318

(22) International Filing Date:
10 February 2017 (10.02.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/293,570   10 February 2016 (10.02.2016)   US
62/457,069   9 February 2017 (09.02.2017)   US
15/429,094   9 February 2017 (09.02.2017)   US

(72) Inventors; and
(71) Applicants : ZHANG, Yong [CN/US]; 1465 Edvalson St., Dept. 2401, Ogden, Utah 84408 (US). LIU, Ning [CN/CN]; 5-1-1102, No. 8, Huahugou, Chaoyang, Beijing, Beijing 100190 (CN).

(74) Agents: STRINGHAM, John C. et al.; 60 East South Temple, Suite 1000, Salt Lake City, Utah 84111 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: 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) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (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,
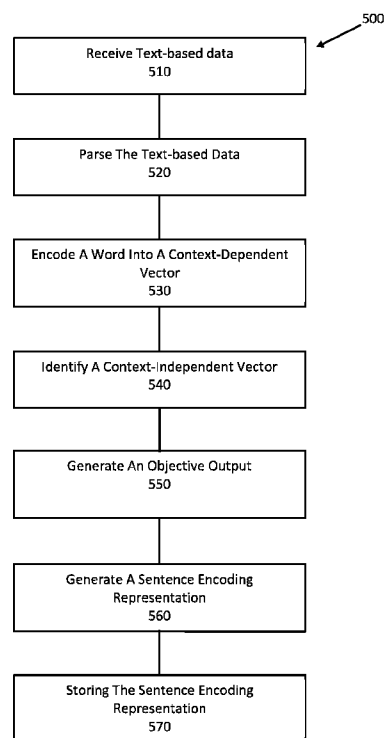
(54) Title: ELECTRONIC MESSAGE INFORMATION RETRIEVAL SYSTEM

Fig. 5

(57) Abstract: A computer system for parsing bulk message information into intelligent question retrieval models receives text-based data associated with a particular user. The system encodes the word into a context-dependent vector, wherein the context-dependent vector indicates the meaning of the word across a semantic space. The system also identifies within a context-independent database a context-independent vector that is associated with the word. Further, the system generates an objective output by combining the context-dependent vector and the context-independent vector. Further still, the system generates a sentence encoding representation by processing at least a portion of the text-based data through a high-level feature embedded convolutional semantic model to generate numerical representations of questions and answers within the text-based dataset. The sentence encoding representation is generated at least in part based upon the objective output.

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(88) Date of publication of the international search report:
19 October 2017

Published:

— *with international search report (Art. 21(3))*

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

IPC - G06F 17/27 (2017.01)
CPC - G06F 17/2785, 17/30684; Y10S 707/99933

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)

See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2011/0270604 A1 (QI, Y et al.) 3 November 2011; abstract; paragraphs [0007], [0016], [0018], [0020], [0054], [0055], [0068], [0071], [0081]-[0085], [0097]; claim 6	1-4, 6, 7, 9-14, 16-19
----		----
Y		5, 8, 15, 20
Y	US 2013/0138641 A1 (KOROLEV, D et al.) 30 May 2013; paragraphs [0006], [0025]	5, 15
Y	US 7444356 B2 (CALISTRI-YEH, R et al.) 28 October 2008; column 3, lines 32-34; claim 3	8
Y	US 2014/0095418 A1 (SCHOLAND, A et al.) 3 April 2014; claim 24	20

☐ 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	
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"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
"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	"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 May 2017 (14.05.2017)	**08 JUN 2017**

Name and mailing address of the ISA/	Authorized officer
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450	Shane Thomas
Facsimile No. 571-273-8300	PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

| Box No. II | Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet) |

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
   because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
   because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
   because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

| Box No. III | Observations where unity of invention is lacking (Continuation of item 3 of first sheet) |

This International Searching Authority found multiple inventions in this international application, as follows:
See extra sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☒ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (January 2015)

-***-Continued from Box No. III Observations where unity of invention is lacking-***-

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fee must be paid.

Group I: Claims 1-19 are directed towards parsing bulk message information, comprising generating a sentence encoding representation.
Group II: Claim 20 is directed towards parsing bulk message information, comprising identifying adjacency pairs.

The inventions listed as Groups I-II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The special technical features of Group I include at least parsing a word from the text-based data; generating a sentence encoding representation by processing at least a portion of the text-based data through a high-level feature embedded convolutional semantic model to generate numerical representations of questions and answers within the text-based dataset, wherein the sentence encoding representation is generated at least in part based upon the objective output., which are not present in Group II.

The special technical features of Group II include at least identifying adjacency pairs within the words, wherein the adjacency pairs comprise two parts that are communicated by different parties in which a first part is conditionally related to a second part, which are not present in Group I.

The common technical features shared by Groups I-II are a computer system for parsing bulk message information into intelligent question retrieval models, comprising: one or more processors; and one or more computer-readable media having stored thereon executable instructions that when executed by the one or more processors configure the computer system to perform at least the following: receive text-based data associated with a particular user; encode a word into a context-dependent vector, wherein the context-dependent vector indicates the meaning of the word across a semantic space; identify within a context-independent database a context• independent vector that is associated with the word, wherein the context• independent database comprises mappings of various words to various context-independent vectors based upon definitions of the various words; and generate an objective output by combining the context• dependent vector and the context-independent vector.

However, these common features are previously disclosed by US 5,455,889 A to BAHL et al. (hereinafter "Bahl"). Bahl discloses a computer system for parsing bulk message information into intelligent question retrieval models, comprising: one or more processors (CPU 116; Fig. 1); and one or more computer-readable media having stored thereon executable instructions that when executed by the one or more processors configure the computer system (micro instructions code 110, inherently residing on computer-readable media), executed by the CPU; Fig. 1; col 4, ln 10-23) to perform at least the following: receive text-based data associated with a particular user (the system having an input comprising a sequence of phones from a training text; claim 1); encode a word into a context-dependent vector (model a word as a sequence of context-dependent models; col 4, ln 3-9), wherein the context-dependent vector indicates the meaning of the word across a semantic space (each word maps onto a string of phonemes that make up a word in the English language (semantic space) and determine the most likely label sequence for each word in the system's vocabulary; col 7, ln 19-22); identify within a context-independent database a context•independent vector that is associated with the word (context-independent labelling of vectors occurs in the context-independent labeler, and a label alphabet is stored in memory; col 5, ln 13-51), wherein the context•independent database comprises mappings of various words to various context-independent vectors based upon definitions of the various words (each acoustic parameter vector is associated with a context-independent label and after the correspondence between the labels and each phone is known, it follows that the acoustic parameter vector corresponding to each phone in the text string is also known; col 6, ln 38-63); and generate an objective output by combining the context•dependent vector and the context-independent vector (the context-independent label prototype vectors are based upon each individual phoneme without regard to the preceding or following phoneme, and the phonetic context of the target phoneme is the combination of a target phoneme and neighboring phonemes; col 6, ln 64 to col 7, ln 28).

Since the common technical features are previously disclosed by the Bahl reference, these common features are not special and so Groups I-II lack unity.