Title (en)
EXTRACTION OF ATTRIBUTES AND VALUES FROM NATURAL LANGUAGE DOCUMENTS

Title (de)
EXTRAHIERUNG VON ATTRIBUTEN UND WERTEN AUS DOKUMENTEN IN NATÜRLICHER SPRACHE

Title (fr)
EXTRACTION D'ATTRIBUTS ET VALEURS DE DOCUMENTS EN LANGAGE NATUREL

Publication
EP 2024915 A2 20090218 (EN)

Application
EP 07812022 A 20070605

Priority
- US 2007070427 W 20070605
- US 80394006 P 20060605
- US 74221507 A 20070430
- US 74224407 A 20070430

Abstract (en)
[origin: WO2007143658A2] One or more classification algorithms are applied to at least one natural language document in order to extract both attributes and values of a given product. Supervised classification algorithms, semi-supervised classification algorithms, unsupervised classification algorithms or combinations of such classification algorithms may be employed for this purpose. The at least one natural language document may be obtained via a public communication network. Two or more attributes (or two or more values) thus identified may be merged to form one or more attribute phrases or value phrases. Once attributes and values have been extracted in this manner, association or linking operations may be performed to establish attribute -value pairs that are descriptive of the product. In a presently preferred embodiment, an (unsupervised) algorithm is used to generate seed attributes and values which can then support a supervised or semi-supervised classification algorithm.

IPC 8 full level
G06Q 10/00 (2006.01)

CPC (source: EP)
G06F 40/258 (2020.01); **G06Q 10/00** (2013.01)

Citation (search report)
See references of WO 2007143658A2

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 MT NL PL PT RO SE SI SK TR

Designated extension state (EPC)
AL BA HR MK RS

DOCDB simple family (publication)
WO 2007143658 A2 20071213; **WO 2007143658 A3 20080424**; **WO 2007143658 A9 20080207**; EP 2024915 A2 20090218

DOCDB simple family (application)
US 2007070427 W 20070605; EP 07812022 A 20070605