

Title (en)

EFFICIENT CRAWLING USING PATH SCHEDULING, AND APPLICATIONS THEREOF

Title (de)

EFFIZIENTES CRAWLING UNTER VERWENDUNG VON PFADPLANUNG UND ANWENDUNGEN DAVON

Title (fr)

EXPLORATION EFFICACE FAISANT APPEL À LA PLANIFICATION DE CHEMINS, ET APPLICATIONS CORRESPONDANTES

Publication

EP 4052145 A4 20231101 (EN)

Application

EP 20883009 A 20201030

Priority

- US 201916668544 A 20191030
- US 201916668524 A 20191030
- US 2020058286 W 20201030

Abstract (en)

[origin: WO2021087308A1] The present disclosure is directed to systems and methods for extracting unstructured data from a data source in a structure manner. Embodiments provide ways to retrieve unstructured data along from data sources not optimized for automated retrieval. For example, embodiments may generate a branched tree for each data source that maps out paths to individual sites of, for example, a healthcare provider listing the unstructured data. Using this branched tree, tasks can be generated to navigate along a path with the data source to each site and extract the unstructured data from the data source. In this way, embodiments provide the ability to navigate through a site from a base site to a site that has the relevant data.

IPC 8 full level

G06F 17/00 (2019.01); **G06F 16/951** (2019.01); **G06Q 30/0201** (2023.01)

CPC (source: EP)

G06F 16/951 (2019.01); **G06Q 30/0201** (2013.01); **G06Q 50/22** (2013.01)

Citation (search report)

- [Y] US 2013166207 A1 20130627 - SHAO WEILI [US], et al
- [Y] US 2016188717 A1 20160630 - ROSENBERG NAOR [IL], et al
- [A] US 2018150562 A1 20180531 - GUNDIMEDA VENUGOPAL [IN], et al
- See also references of WO 2021087308A1

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

DOCDB simple family (publication)

WO 2021087308 A1 20210506; CN 114761945 A 20220715; EP 4052145 A1 20220907; EP 4052145 A4 20231101

DOCDB simple family (application)

US 2020058286 W 20201030; CN 202080076024 A 20201030; EP 20883009 A 20201030