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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 July 2009 (09.07.2009)

PCT

(10) International Publication Number
WO 2009/085555 A3

(51) International Patent Classification:
G06F 7/00 (2006.01)

(21) International Application Number:
PCT/US2008/085585

(22) International Filing Date:
4 December 2008 (04.12.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/967,494 31 December 2007 (31.12.2007) US

(71) Applicant *(for all designated States except US)*: **MASTERCARD INTERNATIONAL INCORPORATED** [US/US]; 2000 Purchase Street, Purchase, NY 10577 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **MERZ, Christopher, J.** [US/US]; 10 Flowery Branch Place, Wildwood, MO 63040 (US). **MCGEEHAN, Thomas** [US/US]; 1111 Sandstone Drive, St. Louis, MO 63146 (US).

(74) Agents: **FITZGERALD, Daniel, M.** et al.; Armstrong Teasdale Llp, One Metropolitan Square, Suite 2600, St. Louis, MO 63102 (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, 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, ST, 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, 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).

Published:

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

(88) Date of publication of the international search report:
7 January 2010

(54) Title: METHODS AND SYSTEMS FOR IMPLEMENTING APPROXIMATE STRING MATCHING WITHIN A DATABASE

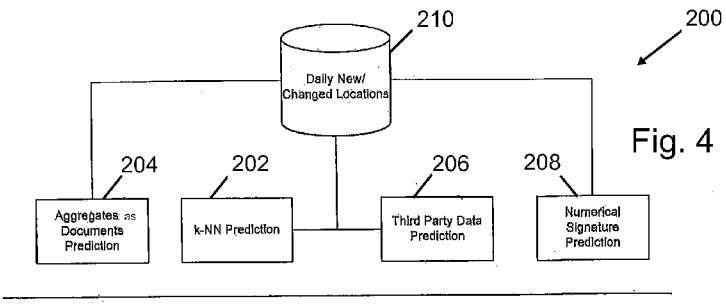


Fig. 4

(57) Abstract: A computer-based method for character string matching of a candidate character string with a plurality of character string records stored in a database is described. The method includes a) identifying a set of reference character strings in the database, the reference character strings identified utilizing an optimization search for a set of dissimilar character strings, b) generating an n-gram representation for one of the reference character strings in the set of reference character strings, c) generating an n-gram representation for the candidate character string, d) determining a similarity between the n-gram representations, e) repeating steps b) and d) for the remaining reference character strings in the set of identified reference character strings, and f) indexing the candidate character string within the database based on the determined similarities between the n-gram representation of the candidate character string and the reference character strings in the identified set.

WO 2009/085555 A3