

(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 4 月 3 日(03.04.2014)



(10) 国際公開番号

WO 2014/050981 A1

- (51) 国際特許分類:
G06F 17/27 (2006.01) G06F 17/30 (2006.01)
- (21) 国際出願番号: PCT/JP2013/076094
- (22) 国際出願日: 2013 年 9 月 26 日(26.09.2013)
- (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 2012-213536 2012 年 9 月 27 日(27.09.2012) 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: 大西 貴士(ONISHI Takashi); 〒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). 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: 宇高 克己(UDAKA Katsuki); 〒1010025
東京都千代田区神田佐久間町 1 - 1 4 第二東ビ
ル 5 階 Tokyo (JP).

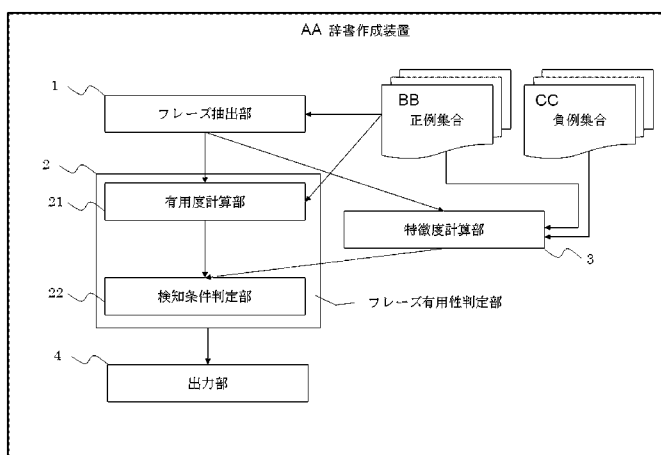
- (81) 指定国 (表示のない限り、全ての種類の国内保
護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA,
BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN,
CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES,
FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN,
IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS,
LT, LU, LY, MA, MD, ME, 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))

(54) Title: DICTIONARY CREATION DEVICE FOR MONITORING TEXT INFORMATION, DICTIONARY CREATION METHOD FOR MONITORING TEXT INFORMATION, AND DICTIONARY CREATION PROGRAM FOR MONITORING TEXT INFORMATION

(54) 発明の名称: テキスト情報監視用辞書作成装置、テキスト情報監視用辞書作成方法、及び、テキスト情報監視用辞書作成プログラム



- 1 Phrase extraction unit
2 Phrase usefulness determination unit
3 Feature degree calculation unit
4 Output unit
21 Usefulness degree calculation unit
22 Detection condition determination unit
AA Dictionary creation device
BB Positive example group
CC Negative example group

(57) Abstract: The purpose of the present invention is to create a dictionary for monitoring text information such that it is possible to achieve high-precision detection compared to prior art. A feature degree calculation unit (3) compares the statistics of a positive example group and a negative example group, and calculates the degree by which a given phrase appears in the positive example group as the feature degree. A usefulness degree calculation unit (21) calculates a usefulness degree by using indices pertaining to the length of the phrase, the frequency at which the phrase appears within the positive example group, and the inclusion relationship between phrases for each phrase extracted by means of a phrase extraction unit (1). A detection condition determination unit (22) uses the usefulness degree calculated by means of the usefulness degree calculation unit (21) and the feature degree calculated by means of the feature degree calculation unit (3) to evaluate the appropriateness of each phrase as a detection condition by means of the product of the usefulness degree and the feature degree, and determines that the phrase is appropriate as a detection condition when the value of the product is greater than a threshold value.

(57) 要約:

[続葉有]

従来技術に比べて高精度な検知を実現可能にするテキスト情報監視用辞書を作成することを目的とする。特徴度計算部 3 は、正例集合と負例集合の統計量を比較し、着目するフレーズが正例集合に出現する度合いを特徴度として計算する。有用度計算部 2 1 は、フレーズ抽出部 1 で抽出された各フレーズに対して、フレーズの長さ、フレーズの正例集合中の頻度、フレーズ間の包含関係に関する指標を用いて有用度を計算する。検知条件判定部 2 2 は、各フレーズに対して、有用度計算部 2 1 で計算した有用度と、特徴度計算部 3 で計算した特徴度とを用いて、有用度と特徴度の積によって検知条件として適切さを評価し、その値が閾値よりも大きい場合に検知条件として適切であると判定する。