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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 December 2005 (08.12.2005)

PCT

(10) International Publication Number
WO 2005/116866 A1

(51) International Patent Classification[7]: G06F 17/28, 15/18

(21) International Application Number: PCT/SG2005/000169

(22) International Filing Date: 28 May 2005 (28.05.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
200403036-7 28 May 2004 (28.05.2004) SG

(71) Applicant *(for all designated States except US)*: **AGENCY FOR SCIENCE, TECHNOLOGY AND RESEARCH** [SG/SG]; 20 Biopolis Way, #07-01 Centros, Singapore, 138668 (SG).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SU, Jian** [CN/SG]; c/o Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore, 119613 (SG). **SHEN, Dan** [CN/DE]; Department of Computational Linguistics Saarland University Postfach 15 11 50, 66041 Saarbruecken (DE). **ZHANG, Jie** [CN/SG]; c/o Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613 (SG). **ZHOU, Guo, Dong** [CN/SG]; c/o Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613 (SG).

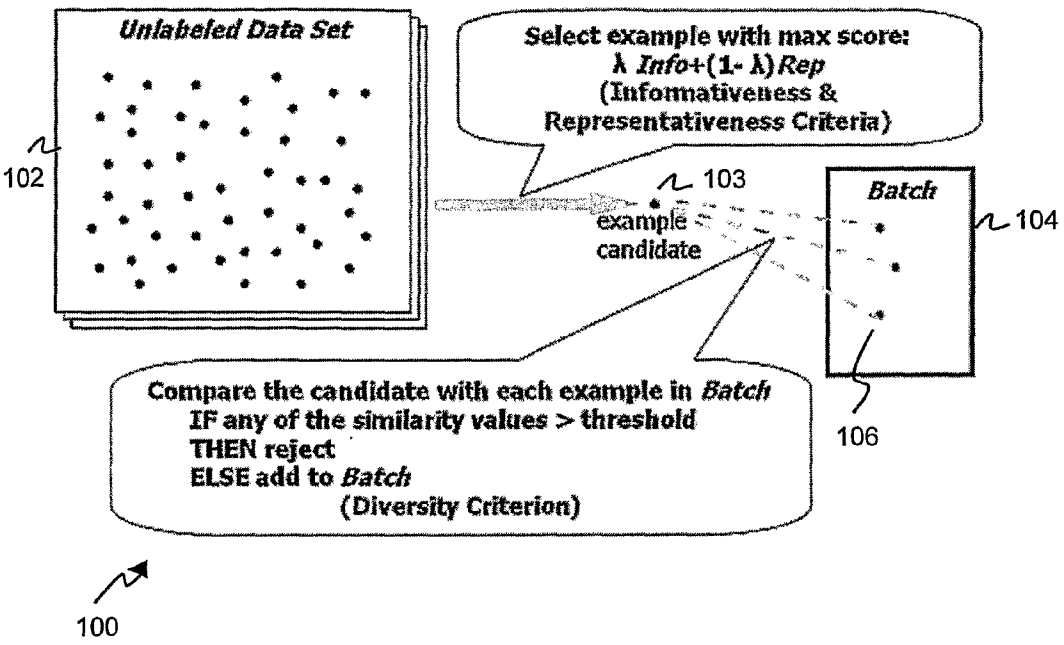
(74) Agent: **ELLA CHEONG SPRUSON & FERGUSON (SINGAPORE) PTE LTD**; P.O. Box 1531, Robinson Road Post Office, Singapore 903031 (SG).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,

[Continued on next page]

(54) Title: METHOD AND SYSTEM FOR WORD SEQUENCE PROCESSING



(57) Abstract: A method and system of conducting named entity recognition. One method comprises selecting one or more examples for human labelling, each example comprising a word sequence containing a named entity and its context; and retraining a model for the named entity recognition based on the labelled examples as training data.

WO 2005/116866 A1

ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:
— *with international search report*