Title (en)
METHOD FOR AUTOMATIC THEMATIC CLASSIFICATION OF A DIGITAL TEXT FILE

Title (de)
VERFAHREN ZUR AUTOMATISCHEN THEMATISCHEN KLASSIFIKATION EINER DIGITALEN TEXTDATEI

Title (fr)
PROCEDE DE CLASSIFICATION THEMATIQUE AUTOMATIQUE D'UN FICHIER DE TEXTE NUMERIQUE

Publication
**EP 3008635 A1 20160420 (FR)**

Application
**EP 14728537 A 20140604**

Priority
- FR 1355596 A 20130614
- EP 2014061535 W 20140604

Abstract (en)
[origin: WO2014198595A1] The invention primarily relates to a method for the thematic classification of a digital text file (1) from an encyclopaedic database (5) comprising a category graph (G), said method comprising, during a learning phase (PA) making it possible to develop a thematic classification model (3), the step of grouping together, for each category node, all of the items directly attached to that category node so as to obtain a "word bag" for each category node; determining a so-called term-frequency vector characteristic of the category node; combining, on each category node, the term-frequency vector directly connected to it with term-frequency vectors of more specific nodes; and in that it comprises, during a production phase (PP), a step for calculating the term-frequency vector (V) of said digital text file (1) and selecting, in said thematic classification model (3), N category nodes having the term-frequency vectors (V') closest to the term-frequency vector (V) of the digital text file.

IPC 8 full level
**G06F 17/30** (2006.01); **G06N 5/02** (2006.01)

CPC (source: EP US)
**G06F 16/353** (2018.12 - EP US); **G06F 16/367** (2018.12 - EP US); **G06F 16/9024** (2018.12 - EP US)

Citation (search report)
See references of WO 2014198595A1

Cited by
CN107391593A; CN110222175A

Designated contracting state (EPC)
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

Designated extension state (EPC)
BA ME

DOCDB simple family (publication)
**WO 2014198595 A1 20141218**; EP 3008635 A1 20160420; FR 3007164 A1 20141219; FR 3007164 B1 20161007; US 2016140220 A1 20160519

DOCDB simple family (application)
**EP 2014061535 W 20140604**; EP 14728537 A 20140604; FR 1355596 A 20130614; US 201414898141 A 20140604