

(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
2012 年 12 月 13 日(13.12.2012)



(10) 国際公開番号
WO 2012/169128 A1

- (51) 国際特許分類:
G06F 17/30 (2006.01) *G06F 17/21* (2006.01)
- (21) 国際出願番号: PCT/JP2012/003357
- (22) 国際出願日: 2012 年 5 月 23 日(23.05.2012)
- (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 2011-128731 2011 年 6 月 8 日(08.06.2011) JP
- (71) 出願人 (米国を除く全ての指定国について): 株式会社 東芝 (KABUSHIKI KAISHA TOSHIBA) [JP/JP]; 〒1058001 東京都港区芝浦一丁目 1 番 1 号 Tokyo (JP). 東芝ソリューション株式会社 (TOSHIBA SOLUTIONS CORPORATION) [JP/JP]; 〒1058001 東京都港区芝浦一丁目 1 番 1 号 Tokyo (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてのみ): 小林 幹門 (KOBAYASHI, Mikito). 篠崎 政久 (SHINOZAKI,

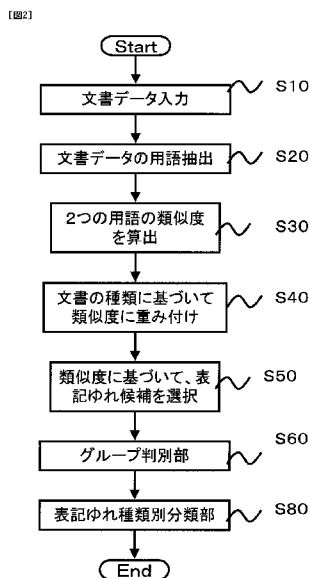
Masahisa). 加納 敏行 (KANO, Toshiyuki). 齋藤 佳美 (SAITO, Yoshimi).

- (74) 代理人: 砂井 正之 (SAGOI, Masayuki); 〒1058001 東京都港区芝浦一丁目 1 番 1 号東芝テクノロジー株式会社内 Tokyo (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, 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, IS, 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, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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

[続葉有]

(54) Title: ORTHOGRAPHICAL VARIANT DETECTION DEVICE AND ORTHOGRAPHICAL VARIANT DETECTION PROGRAM

(54) 発明の名称: 表記ゆれ検出装置及び表記ゆれ検出プログラム



- S10 Input document data
S20 Extract document data vocabulary
S30 Compute similarity of two vocabulary terms
S40 Apply weighting to similarity on basis of document type
S50 Select orthographical variant candidate on basis of similarity
S60 Group assessment unit
S80 Orthographical variant type classification unit

(57) Abstract: Provided is an orthographical variant detection device with which an orthographical variant candidate is detected with good precision. An orthographical variant detection device of an embodiment comprises: a vocabulary extraction unit which extracts vocabulary from document data; a similarity computation unit which computes the similarity of an arbitrary pair of terms of the extracted vocabulary; an orthographical variant candidate assessment unit which assesses whether the pair of vocabulary terms is an orthographical variant candidate on the basis of the similarity; and a group classification unit which groups the orthographical variant candidate on the basis of text character strings which are included in both terms in the orthographical variant vocabulary term pair which is the orthographical variant candidate.

(57) 要約: 精度よく表記ゆれ候補を検出する表記ゆれ検出装置を提供することである。実施形態の表記ゆれ検出装置は、文書データから用語を抽出する用語抽出部と、抽出された用語の任意のペアの類似度を算出する類似度算出部と、類似度に基づいて、用語のペアが表記ゆれ候補

であるかを判定する表記ゆれ候補判定部と、表記ゆれ候補である用語のペアに共通に含まれる文字列に基づいて表記ゆれ候補をグループ分けするグループ分類部と、を備える。

WO 2012/169128 A1



(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, ML, MR,
NE, SN, TD, TG).

添付公開書類:

— 国際調査報告 (条約第 21 条(3))