

Title (en)

METHOD FOR RETRIEVING TEXT BLOCKS IN DOCUMENTS

Title (de)

VERFAHREN ZUM WIEDERAUFFINDEN VON TEXTBLÖCKEN IN DOKUMENTEN

Title (fr)

PROCEDE POUR RETROUVER DES BLOCS DE TEXTE DANS DES DOCUMENTS

Publication

EP 1917626 A1 20080507 (DE)

Application

EP 06776758 A 20060811

Priority

- EP 2006007939 W 20060811
- DE 102005040687 A 20050826

Abstract (en)

[origin: CA2620180A1] The invention relates to a method for retrieving text blocks in documents, preferably for postal mailings that are to be sorted, e.g. mass mailings. The aim of the invention is to retrieve or identify reference text blocks in all types of documents with the aid of distinctive characteristic data records of said reference text blocks. According to said method, structure-related characteristics of the text block are extracted as distinctive characteristics and compared with characteristics of a characteristic data record of a reference text block, allowing a simple recognition of similar characteristics in several text blocks to take place. A first extraction of structure-related characteristics can be carried out by the division of a text block into several lines, whose height or spacing is saved to a characteristic data record of a mailing. Different text blocks can be analysed for their similarities by comparing the characteristic data records.

IPC 8 full level

G06V 30/10 (2022.01)

CPC (source: EP US)

G06V 30/147 (2022.01 - EP US); **G06V 30/424** (2022.01 - EP US); **G06V 30/10** (2022.01 - EP US)

Citation (search report)

See references of WO 2007022877A1

Designated contracting state (EPC)

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU LV MC NL PL PT RO SE SI SK TR

DOCDB simple family (publication)

DE 102005040687 A1 20070301; CA 2620180 A1 20070301; CN 101263512 A 20080910; EP 1917626 A1 20080507; US 2009252415 A1 20091008; WO 2007022877 A1 20070301

DOCDB simple family (application)

DE 102005040687 A 20050826; CA 2620180 A 20060811; CN 200680031129 A 20060811; EP 06776758 A 20060811; EP 2006007939 W 20060811; US 99105806 A 20060811