

(12) **UK Patent Application** (19) **GB** (11) **2625485** (13) **A**

(43) Date of Reproduction by UK Office **19.06.2024**

(21) Application No: **2404718.5**  
 (22) Date of Filing: **07.11.2022**  
 Date Lodged: **03.04.2024**

(30) Priority Data:  
 (31) **63263728** (32) **08.11.2021** (33) **US**  
 (31) **18052694** (32) **04.11.2022** (33) **US**

(86) International Application Data:  
**PCT/US2022/049164 En 07.11.2022**

(87) International Publication Data:  
**WO2023/081483 En 11.05.2023**

(71) Applicant(s):  
**Oracle International Corporation**  
**(Incorporated in USA - Delaware)**  
**500 Oracle Parkway, M/S 50P7, Redwood Shores,**  
**California 94065, United States of America**

(72) Inventor(s):  
**Thanh Tien Vu**  
**Poorya Zaremoondi**  
**Duy Vu**  
**Mark Edward Johnson**  
**Xu Zhong**  
**Vladislav Blinov**  
**Cong Duy Vu Hoang**  
**Yu-Heng Hong**  
**Vinamr Goel**  
**Philip Victor Orgren**  
**Srinivasa Phani Kumar Gadde**  
**Vishal Vishnoi**  
**Thanh Long Duong**

(continued on next page)

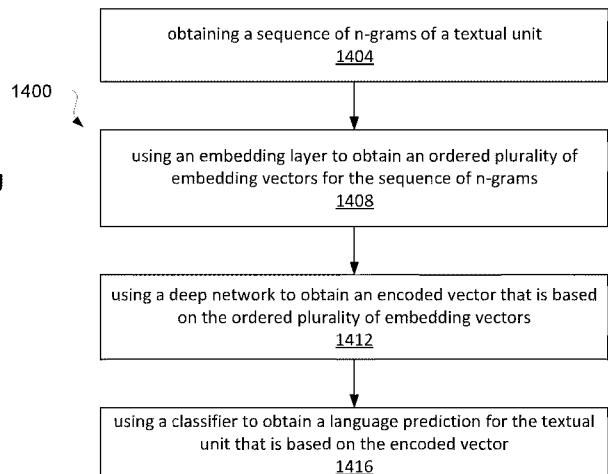
(51) INT CL:  
**G06F 40/263** (2020.01) **G06F 16/332** (2019.01)  
**G06F 18/2415** (2023.01) **G06F 40/216** (2020.01)  
**G06F 40/274** (2020.01) **G06N 3/0464** (2023.01)  
**G06N 3/0499** (2023.01) **G06N 3/08** (2023.01)  
**H04L 51/02** (2022.01)

(56) Documents Cited:  
**CN 110955745 A** **US 20210294781 A1**  
**US 20190019503 A1** **US 20170300691 A1**  
**CAOJIN ZHANG et al; "Model Size Reduction Using**  
**Frequency Based Double Hashing for Recommender**  
**Systems", arXiv:2007.14523, July 2020, pp.1-8**  
**[retrieved on 2023.03.23]. Retrieved from <https://**  
**arxiv.org/abs/2007.14523> pages 1-7 (20200**

(58) Field of Search:  
 INT CL **G06F, G06N, G10L, H04L**  
 Other: **eKOMPASS (KIPO internal)**

(54) Title of the Invention: **Wide and deep network for language detection using hash embeddings**  
 Abstract Title: **Wide and deep network for language detection using hash embeddings**

(57) Techniques disclosed herein relate generally to language detection. In one particular aspect, a method is provided that includes obtaining a sequence of n-grams of a textual unit; using an embedding layer to obtain an ordered plurality of embedding vectors for the sequence of n-grams; using a deep network to obtain an encoded vector that is based on the ordered plurality of embedding vectors; and using a classifier to obtain a language prediction for the textual unit that is based on the encoded vector. The deep network includes an attention mechanism, and using the embedding layer to obtain the ordered plurality of embedding vectors comprises, for each n-gram in the sequence of n-grams: obtaining hash values for the n-gram; based on the hash values, selecting component vectors from among the plurality of component vectors; and obtaining an embedding vector for the n-gram that is based on the component vectors.



**GB 2625485 A**

**GB 2625485 A continuation**

(74) Agent and/or Address for Service:

**Gill Jennings & Every LLP**  
**The Broadgate Tower, 20 Primrose Street, London,**  
**EC2A 2ES, United Kingdom**