(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2014 年 9 月 12 日(12.09.2014)

WIPO | PCT

(10) 国際公開番号
WO 2014/136810 A1

(51) 国際特許分類:
G06F 17/30 (2006.01)

(21) 国際出願番号: PCT/JP2014/055548

(22) 国際出願日: 2014 年 3 月 5 日(05.03.2014)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(30) 優先権データ:
特願 2013-045566 2013 年 3 月 7 日(07.03.2013) JP

(71) 出願人: 日本電気株式会社(NEC CORPORATION) [JP/JP]; 〒1088001 東京都港区芝五丁目７番１号 Tokyo (JP).

(72) 発明者: 土田正明(TSUCHIDA, Masaaki); 〒1088001 東京都港区芝五丁目７番１号日本電気株式会社内 Tokyo (JP). 石川開(ISHIKAWA, Kai); 〒1088001 東京都港区芝五丁目７番１号日本電気株式会社内 Tokyo (JP).

(74) 代理人: 特許業務法人ワンディーＩＰパートナーズ(ONEDEE IP PARTNERS); 〒5320003 大阪府大阪市淀川区宮原四丁目１番４号新大阪センタービル Osaka (JP).

(81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, 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, IR, IS, JP, KE, KG, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, 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.

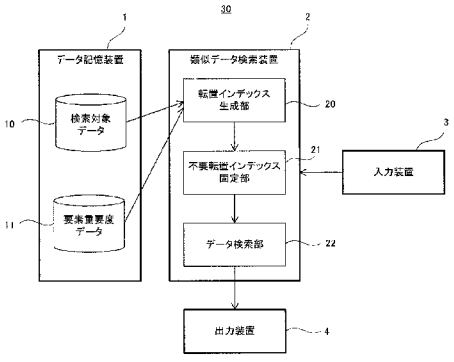
(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:
— 国際調査報告（条約第 21 条(3)）

WO 2014/136810 A1

(54) **Title**: SIMILAR DATA SEARCH DEVICE, SIMILAR DATA SEARCH METHOD, AND COMPUTER-READABLE STORAGE MEDIUM

(54) 発明の名称 : 類似データ検索装置、類似データ検索方法、及びコンピュータ読み取り可能な記録媒体



1 Data storage device
10 Search target data
11 Element importance data
2 Similar data search device
20 Inverted index generating unit
21 Unnecessary inverted index fixing unit
22 Data search unit
3 Input device
4 Output device

(57) **Abstract**: This similar data search device (2) is provided with: an inverted index generating unit (20) which determines a size range of the search target set of each inverted index such that the number of search target sets is greater than or equal to a set number, separates the search target sets for each size range, and generates inverted indexes; an unnecessary inverted index identifying unit (21) which, on the basis of the size of search target sets and the threshold value set for the degree of similarity between the sets, calculates a necessary condition for the degree of similarity to be greater than or equal to the threshold value, and identifies, as inverted indexes not requiring search, the inverted indexes other than those for which the minimum value of the set size fulfills the condition; and a data search unit (22) which performs a search on the inverted indexes not identified.

(57) 要約: 類似データ検索装置２は、検索対象集合の数が設定個数以上になるように、各転置インデックスの検索対象集合のサイズの範囲を決定し、サイズの範囲毎に検索対象集合を分けて、転置インデックスを生成する、転置インデックス生成部２０と、検索条件集合のサイズと集合間類似度に対して設定された閾値とに基づいて、類似度が閾値以上となるために必要な条件を求め、集合のサイズの最小値が条件を満たす転置インデックス以外を、検索不要な転置インデックスとして同定する不要転置インデックス同定部２１と、非同定転置インデックスに対して検索を実行するデータ検索部２２と、を備える。