

Title (en)  
DOMAIN BASED TEXT EXTRACTION

Title (de)  
BEREICHSBASIERTE TEXTEXTRAKTION

Title (fr)  
EXTRACTION DE TEXTE BASÉE SUR LE DOMAINE

Publication  
**EP 4085343 A1 20221109 (EN)**

Application  
**EP 20910797 A 20201230**

Priority  
• IN 201941054421 A 20191230  
• IB 2020062535 W 20201230

Abstract (en)  
[origin: WO2021137166A1] This disclosure relates to a method and system for extracting information from contents of an input file. The method may include identifying text data from the input file, receiving a text input from a user for identifying relevant text entities from the plurality of text entities, and automatically generating a search pattern corresponding to the text input. The method may further include determining a pattern associated with each of the plurality of text entities, and mapping the search pattern corresponding to the text input with patterns associated with the plurality of text entities. The method may further include identifying one or more matching patterns from the patterns associated with the plurality of text entities based on the mapping, and extracting, from the plurality of text entities, relevant text entities corresponding to the one or more matching patterns.

IPC 8 full level  
**G06F 16/00** (2019.01); **G06F 40/00** (2020.01); **G06V 30/262** (2022.01); **G06V 30/10** (2022.01)

CPC (source: EP US)  
**G06F 18/22** (2023.01 - EP); **G06F 18/24133** (2023.01 - EP); **G06F 40/216** (2020.01 - EP); **G06F 40/279** (2020.01 - EP US);  
**G06V 30/262** (2022.01 - EP US); **G06V 30/416** (2022.01 - EP US); **G06V 30/10** (2022.01 - EP US); **G06V 2201/10** (2022.01 - EP)

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

Designated validation state (EPC)  
KH MA MD TN

DOCDB simple family (publication)  
**WO 2021137166 A1 20210708**; AU 2020418619 A1 20220526; CA 3156204 A1 20210708; EP 4085343 A1 20221109; EP 4085343 A4 20240103;  
JP 2023507881 A 20230228

DOCDB simple family (application)  
**IB 2020062535 W 20201230**; AU 2020418619 A 20201230; CA 3156204 A 20201230; EP 20910797 A 20201230; JP 2022525481 A 20201230