

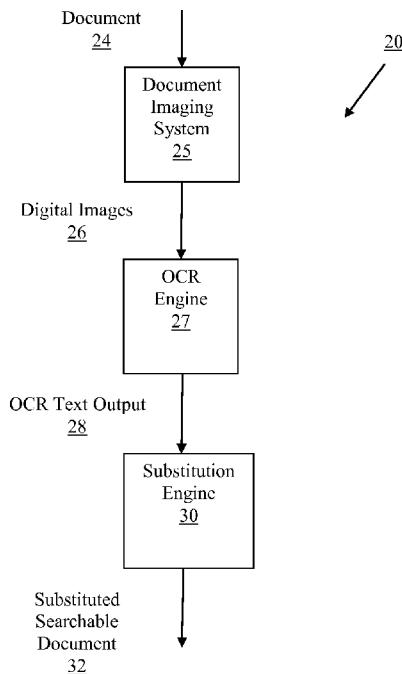


- (51) International Patent Classification:
G06K 9/00 (2006.01)
- (21) International Application Number:
PCT/US2014/030867
- (22) International Filing Date:
17 March 2014 (17.03.2014)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
61/798,223 15 March 2013 (15.03.2013) US
- (72) Inventors; and
- (71) Applicants : SUAREZ, Sergio David Jr. [US/US]; 821 N Neva, Addison, IL 60101 (US). MESKE, Joshua Daniel [US/US]; 5121 N. East River Rd., Apt. 2H, Chicago, IL 60656 (US).
- (74) Agent: PASKY, Jonathan R.; Pasky IP Law, 320 W Ohio St., Ste 300, Chicago, IL 60654 (US).

- (81) Designated States (unless otherwise indicated, for every kind of national protection available): 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) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (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).

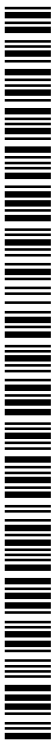
[Continued on next page]

(54) Title: SEARCHING TEXT BY OPTICAL CHARACTER RECOGNITION



(57) Abstract: A method for generating a character-by-character substitution in an optical character recognition (OCR) text output of a document including at least one character, includes: executing on a processor instructions for substituting an OCR key for the at least one character. The instructions include: identifying a class corresponding to the at least one character, wherein the class includes a character shape corresponding to at least a portion of the at least one character; substituting the OCR key including to the character shape for the at least one character; and generating a searchable substituted document including the OCR key.

Fig. 1





Published:

(88) Date of publication of the international search report:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

6 November 2014

INTERNATIONAL SEARCH REPORT

PCT/US14/30867

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06K 9/00 (2014.01)

CPC - G06K 9/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8) Classification(s): G06K 9/00, 9/20, 9/62, 9/48 (2014.01)

CPC Classification(s): G06K 9/00, 9/685, 9/344; USPC Classification(s): 382/181, 382/182, 382/177

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

MicroPatent (US-G, US-A, EP-A, EP-B, WO, JP-bib, DE-C,B, DE-A, DE-T, DE-U, GB-A, FR-A); ProQuest; IEEE/IEEEXplore; Google/Google Scholar; IP.com; Espacenet; Search Terms: character, letter, substitute, replace, OCR, optical character recognition, shape, search, document, classification, cardinality, frequency, identify

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ----- Y	US 7415171 B2 (SNAPP, RF) August 19, 2008, column 3, lines 14-24, column 5, lines 30-40, column 7, lines 39-51, column 7, lines 62-67, column 8, lines 1-14, column 13, lines 34-48, column 15, lines 5-23.	1, 3 ----- 2
Y	US 2010/0246963 A1 (AL-MUHTASEB, HA et al.) September 30, 2010, paragraph [0025]-[0028], [0039], [0040].	2

Further documents are listed in the continuation of Box C.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

2 September 2014 (02.09.2014)

Date of mailing of the international search report

25 SEP 2014

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-3201

Authorized officer:

Shane Thomas

PCT Helpdesk: 571-272-4300

PCT OSP: 571-272-7774