

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
21 June 2007 (21.06.2007)

PCT

(10) International Publication Number
WO 2007/070010 A1

(51) International Patent Classification:
G06K 9/00 (2006.01)

Suat Ling, Jamie [SG/SG]; c/o I2R, 21 Heng Mui Keng Terrace, Singapore 119613 (SG).

(21) International Application Number:
PCT/SG2005/000423

(74) Agent: **GREENE-KELLY, James, Patrick**; Lloyd Wise, Tanjong Pagar, P.O. Box 636, Singapore 910816 (SG).

(22) International Filing Date:
16 December 2005 (16.12.2005)

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US):
AGENCY FOR SCIENCE, TECHNOLOGY AND RESEARCH [SG/SG]; 20 Biopolis Way, #07-01 Centros, Singapore 138668 (SG).

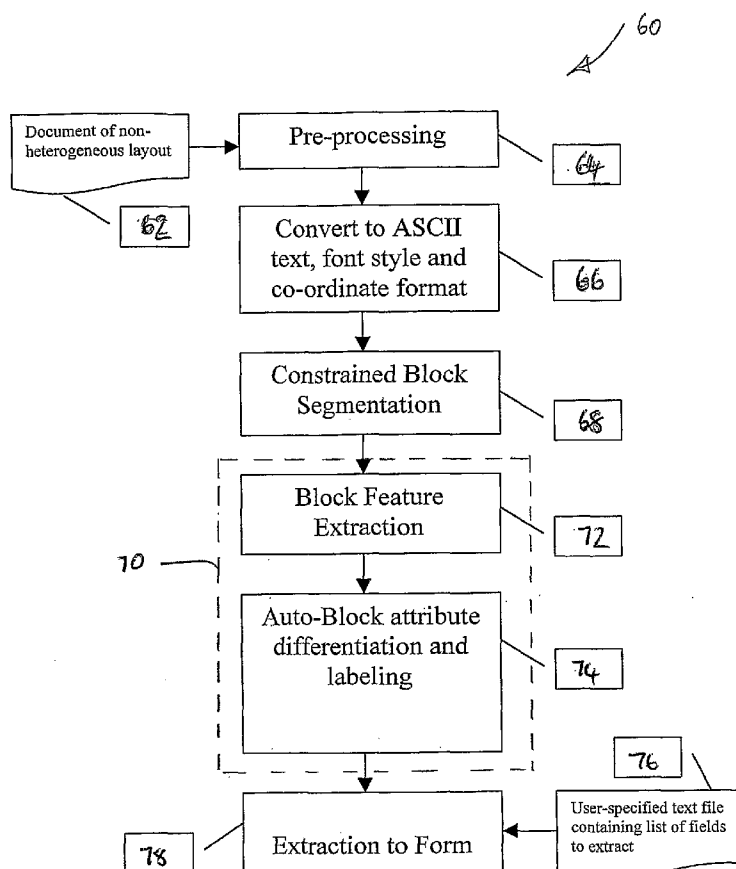
(72) Inventors; and

(75) Inventors/Applicants (for US only): **LAI, Fon Lin** [MY/SG]; c/o I2R, 21 Heng Mui Keng Terrace, Singapore 119613 (SG). **ONG, Hwee Leng** [SG/SG]; c/o I2R, 21 Heng Mui Keng Terrace, Singapore 119613 (SG). **NG,**

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

[Continued on next page]

(54) Title: IMPROVEMENTS IN ELECTRONIC DOCUMENT ANALYSIS



(57) Abstract: Apparatus and methods for electronic document analysis provide for analysing an electronic document having a first block segment to determine a first block segment characteristic from application of a first block characteristic extraction rule to the document. Input parameters of the first block characteristic extraction rule comprise a property of the first block segment and property of a region of the electronic document in which the first block segment is located. Apparatus and methods for electronic document analysis also provide for analysing an electronic document having first and second block segments to determine whether the first and second block segments are arranged in a recognisable pattern.

WO 2007/070010 A1



European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *with international search report*

(48) Date of publication of this corrected version:

23 August 2007

(15) Information about Correction:

see PCT Gazette No. 34/2007 of 23 August 2007

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.