



(51) International Patent Classification:
G06F 17/20 (2006.01) *G10L 15/26* (2006.01)
G06F 17/30 (2006.01)

(21) International Application Number:
PCT/SG2012/000106

(22) International Filing Date:
28 March 2012 (28.03.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
201102176-3 28 March 2011 (28.03.2011) SG

(71) Applicant (for all designated States except US):
**AGENCY FOR SCIENCE, TECHNOLOGY AND RE-
SEARCH** [SG/SG]; 1 Fusionpolis Way, #20-10 Connexis,
Singapore 138632 (SG).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **HUANG, Chien-Lin**
[CN/SG]; c/o PKM, Institute for Infocomm Research, 1
Fusionpolis Way, #21-01 Connexis, South Tower, Singa-
pore 138632 (SG). **MA, Bin** [SG/SG]; c/o PKM, Institute
for Infocomm Research, 1 Fusionpolis Way, #21-01 Con-

nexis, South Tower, Singapore 138632 (SG). **LI, Haizhou**
[SG/SG]; c/o PKM, Institute for Infocomm Research, 1 Fu-
sionpolis Way, #21-01 Connexis, South Tower, Singapore
138632 (SG).

(74) Agent: **ELLA CHEONG SPRUSON & FERGUSON**
(SINGAPORE) PTE LTD; P.O. Box 1531, Robinson
Road Post Office, Singapore 903031 (SG).

(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, BR, BW, BY, BZ,
CA, CH, CL, 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, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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,

[Continued on next page]

(54) Title: A METHOD, AN APPARATUS AND A COMPUTER-READABLE MEDIUM FOR INDEXING A DOCUMENT FOR DOCUMENT RETRIEVAL

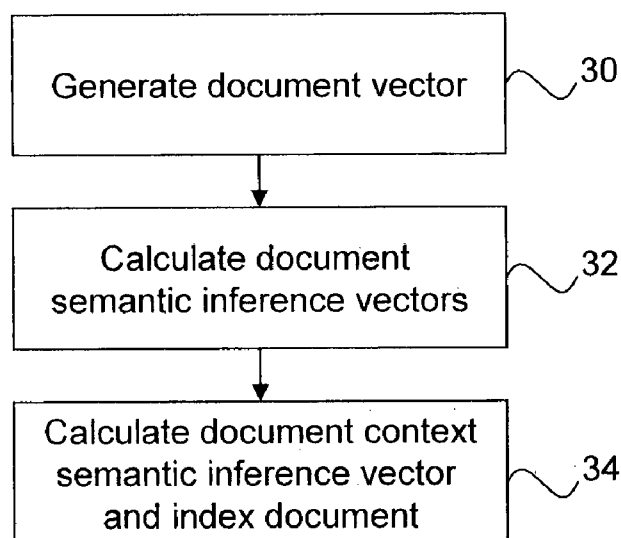


Figure 4B

(57) Abstract: Various embodiments provide a method for indexing a document for document retrieval. The document may include: generating a document vector indicating if each of a plurality of terms are present in the document; calculating a document semantic inference vector for each of the plurality of terms present in the document using the document vector and a semantic relation matrix, the semantic relation matrix identifying semantic relationships between different terms of the plurality of terms; and, indexing the document using a document semantic context inference vector calculated based on the document semantic inference vectors. Various embodiments provide a corresponding apparatus and computer readable medium.



LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, **Published:**

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, — *with international search report (Art. 21(3))*
GW, ML, MR, NE, SN, TD, TG).