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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 May 2007 (31.05.2007)

PCT

(10) International Publication Number
WO 2007/062156 A3

(51) International Patent Classification:
G06F 17/30 (2006.01)

(21) International Application Number:
PCT/US2006/045269

(22) International Filing Date:
22 November 2006 (22.11.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/739,270 23 November 2005 (23.11.2005) US

(71) Applicant *(for US only)*: **DUN & BRADSTREET INC.** [US/US]; 103 JFK PARKWAY, Short Hills, NJ 07078 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SCRIFFIGNANO, Anthony** [US/US]; 37 WOODROW PLACE, West Caldwell, NJ 07006 (US). **NEDD, Kevin** [US/US]; 65 KIM LANE, Long Valley, NJ 07853 (US). **SHAO, Pei** [US/US]; 3F-2, NO. 17, LANE 43, SHAN MING ROAD, Taipei (TW). **PENG, Gan, Sim** [CN/CN]; 18 RAMBIA ROAD, Singapore 424337 (SG). **LU, Sarah** [CN/CN]; ROOM 503, NO 42 LV YIN, YUAN, ALLEY 588, BAO CHUN ROAD, Shanghai 201100 (CN).

(74) Agent: **GREELEY, Paul, D.**; Ohlandt, Greeley, Ruggiero & Perle, One Landmark Square, 10th Floor, Stamford, CT 06901 (US).

(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, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, 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.

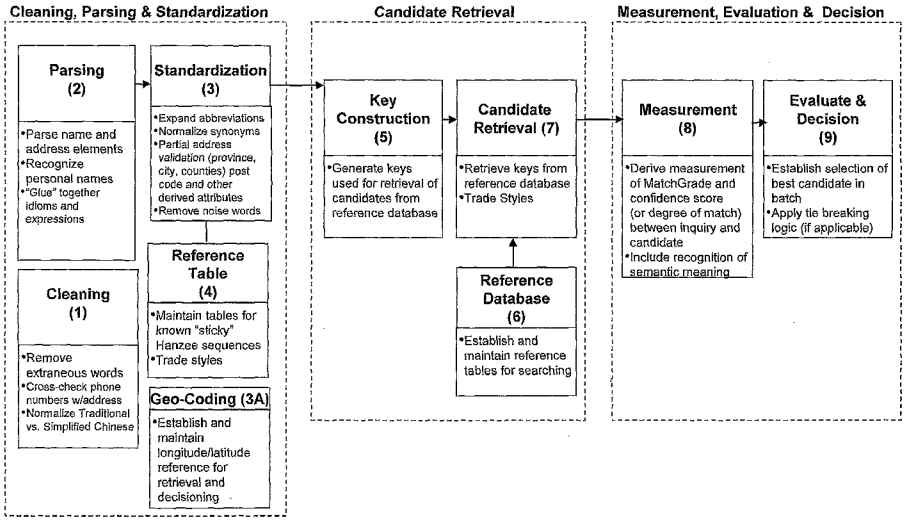
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, 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).

Published:
— *with international search report*

(88) Date of publication of the international search report:
8 November 2007

[Continued on next page]

(54) Title: SYSTEM AND METHOD FOR SEARCHING AND MATCHING DATA HAVING IDEOGRAMMATIC CONTENT



(57) **Abstract:** A method of searching and matching non-phonetic or ideogrammatic input data to stored data, including the steps of receiving input data comprising a search string having a plurality of elements, converting a subset of the elements into a set of terms, generating an optimized plurality of keys from the set of terms, retrieving stored data based on the optimized keys corresponding to most likely candidates for match, and selecting a best match from the plurality of candidates. At least some of the ideogrammatic elements form part of an ideogrammatic writing system. The method may also include dividing the search string into a plurality of overlapping sub-segments and identifying sub-segments having inferred semantic meaning as well as sub-segments having no semantic meaning in the ideogrammatic writing system, and using the various sub- segments to generate the optimized keys.

WO 2007/062156 A3

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.