

Title (en)

WORDBREAK ALGORITHM WITH OFFSET MAPPING

Title (de)

WORTUNTERBRECHUNGsalgorithmus mit offset-abbildung

Title (fr)

ALGORITHME DE COUPURE DE MOTS COMPRENANT UN MAPPAGE DE DÉCALAGE

Publication

EP 4348490 A1 20240410 (EN)

Application

EP 22751806 A 20220505

Priority

- IN 202141023933 A 20210528
- US 202117444347 A 20210803
- IB 2022000257 W 20220505

Abstract (en)

[origin: WO2022248933A1] A computer system is provided, including a processor coupled to a mass storage device that stores instructions, which, upon execution by the processor, cause the processor to store an original string formed of a plurality of characters, perform a wordbreak algorithm on the original string, and tokenize the original string to generate a processed string including a plurality of word tokens separated by spaces. The processor is further configured to generate an offset map between locations within the word tokens in the processed string and corresponding locations in the original string and classify a portion of the processed string as a target. The processor is further configured to identify target characters in the original string that correspond to the target using the offset map and perform a predetermined action on the target characters in the original string.

IPC 8 full level

G06F 40/163 (2020.01); **G06F 21/62** (2013.01); **G06F 40/166** (2020.01); **G06F 40/284** (2020.01); **G06F 40/53** (2020.01)

CPC (source: EP KR)

G06F 21/6254 (2013.01 - EP KR); **G06F 40/109** (2020.01 - KR); **G06F 40/163** (2020.01 - EP KR); **G06F 40/166** (2020.01 - EP KR); **G06F 40/284** (2020.01 - EP KR); **G06F 40/53** (2020.01 - EP KR); **G06F 40/109** (2020.01 - 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 2022248933 A1 20221201; EP 4348490 A1 20240410; JP 2024521833 A 20240604; KR 20240011718 A 20240126

DOCDB simple family (application)

IB 2022000257 W 20220505; EP 22751806 A 20220505; JP 2023573274 A 20220505; KR 20237040866 A 20220505