

(12) International Application Status Report

Received at International Bureau: 02 April 2016 (02.04.2016)

Information valid as of: 05 September 2016 (05.09.2016)

Report generated on: 27 January 2021 (27.01.2021)

(10) Publication number:

WO2016/154466

(43) Publication date:

29 September 2016 (29.09.2016)

(26) Publication language:

English (EN)

(21) Application Number:

PCT/US2016/024069

(22) Filing Date:

24 March 2016 (24.03.2016)

(25) Filing language:

English (EN)

(31) Priority number(s):

201510133507.9 (CN)

(31) Priority date(s):

25 March 2015 (25.03.2015)

(31) Priority status:

Priority document received (in compliance with PCT Rule 17.1)

(51) International Patent Classification:

G06F 17/27 (2006.01); **G06K 9/00** (2006.01); **G06K 9/72** (2006.01); **G06K 9/18** (2006.01); **G06K 9/34** (2006.01); **G06K 9/46** (2006.01); **G06K 9/48** (2006.01)

(71) Applicant(s):

ALIBABA GROUP HOLDING LIMITED [/US]; Fourth Floor, One Capital Place P.O. Box 847 George Town, Grand Cayman (KY) (for all designated states)

(72) Inventor(s):

JIN, Xuan; Alibaba Group Legal Department 5/F, Building 3, No. 969 West Wen Yi Road Yu Hang District, Hangzhou, 311121 (CN)

WANG, Tianzhou; Alibaba Group Legal Department 5/F, Building 3, No. 969 West Wen Yi Road Yu Hang District, Hangzhou, 311121 (CN)

XUE, Qin; Alibaba Group Legal Department 5/F, Building 3, No. 969 West Wen Yi Road Yu Hang District, Hangzhou, 311121 (CN)

(74) Agent(s):

MURABITO, Anthony, C.; Murabito Hao & Barnes LLP 2 N. Market St. 3rd Floor San Jose, CA 95113 (US)

(54) Title (EN): METHOD AND APPARATUS FOR GENERATING TEXT LINE CLASSIFIER

(54) Title (FR): PROCÉDÉ ET APPAREIL DE GÉNÉRATION D'UN CLASSIFICATEUR DE LIGNES DE TEXTE

(57) Abstract:

(EN): A method of generating a text line classifier including generating text line samples by use of a present terminal system font reservoir. The method also includes extracting features from the text line samples and pre-stored marked-up samples. The method further includes training models by use of the extracted features to generate a text line classifier for recognizing text regions. With the system font reservoir being utilized for generating text line samples, the generated text line classifiers can target different scenes or different requirements for text region recognition with a high degree of applicability and wide application in addition to ease of implementation. Together with the combinational use of the marked up samples for extracting features from the text line samples, the generated text line classifiers provide for enhanced classification efficiency and accuracy.

(FR): La présente invention concerne un procédé de génération d'un classificateur de lignes de texte. Le procédé comprend les étapes consistant à : générer des échantillons de lignes de texte en utilisant une réserve actuelle de polices de caractères d'un système de terminal; extraire des caractéristiques à partir des échantillons de lignes de texte et d'échantillons marqués préstockés; et former des modèles en utilisant les caractéristiques extraites de façon à générer un classificateur de lignes de texte permettant de reconnaître des zones de texte. Grâce à la réserve de polices de caractères du système utilisée pour générer des échantillons de lignes de texte, les classificateurs de lignes de texte générés peuvent cibler différentes scènes ou différentes exigences de reconnaissance de zones de texte avec un haut degré d'applicabilité et une large application qui s'ajoutent à la facilité d'implémentation. Associés à l'utilisation combinée des échantillons marqués pour extraire les caractéristiques à partir des échantillons de lignes de texte, les classificateurs de lignes de texte générés assurent une efficacité et une précision accrues de la classification.

International search report:

Received at International Bureau: 25 June 2016 (25.06.2016) [US]

International Report on Patentability (IPRP) Chapter II of the PCT:

Not available

(81) Designated States:

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, 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

European Patent Office (EPO) : 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

African Intellectual Property Organization (OAPI) : BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG

African Regional Intellectual Property Organization (ARIPO) : BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW

Eurasian Patent Organization (EAPO) : AM, AZ, BY, KG, KZ, RU, TJ, TM