

(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
2022年11月17日(17.11.2022)



(10) 国際公開番号

WO 2022/239640 A1

(51) 国際特許分類:
G06F 40/247 (2020.01)

(21) 国際出願番号: PCT/JP2022/018863

(22) 国際出願日: 2022年4月26日(26.04.2022)

(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 2021-080731 2021年5月12日(12.05.2021) JP

(71) 出願人: 株式会社日立製作所(HITACHI, LTD.)
[JP/JP]; 〒1008280 東京都千代田区丸の内一丁目6番6号 Tokyo (JP).

(72) 発明者: 我妻 正太郎 (AGATSUMA, Shotaro);
〒1008280 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内 Tokyo (JP). 照屋 絵理 (TERUYA, Eri); 〒1008280 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内 Tokyo (JP). 竹内 理 (TAKEUCHI, Tadashi); 〒1008280 東京都千代田区丸の内一丁目6番6号 株式会社日立製作所内 Tokyo (JP).

(74) 代理人: 一色国際特許業務法人 (ISSHIKI & CO.); 〒1080073 東京都港区三田三丁目11番36号三田日東ダイビル Tokyo (JP).

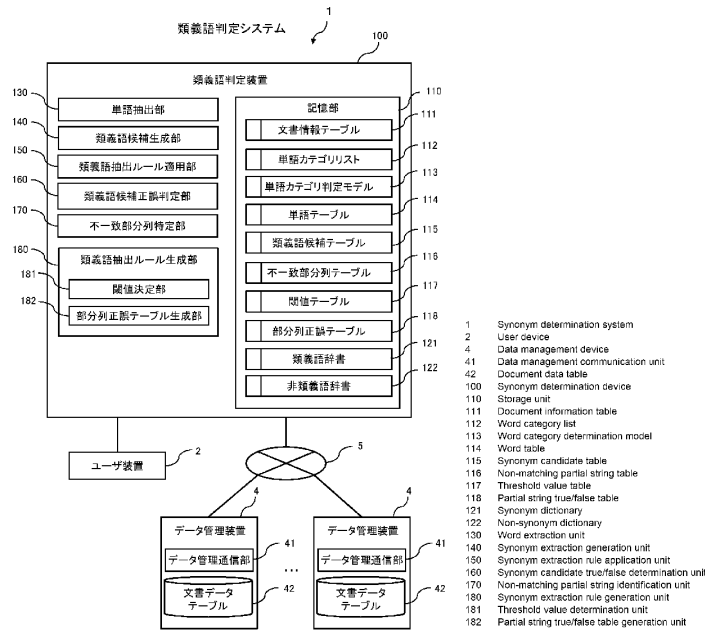
(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH,

(54) Title: SYNONYM DETERMINATION SYSTEM AND SYNONYM DETERMINATION METHOD

(54) 発明の名称: 類義語判定システム、および類義語判定方法



WO 2022/239640 A1



(57) Abstract: In the present invention, synonyms are efficiently extracted from document data with high accuracy. This synonym determination system: acquires, for some of a plurality of synonym candidates consisting of combinations of two words selected from a plurality of words extracted from document data, true/false information, which is information indicating whether the two constituent words are synonyms; generates a synonym extraction rule, which is information for determining whether two words constituting a synonym candidate are synonyms, on the basis of the true/false information

[続葉有]

CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ,
EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
HR, HU, ID, IL, IN, IR, IS, IT, JM, JO, KE, KG, KH,
KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, 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,
ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, WS, 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, ST, 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:

一 国際調査報告 (条約第21条(3))

and a feature amount for the synonym candidates acquired from the document data; and extracts a synonym candidate for which the two constituent words are synonyms by applying the synonym extraction rule to synonym candidates for which true/false information has not been acquired. The true/false information is acquired by, for example, being received from a user via a user interface.

(57) 要約: 文書データから類義語を効率よく高い精度で抽出する。類義語判定システムは、文書データから抽出される複数の単語から選択される2つの単語の組合せである複数の類義語候補の一部について、構成する2つの単語が類義語の関係となるかを示す情報である正誤情報を取得し、文書データから取得される類義語候補についての特徴量と正誤情報とに基づき、類義語候補を構成する2つの単語が類義語の関係となるかを判定するための情報である類義語抽出ルールを生成し、正誤情報を取得していない類義語候補に類義語抽出ルールを適用することにより、構成する2つの単語が類義語の関係となる類義語候補を抽出する。上記の正誤情報は、例えば、ユーザインタフェースを介してユーザから受け付けることにより取得する。