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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 October 2009 (29.10.2009)

PCT

(10) International Publication Number
WO 2009/132263 A3

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

(21) International Application Number:
PCT/US2009/041649

(22) International Filing Date:
24 April 2009 (24.04.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/047,570 24 April 2008 (24.04.2008) US
61/053,202 14 May 2008 (14.05.2008) US

(71) Applicant *(for all designated States except US)*: **LEXISNEXIS RISK & INFORMATION ANALYTICS GROUP, INC.** [US/US]; 6601 Park Of Commerce Boulevard, Boca Raton, FL 33487 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **BAYLISS, David** [US/US]; 7219 Serrano Terrace, Delray Beach, FL 33446 (US).

(74) Agents: **LEANING, Jeffrey, S.** et al.; Hunton & Williams, Intellectual Property Department, 1900 K Street, N.W., Washington, DC 20006 (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,

[Continued on next page]

(54) Title: DATABASE SYSTEMS AND METHODS

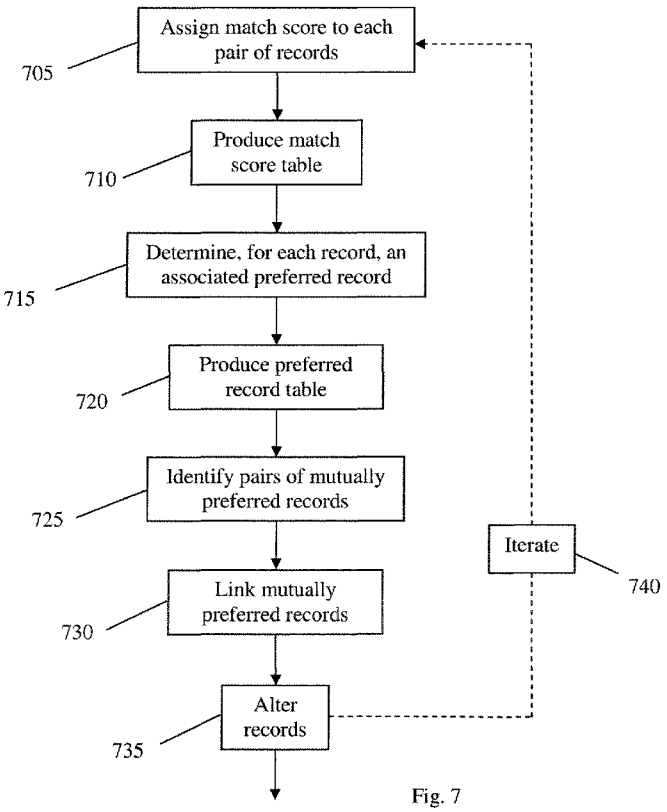

Fig. 7

(57) Abstract: Disclosed is a system for, and method of, calculating parameters used to determine whether records and entity representations should be linked. The system and method apply iterative techniques such that parameters from each linking iteration are used in the next linking iteration. The system and method need no human interaction in order to calibrate and utilize record matching formulas used for the linking decisions.

WO 2009/132263 A3

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, MK, 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

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 09/41649

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06F 7/00 (2009.01)
USPC - 707/7
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)
USPC: 707/7

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC: 707/1, 3, 4, 6, 100, 101

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST(USPT,PGPB,EPAB,JPAB); Google Scholar
Search Terms: database, score, linking, matching, record, threshold. field, value, weight

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2003/0126156 A1 (Stoltenberg et al.) 03 July 2003 (03.07.2003), para [0010], [0011], [0015], [0053], [0060]	1-5, 8, 9, 11-16, 19-20, 22, 23 --- 6, 7, 10, 17, 18, 21, 24
Y	US 6,374,241 B1 (Lamburt et al.) 16 April 2002 (16.04.2002), col 64, ln 44-50	6, 7, 17, 18
Y	US 2002/0156793 A1 (Jaro) 24 October 2002 (24.10.2002), para [0074]	10, 21
Y	US 6,968,335 B2 (Bayliss et al.) 22 November 2005 (22.11.2005), col 32, ln 49-52	24

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"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
"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)
"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
"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
"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
"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
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
02 August 2009 (02.08.2009)	13 AUG 2009
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (April 2007)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 09/41649

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---

see p.8 of this form

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.:
1-24

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.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 09/41649

Continuation of Box No. 3 - Unity of Invention

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 searched, the appropriate additional search fees must be paid.

Group I, Claims 1-24, drawn to a computer implemented method of linking a first record in a database to a second record in a database upon a determination that the first record and the second record correspond to the same individual, the method comprising calculating a match score based on a plurality of terms, each of the plurality of terms corresponding to a different field common to the first record and the second record, each of the plurality of terms comprising: (1) a probability that a field value in a corresponding field of the first record matches a field value in a corresponding field in the second record, and (2) a quantity based on a corresponding one of the plurality of match probabilities

Group II, Claims 25-60, 164-184, drawn to a computer implemented iterative process for generating entity representations in a computer implemented database using a record matching formula and for generating parameters for the record matching formula, the database comprising a plurality of records, each record comprising a plurality of fields, each field capable of containing a field value, wherein at least a portion of parameters for the record matching formula are configured for a particular field value appearing in a selected field of at least one record, the process comprising: calculating a revised field value weight, the revised field value weight reflecting a likelihood that an arbitrary entity representation in the database comprises the particular field value in the selected field of a record in the arbitrary entity representation

Group III, Claims 61-76, drawn to a computer implemented iterative process for generating entity representations in a computer implemented database using a record matching formula and for generating parameters for the record matching formula, the database comprising a plurality of records, each record comprising a plurality of fields, each field capable of containing a field value, wherein at least a portion of parameters for the record matching formula are configured for a particular field value associated with a selected field, and wherein the process provides for linking records or entity representations with non-identical field values, the process comprising: applying a symmetric, reflexive and transitive function to each field value in the selected field of each of a plurality of records in the database, whereby applying the symmetric, reflexive and transitive function to each field value in the selected field of each of a plurality of records in the database defines a partition of the plurality of records

Group IV, Claims 77-92 drawn to a computer implemented iterative process for generating entity representations in a computer implemented database using a record matching formula and for generating parameters for the record matching formula, the database comprising a plurality of records, each record comprising a plurality of fields, each field capable of containing a field value, wherein at least a portion of the parameters for the record matching formula are specific to a particular plurality of field values associated with a particular plurality of fields, the process comprising: adding, in the database, a supplemental field to each of the plurality of records; populating each supplemental field of each of the plurality of records with a supplemental field value, each supplemental field value representative of field values from the particular plurality of fields of that record

Group V, Claims 93-114 drawn to a computer implemented method of detecting and treating one or more field values as null field values in records of an electronic database, the computer implemented method comprising: determining, using a programmed computer, a frequency for each field value of the field, the frequency comprising an amount of the records in which the field value of the field appears; calculating, using a programmed computer, a critical frequency for the field based on the frequency of each field value of the field; and treating, using a programmed computer, each field value of the field with a frequency that is greater than the critical frequency of the field as a null field value and each field value of the field with a frequency that is less than the critical frequency of the field as a non-null field value.

Group VI, Claims 115-132 drawn to a computer implemented iterative process for generating entity representations by identifying and linking related records in a computer implemented database using a record matching formula, each record and entity representation electronically stored in the database, each record comprising a plurality of fields, each field capable of containing a field value, the process comprising: identifying mutually preferred pairs of records from the plurality of records, each mutually preferred pair of records consisting of a first record and a second record, the first record consisting of a preferred record associated with the second record and the second record consisting of a preferred record associated with the first record.

Group VII, Claims 133-163, drawn to a computer implemented method of identifying a set of fields applicable to partition a plurality of records in an electronic database into one or more blocks based on a desired block size and independent of specific queries against the database, the method comprising: determining a set of fields wherein a product of the associated field probabilities and the number of records in the database is approximately equal to the desired block size; and outputting the set of fields, the set of fields independent of specific queries against the database.

---see next extra sheet---

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 09/41649

Continuation of Box No. 3 - Unity of Invention

The inventions listed as Groups I-VII 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 feature of the Group I invention is calculating a match score based on a plurality of terms, each of the plurality of terms corresponding to a different field common to the first record and the second record. The special technical feature of the Group II invention is calculating a revised field value weight, the revised field value weight reflecting a likelihood that an arbitrary entity representation in the database comprises the particular field value in the selected field of a record in the arbitrary entity representation. The special technical feature of the Group III invention is applying a symmetric, reflexive and transitive function to each field value in the selected field of each of a plurality of records in the database, whereby applying the symmetric, reflexive and transitive function to each field value in the selected field of each of a plurality of records in the database defines a partition of the plurality of records. The special technical feature of the Group IV invention is adding, in the database, a supplemental field to each of the plurality of records; populating each supplemental field of each of the plurality of records with a supplemental field value, each supplemental field value representative of field values from the particular plurality of fields of that record. The special feature of the Group V invention is a critical frequency for the field based on the frequency of each field value of the field; and treating, using a programmed computer, each field value of the field with a frequency that is greater than the critical frequency of the field as a null field value and each field value of the field with a frequency that is less than the critical frequency of the field as a non-null field value. The special feature of the Group VI invention is identifying mutually preferred pairs of records from the plurality of records, each mutually preferred pair of records consisting of a first record and a second record, the first record consisting of a preferred record associated with the second record and the second record consisting of a preferred record associated with the first record. The special feature of the Group VII invention is determining a set of fields wherein a product of the associated field probabilities and the number of records in the database is approximately equal to the desired block size. None of these special technical features are common to the other groups, nor do they correspond to a special technical feature in the other groups. Therefore, unity of invention is lacking.