(12) UK Patent Application (19) GB (11) 2627092 (13) A

(43) Date of Reproduction by UK Office 14.08.2024

(21) Application No: 2405984.2

(22) Date of Filing: 15.08.2022

Date Lodged: 29.04.2024

(30) Priority Data:
(31) 63273761 (32) 29.10.2021 (33) US
(31) 17819445 (32) 12.08.2022 (33) US

(86) International Application Data:
PCT/US2022/074974 En 15.08.2022

(87) International Publication Data:
WO2023/076754 En 04.05.2023

(71) Applicant(s):
ORACLE INTERNATIONAL CORPORATION
500 Oracle Parkway, Mail Stop 5OP7,
Redwood Shores, CA 94065, United States of America

(72) Inventor(s):
Xu Zhong
Yakupitiyage Don Thanuja Samodhye Dharmasiri
Thanh Long Duong
Mark Edward Johnson

(74) Agent and/or Address for Service:
Gill Jennings & Every LLP
The Broadgate Tower, 20 Primrose Street, London,
EC2A 2ES, United Kingdom

(51) INT CL:
G06F 16/332 (2019.01) ***G06F 16/33*** (2019.01)
G06F 16/35 (2019.01) ***G06N 3/08*** (2023.01)
H04L 51/02 (2022.01)

(56) Documents Cited:
JP 2004326600 A **US 20200073882 A1**
US 20190073420 A1
KR 1020190058935
US 10607042

(58) Field of Search:
INT CL **G06F, G06N**
Other: **eKOMPASS(KIPO internal)**

(54) Title of the Invention: **Deep learning techniques for extraction of embedded data from documents**
Abstract Title: **Deep learning techniques for extraction of embedded data from documents**

(57) Deep learning techniques are disclosed for extraction of embedded data from documents. In an exemplary technique, a set of unstructured text data is received. One or more text groupings are generated by processing the set of unstructured text data. One or more text grouping embeddings are generated in a format for input to a machine learning model based on the one or more generated text groupings. One or more output predictions are generated by inputting the one or more text grouping embeddings into the machine learning model. Each output prediction of the one or more output predictions correspond to a predicted aspect of a text grouping of the one or more text groupings.

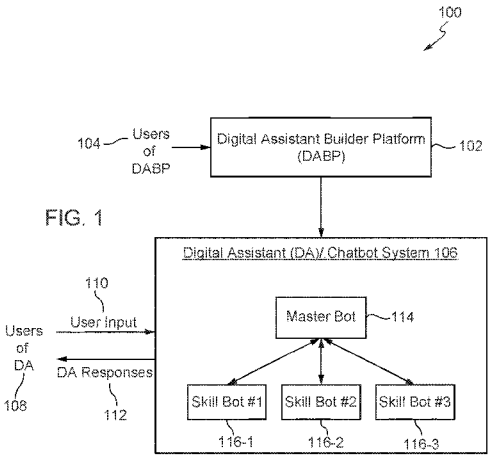


FIG. 1

GB 2627092 A