

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2026 年 3 月 5 日 (05.03.2026)



(10) 国际公布号
WO 2026/045027 A1

- (51) 国际专利分类号:
G06N 3/08 (2023.01) **G06F 40/10** (2020.01)
- (21) 国际申请号: PCT/CN2024/140674
- (22) 国际申请日: 2024 年 12 月 19 日 (19.12.2024)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202411191440.X 2024年8月28日 (28.08.2024) CN
- (71) 申请人: 无限光年(上海)技术有限公司(INFLY TECH (SHANGHAI) CO., LTD.) [CN/CN]; 中国上海市徐汇区虹漕路25-1号2层1325室 200233 (CN)。
- (72) 发明人: 褚巍 (CHU, Wei); 中国上海市徐汇区虹漕路25-1号2层1325室 200233 (CN)。

- (74) 代理人: 北京智信禾专利代理有限公司(CHISPO ATTORNEYS AT LAW); 中国北京市东城区崇文门外大街8号院哈德门广场西塔901 100062 (CN)。
- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。
- (84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ,

(54) **Title:** LANGUAGE MODEL TRAINING METHOD AND APPARATUS, AND INFORMATION PROCESSING METHOD AND APPARATUS

(54) 发明名称: 语言模型训练方法及装置、信息处理方法及装置

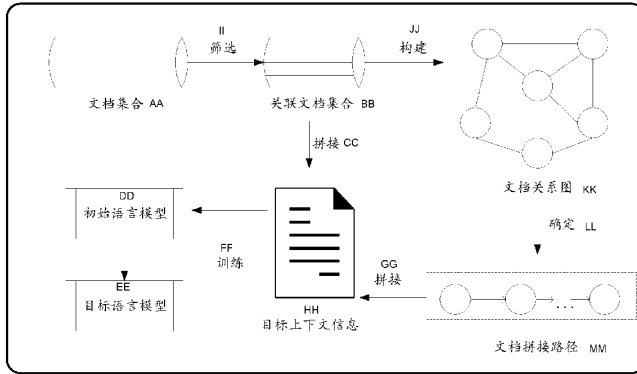


图 1

AA	Document collection
BB	Associated document collection
CC, GG	Concatenate
DD	Initial language model
EE	Target language model
FF	Train
HH	Target context information
II	Select
JJ	Construct
KK	Document relationship graph
LL	Determine
MM	Document concatenation path

(57) **Abstract:** Provided are a language model training method and apparatus, and an information processing method and apparatus. The language model training method comprises: acquiring a document collection of an initial language model, and selecting associated documents from the document collection to form an associated document collection (S202); on the basis of the associated documents in the associated document collection and the similarity between the associated documents, constructing a document relationship graph corresponding to the associated document collection (S204); determining a document concatenation path in the document relationship

[见续页]

NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚
(AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE,
BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR,
HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO,
PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN,
TD, TG)。

本国际公布：
— 包括国际检索报告(条约第21条(3))。

graph, and on the basis of the document concatenation path, extracting target associated documents from the associated document collection and performing concatenation processing on the extracted target associated documents, in order to obtain target context information (S206); and on the basis of the target context information, training the initial language model until a target language model meeting a training stopping condition is obtained (S208).

(57) 摘要: 提供了一种语言模型训练方法及装置、信息处理方法及装置。其中语言模型训练方法包括: 获取初始语言模型的文档集合, 并在文档集合中筛选关联文档组成关联文档集合 (S202); 根据关联文档集合中的关联文档以及关联文档之间的相似度, 构建关联文档集合对应的文档关系图 (S204); 在文档关系图中确定文档拼接路径, 并按照文档拼接路径在关联文档集合中提取目标关联文档进行拼接处理, 获得目标上下文信息 (S206); 基于目标上下文信息对初始语言模型进行训练, 直至获得满足训练停止条件的目标语言模型 (S208)。