



(51) International Patent Classification:
G06F 17/27 (2006.01)

(21) International Application Number:
PCT/CN2019/083723

(22) International Filing Date:
22 April 2019 (22.04.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
201910269532.8 04 April 2019 (04.04.2019) CN

(71) Applicant: **BEIJING DIDI INFINITY TECHNOLOGY AND DEVELOPMENT CO., LTD.** [CN/CN]; Building 34, No. 8 Dongbeiwang West Road, Haidian District, Beijing 100193 (CN).

(72) Inventors: **HU, Juan**; Room 201, 2/F, Building 3, No. 29 Dongbeiwang South Road, Haidian District, Beijing 100193 (CN). **CHEN, Huan**; Room 201, 2/F, Building 3, No. 29 Dongbeiwang South Road, Haidian District, Beijing 100193 (CN). **SONG, Qi**; Building 34, No. 8 Dongbeiwang West Road, Haidian District, Beijing 100193 (CN). **MA, Li**; Room 201, 2/F, Building 3, No. 29 Dongbeiwang South Road, Haidian District, Beijing 100193 (CN).

(74) Agent: **METIS IP (CHENGDU) LLC**; Room 808, 8th Floor, Building B7, Block D, Tianfu Jingrong Center, No. 99 West Section of Hupan Road, Tianfu New District, Chengdu, Sichuan 610213 (CN).

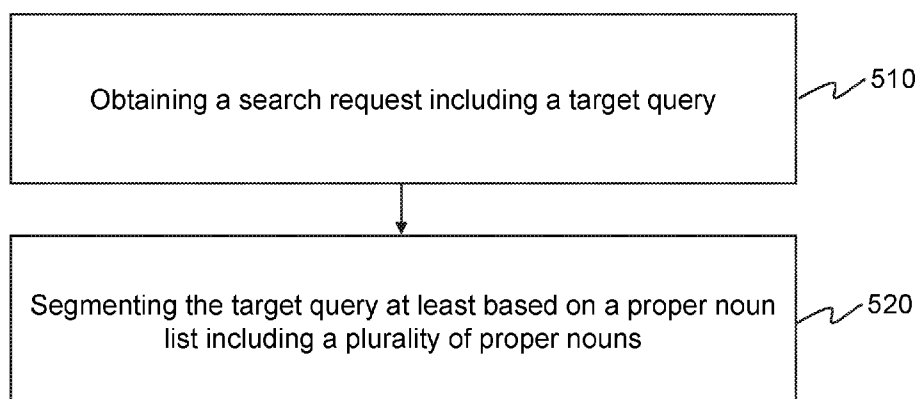
(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, 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).

(54) Title: SYSTEMS AND METHODS FOR IDENTIFYING PROPER NOUNS

FIG. 5

500



(57) Abstract: Systems and methods for identifying proper noun. The method may include obtaining a search request including a target query. The method may also include determining whether the target query includes at least one proper noun based on a trained identification model. The trained identification model may be configured to provide a probability that the target query includes at least one proper noun. In response to a determination that the probability is greater than a predetermined probability threshold, the method may include determining that the target query includes at least one proper noun.

Published:

— *with international search report (Art. 21(3))*

SYSTEMS AND METHODS FOR IDENTIFYING PROPER NOUNS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to Chinese Application No. 201910269532.8 filed on April 4, 2019, the contents of which are incorporated herein by reference in their entirety

TECHNICAL FIELD

[0002] The present disclosure generally relates to systems and methods for identifying proper nouns, and in particular, to systems and methods for segmenting a query based on the proper nouns.

BACKGROUND

[0003] Currently, when a user initiates a search request by inputting a query (e.g., a query that the user intends to search) via a user terminal, after receiving the search request, a system providing search services segments the query, determines, based on the segmented result, one or more terms of interest (TOIs) associated with the query, and recommends at least a portion of the TOIs to the user terminal. This approach may segment a proper noun included in the query, causing the system to fail to identify the proper noun (e.g., a name of a person, a name of a place, a name of an organization) included in the query. For example, if the user intends to search Novel Value Technology Limited (a name of a company), a system providing search services may segment Novel Value Technology into, e.g., Novel value, Value Technology, Technology Limited, etc. Consequently, the system is unable to identify the proper noun "Novel Value Technology Limited." To resolve this issue, traditionally, a proper noun identification method is performed by training the sequence-labeling model using the context information, which is

labeled in sequence. However, due to the low frequency of certain proper nouns in the travel scenario, and inputted short queries and TOI strings, which result in insufficient context information, this traditional method may not work. Therefore, it is desirable to provide new systems and methods for identifying proper nouns in and/or segmenting a query based on a refined proper noun list automatically and efficiently.

SUMMARY

[0004] In a first aspect of the present disclosure, a system for segmenting target queries is provided. The system may include at least one storage medium and at least one processor in communication with the at least one storage medium. The at least one storage medium may include a set of instructions. When executing the set of instructions, the at least one processor may be directed to: obtain a search request including a target query; and segment the target query at least based on a proper noun list including a plurality of proper nouns, wherein the proper noun list is provided by: obtaining a plurality of first historical search records, wherein each of the plurality of first historical search records includes a first query from a first user, and a first Term of Interest (TOI) selected by the first user as a first search term of the first historical search record, and at least a portion of the first TOIs include at least one proper noun; obtaining a trained identification model, wherein the trained identification model is configured to provide a probability that the first TOI includes at least one proper noun; and determining the proper noun list at least based on the trained identification model and the plurality of first historical search records.

[0005] In some embodiments, to segment the target query at least based on a proper noun list including a plurality of proper nouns, the at least one processor may be directed to: segment the target query into a plurality of segments using a segmenting technique, wherein the segmenting technique

includes an N-gram technique; compare the plurality of segments with the proper noun list; and combine at least two segments of the plurality of segments in response to a determination that the at least two combined segments is a proper noun on the proper noun list.

[0006] In some embodiments, to segment the target query at least based on a proper noun list including a plurality of proper nouns, the at least one processor may be directed to: obtain a segmentation model, wherein the segmentation model is configured to segment the target query; merge the proper noun list into the segmentation model by allocating a weighting coefficient to the proper noun list; and segment the target query based on the merged segmentation model.

[0007] In some embodiments, the determining the proper noun list at least based on the trained identification model and the plurality of first historical search records may include: for each of the plurality of first historical search records; determining a first probability that the first TOI includes at least one proper noun; and determining whether the first probability is greater than a first predetermined probability threshold; and adding the proper nouns into the proper noun list if the first probability is greater than the first predetermined probability threshold.

[0008] In some embodiments, the trained identification model may be determined with a training process, and the training process may include: obtaining a plurality of second historical search records, wherein each of the plurality of second historical search records includes a second query from a second user, and a second TOI selected by the second user as a second search term of the second historical search record, and at least a portion of the second TOIs include at least one proper noun; obtaining a plurality of training samples based on the plurality of second historical search records; and training a preliminary identification model based on the plurality of training samples to generate the trained identification model.

[0009] In some embodiments, the obtaining a plurality of training samples based on the plurality of second historical search records may include: for each of the plurality of second historical search records, segmenting the second TOI into a plurality of second segments in the second historical search record; extracting feature information associated with the plurality of second segments; determining whether the feature information satisfies a first predetermined condition; and designating the second historical search record as a candidate search record if the feature information satisfies the first predetermined condition; and determining the plurality of training samples based on the candidate search records.

[0010] In some embodiments, the feature information of each of the plurality of second segments may include at least one of a first consistent parameter associated with the second query and the second segment, a second consistent parameter associated with the second TOI and the second segment, a cohesive parameter associated with the second segment, a left entropy associated with the second segment, a right entropy associated with the second segment, a degree of freedom associated with the second segment, a probability that the second segment is at a start of the second TOI or an end of the second TOI, and a frequency associated the second segment.

[0011] In some embodiments, the system may further include: manually filtering out at least one candidate search record among the candidate search records; and determining the plurality of training samples based on the filtered candidate search records.

[0012] In some embodiments, the training a preliminary identification model based on the plurality of training samples to generate the trained identification model may include: determining a plurality of sample probabilities corresponding to the plurality of training samples based on the preliminary identification model and the feature information for each of the plurality of

training samples; determining whether the plurality of sample probabilities satisfy a second predetermined condition; and designating the preliminary identification model as the trained identification model in response to the determination that the plurality of sample probabilities satisfy the second predetermined condition.

[0013] In some embodiments, the trained identification model may include a support vector machine model.

[0014] In a second aspect of the present disclosure, a system for identifying proper noun included in target queries is provided. The system may include at least one storage medium and at least one processor in communication with the at least one storage medium. The at least one storage medium may include a set of instructions. When executing the set of instructions, the at least one processor may be directed to: obtain a search request including a target query; and determine whether the target query includes at least one proper noun based on a trained identification model, wherein the trained identification model is configured to provide a probability that the target query includes at least one proper noun; and in response to a determination that the probability is greater than a predetermined probability threshold, determine that the target query includes at least one proper noun.

[0015] In some embodiments, the trained identification model may be determined with a training process, and the training process may include: obtaining a plurality of first historical search records, wherein each of the plurality of first historical search records includes a first query from a first user, and a first TOI selected by the first user as a first search term of the first historical search record, and at least a portion of the first TOIs includes at least one proper noun; obtaining a plurality of training samples based on the plurality of first historical search records; and training a preliminary identification model based on the plurality of training samples to generate the trained identification model.

[0016] In some embodiments, the obtaining a plurality of training samples based on the plurality of first historical search records may include: for each of the plurality of first historical search records, segmenting the first TOI into a plurality of segments in the first historical search record; extracting feature information associated with the plurality of segments; determining whether the feature information satisfies a first predetermined condition; and designating the first historical search records as a candidate search record if the feature information satisfies a first predetermined condition; and determining the plurality of training samples based on the candidate search records.

[0017] In some embodiments, the feature information of each of the plurality of segments may include at least one of a first consistent parameter associated with the first query and the segment, a second consistent parameter associated with the first TOI and the segment, a cohesive parameter associated with the segment, a left entropy associated with the segment, a right entropy associated with the segment, a degree of freedom associated with the segment, a probability that the segment is at a start of the first TOI or an end of the first TOI, or a frequency associated the segment.

[0018] In some embodiments, the system may further include: manually filtering out at least one candidate search record among the candidate search records; and determining the plurality of training samples based on the filtered candidate search records.

[0019] In some embodiments, the training a preliminary identification model based on the plurality of training samples to generate the trained identification model may include: determining a plurality of sample probabilities corresponding to the plurality of training samples based on the preliminary identification model and the feature information for each of the plurality of training samples; determining whether the plurality of sample probabilities satisfy a second predetermined first condition; and designating the preliminary identification model as the trained identification model in response to the

determination that the plurality of sample probabilities satisfy the second predetermined condition.

[0020] In some embodiments, the trained identification model may include a support vector machine model.

[0021] In some embodiments, the at least one processor may be directed further to: segment the target query based on the at least one proper noun in response to the determination that the probability is greater than a predetermined probability threshold.

[0022] In a third aspect of the present disclosure, a system for segmenting target queries is provided. The system may include at least one storage medium and at least one processor in communication with the at least one storage medium. The at least one storage medium may include a set of instructions. When executing the set of instructions, the at least one processor may be directed to: obtain a search query including a target query; determine one or more terms of interest (TOIs) associated with the target query; and rank at least a portion of the one or more TOIs based on a proper noun list including a plurality of proper nouns.

[0001] In a fourth aspect of the present disclosure, a method for segmenting target queries is provided. The method may be implemented on a computing device having at least one processor, at least one storage medium, and a communication platform connected to a network. The method may include: obtaining a search request including a target query; and segmenting the target query at least based on a proper noun list including a plurality of proper nouns, wherein the proper noun list is provided by: obtaining a plurality of first historical search records, wherein each of the plurality of first historical search records includes a first query from a first user, and a first TOI selected by the user as a first search term of the first historical search record, and at least a portion of the first TOIs include at least one proper noun; obtaining a trained identification model, wherein the trained identification model is configured to

provide a probability that the first TOI includes at least one proper noun; and determining the proper noun list at least based on the trained identification model and the plurality of first historical search records.

[0002] In some embodiments, the segmenting the target query at least based on a proper noun list including a plurality of proper nouns may include:

segmenting the target query into a plurality of segments based on a segmenting technique, wherein the segmenting technique includes an N-gram technique; comparing the plurality of segments with the proper noun list; and combining at least two segments in response to a determination that the at least two combined segments is a proper noun on the proper noun list.

[0003] In some embodiments, the segmenting the target query at least based on a proper noun list including a plurality of proper nouns may include:

obtaining a segmentation model, wherein the segmentation model is configured to segment the target query; merging the proper noun list into the segmentation model by allocating a weighting coefficient to the proper noun list; and segmenting the target query based on the merged segmentation model.

[0004] In some embodiments, the determining the proper noun list at least based on the trained identification model and the plurality of first historical search records may include: for each of the plurality of first historical search records; determining a first probability that the first TOI includes at least one proper noun; and determining whether the first probability is greater than a first predetermined probability threshold; and adding the proper nouns into the proper noun list if the first probability is greater than the first predetermined probability threshold.

[0005] In some embodiments, the trained identification model may be determined with a training process, and the training process may include: obtaining a plurality of second historical search records, wherein each of the plurality of second historical search records includes a second query from a

second user, and a second TOI selected by the user as a second search term of the second historical search record, and at least a portion of the second TOIs include at least one proper noun; obtaining a plurality of training samples based on the plurality of second historical search records; and training a preliminary identification model based on the plurality of training samples to generate the trained identification model.

[0006] In some embodiments, the obtaining a plurality of training samples based on the plurality of second historical search records may include: for each of the plurality of second historical search records, segmenting the second TOI into a plurality of second segments in the second historical search record; extracting feature information associated with the plurality of second segments; determining whether the feature information satisfies a first predetermined condition; and designating the second historical search record as a candidate search record if the feature information satisfies the first predetermined condition; and determining the plurality of training samples based on the candidate search records.

[0007] In some embodiments, the feature information of each of the plurality of segments may include at least one of a first consistent parameter associated with the second query and the segment, a second consistent parameter associated with the second TOI and the segment, a cohesive parameter associated with the segment, a left entropy associated with the segment, a right entropy associated with the segmentation, a degree of freedom associated with the segment, a probability that the segment is at a start of the query or an end of the second TOI, and a frequency associated with the segment.

[0008] In some embodiments, the determining the plurality of training samples based on the candidate search records may include: filtering out at least one candidate search record among the second historical search records manually; and determining the plurality of training samples based on

the filtered candidate search records.

[0009] In some embodiments, the training a preliminary identification model based on the plurality of training samples to generate the trained identification model may include: determining a plurality of sample probabilities corresponding to the plurality of training samples based on the preliminary identification model and the feature information for each of the plurality of training samples; determining whether the plurality of sample probabilities satisfy a predetermined second condition; and designating the preliminary identification model as the trained identification model in response to the determination that the plurality of sample probabilities satisfy the predetermined condition.

[0010] In some embodiments, the trained identification model may include a support vector machine model.

[0011] In a fifth aspect of the present disclosure, a method for segmenting target queries is provided. The method may be implemented on a computing device having at least one processor, at least one storage medium, and a communication platform connected to a network. The method may include: obtaining a search request including a target query; and determining whether the target query includes at least one proper noun based on a trained identification model, wherein the trained identification model is configured to provide a probability that the target query includes at least one proper noun; and in response to a determination that the probability is greater than a predetermined probability threshold, determining that the target query includes at least one proper noun.

[0012] In some embodiments, the trained identification model may be determined with a training process, and the training process may include: obtaining a plurality of first historical search records, wherein each of the plurality of first historical search records includes a first query from a first user, and a first TOI selected by the user as a first search term of the first historical

search record, and at least a portion of the first TOIs includes at least one first proper noun; obtaining a plurality of training samples based on the plurality of first historical search records; and training a preliminary identification model based on the plurality of training samples to generate the trained identification model.

[0013] In some embodiments, the obtaining a plurality of training samples based on the plurality of first historical search records may include: for each of the plurality of first historical search records, segmenting the first TOI into a plurality of segments in the first historical search record; extracting feature information associated with the plurality of segments; determining whether the feature information satisfies a first predetermined condition; and designating the first historical search records as a candidate search record if the feature information satisfies a first predetermined condition; and determining the plurality of training samples based on the candidate search records.

[0014] In some embodiments, the feature information of each of the plurality of segments may include at least one of a first consistent parameter associated with the first query and the segment, a second consistent parameter associated with the first TOI and the segment, a cohesive parameter associated with the segment, a left entropy associated with the segment, a right entropy associated with the segment, a degree of freedom associated with the segment, a probability that the segment is at a start of the query or an end of the first TOI, or a frequency associated the segment.

[0015] In some embodiments, the determining the plurality of training samples based on the candidate search records may include: filtering out at least one candidate search record among the first historical search records manually; and determining the plurality of training samples based on the filtered candidate search records.

[0016] In some embodiments, the training a preliminary identification model based on the plurality of training samples to generate the trained identification

model may include: determining a plurality of sample probabilities corresponding to the plurality of training samples based on the preliminary identification model and the feature information for each of the plurality of training samples; determining whether the plurality of sample probabilities satisfy a predetermined first condition; and designating the preliminary identification model as the trained identification model in response to the determination that the plurality of sample probabilities satisfy the predetermined condition.

[0017] In some embodiments, the trained identification model may include a support vector machine model.

[0018] In some embodiments, the method may further include: segmenting the target query based on the at least one proper noun in response to the determination that the probability is greater than a predetermined probability threshold.

[0019] In a sixth aspect of the present disclosure, a method for segmenting target queries is provided. The method may include: obtaining a search request including a target query; determining one or more terms of interest (TOIs) associated with the target query; and ranking at least a portion of the one or more TOIs based on a proper noun list including a plurality of proper nouns.

[0020] In a seventh aspect of the present disclosure, a non-transitory computer readable medium for segmenting target queries is provided. The non-transitory computer readable medium, including executable instructions that, when executed by at least one processor, may direct the at least one processor to perform a method. The method may include: obtaining a search request including a target query; and segmenting the target query at least based on a proper noun list including a plurality of proper nouns, wherein the proper noun list is provided by: obtaining a plurality of first historical search records, wherein each of the plurality of first historical search

records includes a first query from a first user, and a first TOI selected by the user as a first search term of the first historical search record, and at least a portion of the first TOIs include at least one proper noun; obtaining a trained identification model, wherein the trained identification model is configured to provide a probability that the first TOI includes at least one proper noun; and determining the proper noun list at least based on the trained identification model and the plurality of first historical search records.

[0021] In an eighth aspect of the present disclosure, a non-transitory computer readable medium for identifying proper noun included in target queries is provided. The non-transitory computer readable medium, including executable instructions that, when executed by at least one processor, may direct the at least one processor to perform a method. The method may include: obtaining a search request including a target query; and determining whether the target query includes at least one proper noun based on a trained identification model, wherein the trained identification model is configured to provide a probability that the target query includes at least one proper noun; and in response to a determination that the probability is greater than a predetermined probability threshold, determining that the target query includes at least one proper noun.

[0022] In a ninth aspect of the present disclosure, a non-transitory computer readable medium for segmenting target queries is provided. The non-transitory computer readable medium, including executable instructions that, when executed by at least one processor, may direct the at least one processor to perform a method. The method may include: obtaining a search query including a target query; determining one or more terms of interest (TOIs) associated with the target query; and ranking at least a portion of the one or more TOIs based on a proper noun list including a plurality of proper nouns.

[0023] Additional features will be set forth in part in the description which

follows, and in part will become apparent to those skilled in the art upon examination of the following and the accompanying drawings or may be learned by production or operation of the examples. The features of the present disclosure may be realized and attained by practice or use of various aspects of the methodologies, instrumentalities, and combinations set forth in the detailed examples discussed below.

BRIEF DESCRIPTION OF THE DRAWINGS

[0024] The present disclosure is further described in terms of exemplary embodiments. These exemplary embodiments are described in detail with reference to the drawings. These embodiments are non-limiting exemplary embodiments, in which like reference numerals represent similar structures throughout the several views of the drawings, and wherein:

[0025] FIG. 1 is a schematic diagram illustrating an exemplary proper noun identification system according to some embodiments of the present disclosure;

[0026] FIG. 2 is a schematic diagram illustrating exemplary hardware and/or software components of a computing device according to some embodiments of the present disclosure;

[0027] FIG. 3 is a schematic diagram illustrating exemplary hardware and/or software components of a mobile device according to some embodiments of the present disclosure;

[0028] FIG. 4 is a block diagram illustrating an exemplary processing engine according to some embodiments of the present disclosure;

[0029] FIG. 5 is a flowchart illustrating an exemplary process for segmenting a target query at least based on a proper noun list including a plurality of proper nouns according to some embodiments of the present disclosure;

[0030] FIG. 6 is a flowchart illustrating an exemplary process for determining a proper noun list including a plurality of proper nouns according to some

embodiments of the present disclosure;

[0031] FIG. 7 is a flowchart illustrating an exemplary process for determining a trained identification model according to some embodiments of the present disclosure;

[0032] FIG. 8 is a flowchart illustrating an exemplary process for identifying at least one proper noun included in a target query according to some embodiments of the present disclosure; and

[0033] FIG. 9 is a flowchart illustrating an exemplary process for ranking TOIs based on a proper noun list including a plurality of proper nouns according to some embodiments of the present disclosure.

DETAILED DESCRIPTION

[0034] The following description is presented to enable any person skilled in the art to make and use the present disclosure and is provided in the context of a particular application and its requirements. Various modifications to the disclosed embodiments will be readily apparent to those skilled in the art, and the general principles defined herein may be applied to other embodiments and applications without departing from the spirit and scope of the present disclosure. Thus, the present disclosure is not limited to the embodiments shown but is to be accorded the widest scope consistent with the claims.

[0035] The terminology used herein is for the purpose of describing particular example embodiments only and is not intended to be limiting. As used herein, the singular forms “a,” “an,” and “the” may be intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms “comprise,” “comprises,” and/or “comprising,” “include,” “includes,” and/or “including,” when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations,

elements, components, and/or groups thereof.

[0036] These and other features, and characteristics of the present disclosure, as well as the methods of operation and functions of the related elements of structure and the combination of parts and economies of manufacture, may become more apparent upon consideration of the following description with reference to the accompanying drawings, all of which form a part of this disclosure. It is to be expressly understood, however, that the drawings are for the purpose of illustration and description only and are not intended to limit the scope of the present disclosure. It is understood that the drawings are not to scale.

[0037] The flowcharts used in the present disclosure illustrate operations that systems implement according to some embodiments of the present disclosure. It is to be expressly understood, the operations of the flowchart may be implemented not in order. Conversely, the operations may be implemented in inverted order, or simultaneously. Moreover, one or more other operations may be added to the flowcharts. One or more operations may be removed from the flowcharts.

[0038] The terms “requester,” “requestor,” “service requester,” and “service requestor,” “search requester,” and “search requestor” in the present disclosure are used interchangeably to refer to an individual, an entity or a tool that may request or order a service. The term “user” in the present disclosure may refer to an individual, an entity or a tool that may request a service, order a service, provide a service, or facilitate the providing of the service. In the present disclosure, terms “requester” and “requester terminal” may be used interchangeably.

[0039] The terms “request,” “service,” “service request,” and “search request” in the present disclosure are used interchangeably to refer to a request that may be initiated by a requester, a service requester, a search requester, or the like, or any combination thereof. The search request or the service request

may be chargeable or free.

[0040] The present invention is to provide systems and methods, which are more efficient than traditional ones, to identify proper nouns in inputted queries. Those proper nouns are not frequently used, or appear in short queries and TOI strings, which result in insufficient context information.

[0041] An aspect of the present disclosure relates to systems for identifying at least one proper noun included in a target query associated with a search request. The system may identify the at least one proper noun based on a trained identification model. The trained identification model may be configured to determine a probability that a term (e.g., the target query) include the at least one proper noun. In response to a determination that the probability is greater than a predetermined threshold, the system may determine that the target query may include the at least one proper noun. In this disclosure, the trained identification model may be trained based on a plurality of first historical search records.

[0042] In addition, the system may determine a proper noun list including a plurality of proper nouns offline based on the trained identification model and a plurality of second historical search records. The system may determine the proper noun list based on a plurality of probabilities corresponding to the plurality of second historical search records determined by the trained identification model. Further, the system may use the proper noun list for segmenting the target query based on a segmenting technique or a segmentation model.

[0043] FIG. 1 is a schematic diagram illustrating an exemplary proper noun identification system according to some embodiments of the present disclosure. For example, the proper noun identification system 100 may be an online search platform for services, e.g., a transportation service, an online shopping service, a map (e.g., GOOGLE Map, BAIDU Map, TENCENT Map) navigation service, a meal ordering service, etc. The proper noun

identification system 100 may include a server 110, a network 120, a user terminal 130, and a storage 140.

[0044] In some embodiments, the server 110 may be a single server, or a server group. The server group may be centralized, or distributed (e.g., server 110 may be a distributed system). In some embodiments, the server 110 may be local or remote. For example, the server 110 may access information and/or data stored in the user terminal 130 and/or the storage 140 via the network 120. As another example, the server 110 may be directly connected to the user terminal 130 and/or the storage 140 to access stored information and/or data. In some embodiments, the server 110 may be implemented on a cloud platform. Merely by way of example, the cloud platform may include a private cloud, a public cloud, a hybrid cloud, a community cloud, a distributed cloud, an inter-cloud, a multi-cloud, or the like, or any combination thereof. In some embodiments, the server 110 may be implemented on a computing device 200 having one or more components illustrated in FIG. 2.

[0045] In some embodiments, the server 110 may include a processing engine 112. The processing engine 112 may process information and/or data relating to a search request to perform one or more functions described in the present disclosure. For example, the processing engine 112 may segment a target query at least based on a proper noun list including a plurality of proper nouns. In some embodiments, the processing engine 112 may include one or more processing engines (e.g., single-core processing engine(s) or multi-core processor(s)). The processing engine 112 may include a central processing unit (CPU), an application-specific integrated circuit (ASIC), an application-specific instruction-set processor (ASIP), a graphics processing unit (GPU), a physics processing unit (PPU), a digital signal processor (DSP), a field programmable gate array (FPGA), a programmable logic device (PLD), a controller, a microcontroller unit, a

reduced instruction-set computer (RISC), a microprocessor, or the like, or any combination thereof.

[0046] The network 120 may facilitate exchange of information and/or data. In some embodiments, one or more components of the proper noun identification system 100 (e.g., the server 110, the user terminal 130, or the storage 140) may transmit information and/or data to other component(s) of the proper noun identification system 100 via the network 120. For example, the server 110 may obtain a search request from the user terminal 130 via the network 120. In some embodiments, the network 120 may be any type of wired or wireless network, or any combination thereof. Merely by way of example, the network 120 may include a cable network, a wireline network, an optical fiber network, a telecommunications network, an intranet, an Internet, a local area network (LAN), a wide area network (WAN), a wireless local area network (WLAN), a metropolitan area network (MAN), a public telephone switched network (PSTN), a Bluetooth network, a ZigBee network, a near field communication (NFC) network, or the like, or any combination thereof. In some embodiments, the network 120 may include one or more network access points. For example, the network 120 may include wired or wireless network access points such as base stations and/or internet exchange points 120-1, 120-2, ... , through which one or more components of the proper noun identification system 100 may be connected to the network 120 to exchange data and/or information.

[0047] In some embodiments, a search requester may be a user of the user terminal 130. In some embodiments, the user of the user terminal 130 may be someone other than the search requester. For example, a user A of the user terminal 130 may use the user terminal 130 to send a search request for a user B or receive a search confirmation and/or information or instructions from the server 110.

[0048] In some embodiments, the user terminal 130 may include a mobile

device 130-1, a tablet computer 130-2, a laptop computer 130-3, a built-in device in a vehicle 130-4, or the like, or any combination thereof. In some embodiments, the mobile device 130-1 may include a smart home device, a wearable device, a smart mobile device, a virtual reality device, an augmented reality device, or the like, or any combination thereof. In some embodiments, the smart home device may include a smart lighting device, a control device of an intelligent electrical apparatus, a smart monitoring device, a smart television, a smart video camera, an interphone, or the like, or any combination thereof. In some embodiments, the wearable device may include a smart bracelet, a smart footgear, a smart glass, a smart helmet, a smart watch, a smart clothing, a smart backpack, a smart accessory, or the like, or any combination thereof. In some embodiments, the smart mobile device may include a smartphone, a personal digital assistance (PDA), a gaming device, a navigation device, a point of sale (POS) device, or the like, or any combination thereof. In some embodiments, the virtual reality device and/or the augmented reality device may include a virtual reality helmet, a virtual reality glass, a virtual reality patch, an augmented reality helmet, an augmented reality glass, an augmented reality patch, or the like, or any combination thereof. For example, the virtual reality device and/or the augmented reality device may include a Google Glass™, an Oculus Rift™, a Hololens™, a Gear VR™, etc. In some embodiments, a built-in device in the vehicle 130-4 may include an onboard computer, an onboard television, etc.

[0049] The storage 140 may store data and/or instructions relating to the search request. In some embodiments, the storage 140 may store data obtained from the user terminal 130. In some embodiments, the storage 140 may store data and/or instructions that the server 110 may execute or use to perform exemplary methods described in the present disclosure. In some embodiments, the storage 140 may include a mass storage, a removable storage, a volatile read-and-write memory, a read-only memory (ROM), or the

like, or any combination thereof. Exemplary mass storage may include a magnetic disk, an optical disk, a solid-state drive, etc. Exemplary removable storage may include a flash drive, a floppy disk, an optical disk, a memory card, a zip disk, a magnetic tape, etc. Exemplary volatile read-and-write memory may include a random access memory (RAM). Exemplary RAM may include a dynamic RAM (DRAM), a double data rate synchronous dynamic RAM (DDR SDRAM), a static RAM (SRAM), a thyristor RAM (T-RAM), and a zero-capacitor RAM (Z-RAM), etc. Exemplary ROM may include a mask ROM (MROM), a programmable ROM (PROM), an erasable programmable ROM (EPROM), an electrically erasable programmable ROM (EEPROM), a compact disk ROM (CD-ROM), and a digital versatile disk ROM, etc. In some embodiments, the storage 140 may be implemented on a cloud platform. Merely by way of example, the cloud platform may include a private cloud, a public cloud, a hybrid cloud, a community cloud, a distributed cloud, an inter-cloud, a multi-cloud, or the like, or any combination thereof.

[0050] In some embodiments, the storage 140 may be connected to the network 120 to communicate with one or more components of the proper noun identification system 100 (e.g., the server 110, the user terminal 130). One or more components of the proper noun identification system 100 may access the data and/or instructions stored in the storage 140 via the network 120. In some embodiments, the storage 140 may be directly connected to or communicate with one or more components of the proper noun identification system 100 (e.g., the server 110, the user terminal 130). In some embodiments, the storage 140 may be part of the server 110.

[0051] In some embodiments, one or more components of the proper noun identification system 100 (e.g., the server 110, the user terminal 130) may have permissions to access the storage 140. In some embodiments, one or more components of the proper noun identification system 100 may read and/or modify information relating to the search requester and/or the public

when one or more conditions are met. For example, the server 110 may read and/or modify one or more search requesters' information after a search service is completed.

[0052] In some embodiments, information exchanging of one or more components of the proper noun identification system 100 may be achieved by way of requesting a search service. The object of the search request may be any product. In some embodiments, the product may be a tangible product or an immaterial product. The tangible product may include food, medicine, commodity, chemical product, electrical appliance, clothing, car, housing, luxury, or the like, or any combination thereof. The immaterial product may include a servicing product, a financial product, a knowledge product, an internet product, or the like, or any combination thereof. The internet product may include an individual host product, a web product, a mobile internet product, a commercial host product, an embedded product, or the like, or any combination thereof. The mobile internet product may be used in a software of a mobile terminal, a program, a system, or the like, or any combination thereof. The mobile terminal may include a tablet computer, a laptop computer, a mobile phone, a personal digital assistance (PDA), a smart watch, a point of sale (POS) device, an onboard computer, an onboard television, a wearable device, or the like, or any combination thereof. For example, the product may be any software and/or application used in the computer or mobile phone. The software and/or application may relate to socializing, shopping, transporting, entertainment, learning, investment, or the like, or any combination thereof. In some embodiments, the software and/or application relating to transporting may include a traveling software and/or application, a vehicle scheduling software and/or application, a mapping software and/or application, etc. In the vehicle scheduling software and/or application, the vehicle may include a horse, a carriage, a rickshaw (e.g., a wheelbarrow, a bike, a tricycle), a car (e.g., a taxi, a bus, a private car), a

train, a subway, a vessel, an aircraft (e.g., an airplane, a helicopter, a space shuttle, a rocket, a hot-air balloon), or the like, or any combination thereof.

[0053] One of ordinary skill in the art would understand that when an element (or component) of the proper noun identification system 100 performs, the element may perform through electrical signals and/or electromagnetic signals. For example, when the user terminal 130 transmits out a search request to the server 110, a processor of the user terminal 130 may generate an electrical signal encoding the search request. The processor of the user terminal 130 may then transmit the electrical signal to an output port. If the user terminal 130 communicates with the server 110 via a wired network, the output port may be physically connected to a cable, which further may transmit the electrical signal to an input port of the server 110. If the user terminal 130 communicates with the server 110 via a wireless network, the output port of the user terminal 130 may be one or more antennas, which convert the electrical signal to electromagnetic signal. Within an electronic device, such as the user terminal 130 and/or the server 110, when a processor thereof processes an instruction, transmits out an instruction, and/or performs an action, the instruction and/or action is conducted via electrical signals. For example, when the processor retrieves or saves data from a storage medium (e.g., the storage 140), it may transmit out electrical signals to a read/write device of the storage medium, which may read or write structured data in the storage medium. The structured data may be transmitted to the processor in the form of electrical signals via a bus of the electronic device. Here, an electrical signal may refer to one electrical signal, a series of electrical signals, and/or a plurality of discrete electrical signals.

[0054] FIG. 2 is a schematic diagram illustrating exemplary hardware and/or software components of a computing device 200 according to some embodiments of the present disclosure. In some embodiments, the server

110 and/or the user terminal 130 may be implemented on the computing device 200. For example, the processing engine 112 may be implemented on the computing device 200 and configured to perform functions of the processing engine 112 disclosed in this disclosure.

[0055] The computing device 200 may be used to implement any component of the proper noun identification system 100 as described herein. For example, the processing engine 112 may be implemented on the computing device 200, via its hardware, software program, firmware, or a combination thereof. Although only one such computer is shown, for convenience, the computer functions relating to the search service as described herein may be implemented in a distributed fashion on a number of similar platforms to distribute the processing load.

[0056] The computing device 200, for example, may include COM ports 250 connected to and from a network connected thereto to facilitate data communications. The computing device 200 may also include a processor 220, in the form of one or more processors (e.g., logic circuits), for executing program instructions. For example, the processor 220 may include interface circuits and processing circuits therein. The interface circuits may be configured to receive electronic signals from a bus 210, wherein the electronic signals encode structured data and/or instructions for the processing circuits to process. The processing circuits may conduct logic calculations, and then determine a conclusion, a result, and/or an instruction encoded as electronic signals. Then the interface circuits may send out the electronic signals from the processing circuits via the bus 210.

[0057] The computing device 200 may further include program storage and data storage of different forms including, for example, a disk 270, and a read only memory (ROM) 230, or a random access memory (RAM) 240, for various data files to be processed and/or transmitted by the computing device. The exemplary computer platform may also include program instructions stored in

the ROM 230, RAM 240, and/or other type of non-transitory storage medium to be executed by the processor 220. The methods and/or processes of the present disclosure may be implemented as the program instructions. The computing device 200 also includes an I/O component 260, supporting input/output between the computer and other components. The computing device 200 may also receive programming and data via network communications.

[0058] Merely for illustration, only one processor is described in FIG. 2. Multiple processors are also contemplated, thus operations and/or method steps performed by one processor as described in the present disclosure may also be jointly or separately performed by the multiple processors. For example, if in the present disclosure the processor of the computing device 200 executes both step A and step B, it should be understood that step A and step B may also be performed by two different CPUs and/or processors jointly or separately in the computing device 200 (e.g., the first processor executes step A and the second processor executes step B, or the first and second processors jointly execute steps A and B).

[0059] FIG. 3 is a schematic diagram illustrating exemplary hardware and/or software components of a mobile device 300 on which the user terminal 130 may be implemented according to some embodiments of the present disclosure. As illustrated in FIG. 3, the mobile device 300 may include a communication platform 310, a display 320, a graphic processing unit (GPU) 330, a central processing unit (CPU) 340, an I/O 350, a memory 360, a mobile operating system (OS) 370, and a storage 390. In some embodiments, any other suitable component, including but not limited to a system bus or a controller (not shown), may also be included in the mobile device 300.

[0060] In some embodiments, the mobile operating system 370 (e.g., iOS™, Android™, Windows Phone™, etc.) and one or more applications 380 may be loaded into the memory 360 from the storage 390 in order to be executed by

the CPU 340. The applications 380 may include a browser or any other suitable mobile apps for receiving and rendering information relating to search services or other information from the proper noun identification system 100. User interactions with the information stream may be achieved via the I/O 350 and provided to the processing engine 112 and/or other components of the proper noun identification system 100 via the network 120.

[0061] FIG. 4 is a block diagram illustrating an exemplary processing engine according to some embodiments of the present disclosure. The processing engine 112 may include a request obtaining module 410, a query segmentation module 420, a list determination module 430, a model training module 440, a proper noun identification module 450, and a TOI determination module 460.

[0062] The request obtaining module 410 may be configured to obtain a search request including a target query associated with a user. In some embodiments, the search request may refer to a search request for searching a target query. Merely by way of example, the search request may include a transportation search request, an online shopping search request, a map (e.g., GOOGLE Map, BAIDU Map, TENCENT Map) navigation search request, a meal ordering search request, etc. In some embodiments, the target query may refer to a term that the user intends to search. Merely by way of example, types of the target query may include a term (e.g., an address) associated with a transportation search request or a map (e.g., GOOGLE Map, BAIDU Map, TENCENT Map) navigation search request, a term (e.g., goods) associated with an online shopping search request, a term (e.g., food) associated with a meal ordering search request, etc.

[0063] The query segmentation module 420 may be configured to segment the target query at least based on a proper noun list including a plurality of proper nouns. As used herein, the proper noun may refer to a name of a person, a name of particular place, a name of a particular organization, etc.

For example, Jay Chou may be a proper noun. As another example, Beijing University may be a proper noun. As a further example, International Committee of the Red Cross (ICRC) may be a proper noun.

[0064] In some embodiments, the query segmentation module 420 may segment the target query based on the proper noun list and a segmenting technique. For example, the segmenting technique may include an N-gram technique, a forward maximum matching technique, a reverse maximum matching technique, a bidirectional maximum matching technique, a minimum matching technique, an optimal matching technique, a hidden Markov model, a maximum entropy model, a conditional random field model, a neural network model, an association-backtracking technique, or the like, or any combination thereof.

[0065] In some embodiments, the query segmentation module 420 may segment the target query into a plurality of segments. Further, the query segmentation module 420 may compare the plurality of segments with the proper noun list. In response to a determination that, at least two adjacent segments can form a proper noun on the proper noun list, the query segmentation module 420 may then combine the at least two segments. For example, Jay Chou is on the proper noun list. If Jay Chou has been segmented into two segments using a segmenting technique, the query segmentation module 420 may then combine the two segments into a single segment, i.e., Jay Chou.

[0066] In some embodiments, the query segmentation module 420 may segment the target query based on the proper noun list and a segmentation model. The segmentation model may be used to segment a term (e.g., the target query). In some embodiments, the query segmentation module 420 may merge the proper noun list into the segmentation model by allocating a weighting coefficient to the proper noun list. The query segmentation module 420 may then segment the target query based on the merged segmentation

model. In some embodiments, the query segmentation module 420 may unsegment at least one proper noun included in the target query when the target query is segmented by the merged segmentation model. As used herein, the weighting coefficient may be default settings of the proper noun identification system 100, or may be adjustable under different situations.

[0067] The segmentation model may include a plurality of segmentation lists associated with different application scenarios, e.g., a transportation service, an online shopping service, a map navigation service, a meal ordering service. As used herein, a segmentation list may refer to a list including a plurality of segments associated with an application scenario. Merely by way of example, the plurality of segmentation lists may include a segmentation list (e.g., including terms of locations) associated with the transportation service or the map navigation service, a segmentation list (e.g., including terms of goods) associated with the online shopping service, a segmentation list (e.g., including terms of food) associated with the meal ordering service, or the like, or any combination thereof. More detailed description of the segmentation model may be described in Chinese Patent Application No. 201810554080.3 filed on July 1, 2018, the contents of which is incorporated herein by reference in their entirety.

[0068] The list determination module 430 may be configured to determine a proper noun list including a plurality of proper nouns offline. As used herein, the proper noun may refer to a name of a person, a name of particular place, a name of a particular organization, etc. For example, Jay Chou may be a proper noun. As another example, Beijing University may be a proper noun. As a further example, International Committee of the Red Cross (ICRC) may be a proper noun.

[0069] In some embodiments, the list determination module 430 may determine the proper noun list based on a plurality of first historical search records and a trained identification model. Each of the plurality of first

historical search records may include a first query from a first user, and a first TOI selected by the user as a first search term of the first historical search record. At least a portion of the first TOIs of the plurality of first historical search records may include at least one proper noun. The trained identification model may be configured to provide a probability that a term (e.g., the first TOI) includes at least one proper noun. In some embodiments, the list determination module 430 may obtain the plurality of first historical search records or the trained identification model from a storage device (e.g., the storage 140) disclosed elsewhere in the present disclosure, a third party (e.g., an external database), etc.

[0070] The model training module 440 may be configured to determine a trained identification model offline based on a plurality of second historical search records. Similar to the first historical search records, each of the plurality of second historical search records may include a second query from a second user, and a second TOI selected by the user as a second search term of the second historical search record. At least a portion of the second TOIs may include at least one proper noun. In some embodiments, the plurality of second historical search records may be partially or totally different from the plurality of first historical search records. More detailed description of the trained identification model may be found elsewhere in the present disclosure (e.g., FIG. 7 and the descriptions thereof).

[0071] The proper noun identification module 450 may be configured to determine whether a target query includes at least one proper noun based on the trained identification model online. In some embodiments, the proper noun identification module 450 may compare the probability with a predetermined probability threshold. In response to a determination that the probability is greater than the predetermined probability threshold, the proper noun identification module 450 may determine that the target query includes at least one proper noun. As used herein, the predetermined probability

threshold may be default settings of the proper noun identification system 100, or may be adjustable under different situations.

[0072] The TOI determination module 460 may be configured to determine one or more terms of interest (TOIs) associated with a target query. As used herein, the TOI may refer to a term associated with a target query that the user may be interested in. In some embodiments, the TOI determination module 460 may first determine at least one of a prefix, a key term, or a phase in the target query and determine the one or more TOIs based on the prefix, the key term, or the phase.

[0073] In some embodiments, the TOI determination module 460 may determine correlation coefficients for the one or more TOIs. As used herein, the correlation coefficient may indicate a similarity between the target query and the TOI. The larger the correlation coefficient is, the higher the similarity between the TOI and the target query may be, and the greater a probability that a user selects the TOI as a search term associated with the target query may be.

[0074] In some embodiments, the TOI determination module 460 may rank at least a portion of the one or more TOIs based on a proper noun list including a plurality of proper nouns. In some embodiments, the TOI determination module 460 may compare the target query with proper nouns on the proper noun list. In response to a determination that the target query includes at least one proper noun on the proper noun list, the TOI determination module 460 may rank TOIs that include the at least one proper noun in front. In some embodiments, the TOI determination module 460 may allocate at least one weighting coefficient to the at least one proper noun. The TOI determination module 460 may rank the at least one of the one or more TOIs based on the at least one weighting coefficient and the correlation coefficients for the TOIs. In response to a determination that the target query does not include at least one proper noun on the proper noun list, the TOI

determination module 460 may rank the at least a portion of the one or more TOIs based on the correlation coefficients for the TOIs (e.g., from large to small).

[0075] The modules in the processing engine 112 may be connected to or communicated with each other via a wired connection or a wireless connection. The wired connection may include a metal cable, an optical cable, a hybrid cable, or the like, or any combination thereof. The wireless connection may include a Local Area Network (LAN), a Wide Area Network (WAN), a Bluetooth, a ZigBee, a Near Field Communication (NFC), or the like, or any combination thereof. Two or more of the modules may be combined into a single module, and any one of the modules may be divided into two or more units. For example, the processing engine 112 may include a storage module (not shown) which may be used to store data generated by the above-mentioned modules, for example, a proper noun list and a trained identification model. As another example, the model training module 440 may be unnecessary and the trained identification model may be obtained from a storage device (e.g., the storage 140) disclosed elsewhere in the present disclosure or an external device in communication with the proper noun identification system 100.

[0076] FIG. 5 is a flowchart illustrating an exemplary process for segmenting a target query at least based on a proper noun list including a plurality of proper nouns according to some embodiments of the present disclosure. In some embodiments, process 500 may be implemented as a set of instructions (e.g., an application) stored in the storage ROM 230 or RAM 240. The processor 220 and/or the modules in FIG. 4 may execute the set of instructions, and when executing the instructions, the processor 220 and/or the modules may be configured to perform process 500. The operations of the illustrated process presented below are intended to be illustrative. In some embodiments, process 500 may be accomplished with one or more

additional operations not described and/or without one or more of the operations herein discussed. Additionally, the order in which the operations of the process as illustrated in FIG. 5 and described below is not intended to be limiting.

[0077] In 510, the processing engine 112 (e.g., the request obtaining module 410) (e.g., the processing circuits of the processor 220) may obtain a search request including a target query associated with a user. In some embodiments, the search request may refer to a search request for searching a target query. Merely by way of example, the search request may include a transportation search request, an online shopping search request, a map (e.g., GOOGLE Map, BAIDU Map, TENCENT Map) navigation search request, a meal ordering search request, etc. In some embodiments, the target query may refer to a term that the user intends to search. Merely by way of example, types of the target query may include a term (e.g., an address) associated with a transportation search request or a map (e.g., GOOGLE Map, BAIDU Map, TENCENT Map) navigation search request, a term (e.g., goods) associated with an online shopping search request, a term (e.g., food) associated with a meal ordering search request, etc.

[0078] In some embodiments, the target query may be inputted by the user via the user terminal 130. For example, the user may input the target query in an application installed on the user terminal 130. In some embodiments, the user may input the target query via a typing input, a hand input, a voice input, a picture input, etc. The input may be done through an application or a device. The input device may be a key board, touch screen, microphone, hand-writing board, scanner, camera, or any combination thereof.

[0079] In 520, the processing engine 112 (e.g., the query segmentation module 420) (e.g., the processing circuits of the processor 220) may segment the target query at least based on a proper noun list including a plurality of proper nouns. As used herein, the proper noun may refer to a name of a

person, a name of particular place, a name of a particular organization, etc. For example, Jay Chou may be a proper noun. As another example, Peking University may be a proper noun. As a further example, International Committee of the Red Cross (ICRC) may be a proper noun.

[0080] In some embodiments, the proper noun list may be provided by the proper noun identification system 100 (e.g., the processing engine 112) or a third party (e.g., an external database). The proper noun identification system 100 may predetermine the proper noun list offline and store the proper noun list in a storage device (e.g., the storage 140) disclosed elsewhere in the present disclosure.

[0081] In some embodiments, the proper noun identification system 100 may predetermine the proper noun list based on a plurality of first historical search records and a trained identification model. Each of the plurality of first historical search records may include a first query from a first user, and a first TOI selected by the user as a first search term of the first historical search record. At least a portion of the first TOIs of the plurality of first historical search records may include at least one proper noun. The trained identification model may be configured to provide a probability that a term (e.g., the first TOI) includes at least one proper noun of the each of the plurality of first historical search records. In some embodiments, if the probability is greater than a predetermined probability threshold, the proper noun identification system 100 may add the at least one proper noun on the proper noun list. More detailed descriptions of predetermining the proper noun list can be found elsewhere in the present disclosure, e.g, FIG. 6 and the descriptions thereof.

[0082] In some embodiments, the plurality of first historical search records and/or the trained identification model may be obtained e.g., from a storage device (e.g., the storage 140), the third party, etc. The proper noun identification system 100 may predetermine the trained identification model

offline and store the trained identification model in a storage device (e.g., the storage 140) disclosed elsewhere in the present disclosure. More detailed descriptions of predetermining the trained identification model can be found elsewhere in the present disclosure, e.g., FIG. 7 and the descriptions thereof.

[0083] In some embodiments, the processing engine 112 may segment the target query based on the proper noun list and a segmenting technique. For example, the segmenting technique may include an N-gram technique, a forward maximum matching technique, a reverse maximum matching technique, a bidirectional maximum matching technique, a minimum matching technique, an optimal matching technique, a hidden Markov model, a maximum entropy model, a conditional random field model, a neural network model, an association-backtracking technique, or the like, or any combination thereof.

[0084] In some embodiments, the processing engine 112 may segment the target query into a plurality of segments. Further, the processing engine 112 may compare the plurality of segments with the proper noun list. In response to a determination that, at least two adjacent segments can form a proper noun on the proper noun list, the processing engine 112 may then combine the at least two segments. For example, Jay Chou is on the proper noun list. If Jay Chou has been segmented into two segments using a segmenting technique, the processing engine 112 may then combine the two segments into a single segment, i.e., Jay Chou.

[0085] In some embodiments, the processing engine 112 may segment the target query based on the proper noun list and a segmentation model. The segmentation model may be used to segment a term (e.g., the target query). In some embodiments, the processing engine 112 may merge the proper noun list into the segmentation model by allocating a weighting coefficient to the proper noun list. The processing engine 112 may then segment the target query based on the merged segmentation model. In some

embodiments, the processing engine 112 may unsegment at least one proper noun included in the target query when the target query is segmented by the merged segmentation model. As used herein, the weighting coefficient may be default settings of the proper noun identification system 100, or may be adjustable under different situations.

[0086] The segmentation model may include a plurality of segmentation lists associated with different application scenarios, e.g., a transportation service, an online shopping service, a map navigation service, a meal ordering service. As used herein, a segmentation list may refer to a list including a plurality of segments associated with an application scenario. Merely by way of example, the plurality of segmentation lists may include a segmentation list (e.g., including terms of locations) associated with the transportation service or the map navigation service, a segmentation list (e.g., including terms of goods) associated with the online shopping service, a segmentation list (e.g., including terms of food) associated with the meal ordering service, or the like, or any combination thereof. More detailed description of the segmentation model may be described in Chinese Patent Application No. 201810554080.3 filed on July 1, 2018, the contents of which is incorporated herein by reference in their entirety.

[0087] It should be noted that the above description is merely provided for the purposes of illustration, and not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations or modifications may be made under the teachings of the present disclosure. However, those variations and modifications do not depart from the scope of the present disclosure.

[0088] FIG. 6 is a flowchart illustrating an exemplary process for determining a proper noun list including a plurality of proper nouns according to some embodiments of the present disclosure. In some embodiments, process 600 may be implemented as a set of instructions (e.g., an application) stored in

the storage ROM 230 or RAM 240. The processor 220 and/or the modules in FIG. 4 may execute the set of instructions, and when executing the instructions, the processor 220 and/or the modules may be configured to perform process 600. The operations of the illustrated process presented below are intended to be illustrative. In some embodiments, process 600 may be accomplished with one or more additional operations not described and/or without one or more of the operations herein discussed. Additionally, the order in which the operations of the process as illustrated in FIG. 6 and described below is not intended to be limiting.

[0089] In 610, the processing engine 112 (e.g., the list determination module 430) (e.g., the processing circuits of the processor 220) may obtain a plurality of first historical search records within a predetermined time period (e.g., the last month, the last three months, the last year). In some embodiments, the processing engine 112 may obtain the plurality of first historical search records from a storage device (e.g., the storage 140) disclosed elsewhere in the present disclosure, an external database, etc.

[0090] Each of the plurality of first historical search records may include a first query from a first user, and a first TOI selected by the user as a first search term of the first historical search record. At least a portion of the first TOIs of the plurality of first historical search records may include at least one proper noun.

[0091] In 620, the processing engine 112 (e.g., the list determination module 430) (e.g., the processing circuits of the processor 220) may obtain a trained identification model (e.g., a support vector machine model). The trained identification model may be configured to provide a probability that a term (e.g., the first TOI) includes at least one proper noun. In some embodiments, the processing engine 112 may obtain the plurality of first historical search records from a storage device (e.g., the storage 140) disclosed elsewhere in the present disclosure, a third party, etc.

[0092] In 630, the processing engine 112 (e.g., the list determination module 430) (e.g., the processing circuits of the processor 220) may determine the proper noun list at least based on the trained identification model and the plurality of first historical search records. In some embodiments, the processing engine 112 may determine the proper noun list based on first probabilities determined based on the trained identification model for each of the plurality of first historical search records. Further, for each of the plurality of first historical search records, the processing engine 112 may determine whether the corresponding first probability is greater than a first predetermined probability threshold. In response to a determination that the corresponding first probability is greater than a first predetermined probability threshold, the processing engine 112 may add the at least one proper noun included in the first TOI into the proper noun list. As used herein, the first predetermined probability threshold may be default settings of the proper noun identification system 100, or may be adjustable under different situations.

[0093] It should be noted that the above description is provided for the purpose of illustration, and is not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations and modifications may be made under the teaching of the present disclosure. However, those variations and modifications do not depart from the scope of the present disclosure. For example, one or more other optional steps (e.g., a storing step) may be added elsewhere in process 600. In the storing step, the processing engine 112 may store the proper noun list in a storage device (e.g., the storage 140) disclosed elsewhere in the present disclosure.

[0094] FIG. 7 is a flowchart illustrating an exemplary process for determining a trained identification model according to some embodiments of the present disclosure. In some embodiments, process 700 may be implemented as a set of instructions (e.g., an application) stored in the storage ROM 230 or

RAM 240. The processor 220 and/or the modules in FIG. 4 may execute the set of instructions, and when executing the instructions, the processor 220 and/or the modules may be configured to perform process 700. The operations of the illustrated process presented below are intended to be illustrative. In some embodiments, process 700 may be accomplished with one or more additional operations not described and/or without one or more of the operations herein discussed. Additionally, the order in which the operations of the process as illustrated in FIG. 7 and described below is not intended to be limiting.

[0095] In 710, the processing engine 112 (e.g., the model training module 440) (e.g., the processing circuits of the processor 220) may obtain a plurality of second historical search records. In some embodiments, the processing engine 112 may obtain the plurality of second historical search records from a storage device (e.g., the storage 140) disclosed elsewhere in the present disclosure, an external database, etc.

[0096] As described in connection with 610, similar to the first historical search records, each of the plurality of second historical search records may include a second query from a second user, and a second TOI selected by the user as a second search term of the second historical search record. At least a portion of the second TOIs may include at least one proper noun. In some embodiments, the plurality of second historical search records may be partially or totally different from the plurality of first historical search records.

[0097] In 720, the processing engine 112 (e.g., the model training module 440) (e.g., the processing circuits of the processor 220) may obtain a plurality of training samples based on the plurality of second historical records. In some embodiments, the processing engine 112 may extract feature information associated with each of the plurality of second historical search records. The processing engine 112 may determine the plurality of training samples based on the feature information and the plurality of second historical

records.

[0098] In some embodiments, the processing engine 112 may first segment the second TOI into a plurality of second segments in the second historical search record for each of the plurality of second historical search records. As described in connection with the operation 520, the processing engine 112 may segment the second TOI based on a segmenting technique or a segmentation model. More detailed descriptions can be found in 520 and not be repeated here.

[0099] The processing engine 112 may then determine the feature information associated with each of the plurality of second historical search records based on the plurality of second segments and the second historical query. Merely by way of example, for a second segment of the plurality of second segments, the feature information may include a first consistent parameter associated with the second query and the second segment, a second consistent parameter associated with the second TOI and the second segment, a cohesive parameter associated with the second segment, a left entropy associated with the second segment, a right entropy associated with the second segment, a degree of freedom associated with the second segment, a probability that the second segment is at a start of the query or an end of the second TOI, a frequency associated the second segment, etc.

[0100] As used herein, the first consistent parameter may indicate a probability that at least two subsequent segments are a single term in the second historical queries. Taking two segments as an example, the probability may be a ratio of a first count of second historical queries that include segment 1 and segment 2, and segment 1 and segment 2 are subsequent and are a single term to a first count of second historical queries that include segment 1 and segment 2. If the first count is n , and the second count is m , the consistent parameter of the at least two segments may be $n/m \times 100\%$. Similar to the first consistent parameter, the second consistent

parameter may indicate a probability that at least two sequential segments are a single term in the second historical TOIs.

[0101] The cohesive parameter may indicate a correlation or a cohesion between at least two segments, i.e., a probability that the at least two segments may constitute of a single term. For example, a cohesive parameter of Lou Zhu is greater than a cohesive parameter of Lou Men. In some embodiments, the processing engine 112 may determine the cohesive parameter according to Equation (1) below:

$$C = \min\left(\frac{P_1}{P_{left1} \times P_{right2}}, \frac{P_1}{P_{left3} \times P_{right4}}\right) \quad (1)$$

wherein C refers to the cohesive parameter associated with a term including segment 1 and segment 2 or segment 3 and segment 4, P_1 refers to a probability of the term included in a corpus database, P_{left1} refers to a probability of the segment 1 included in the corpus database, P_{right2} refers to a probability of the segment 2 included in the corpus database, P_{left3} refers to a probability of the segment 3 included in the corpus database, and P_{right4} refers to a probability of the segment 4 is included in the corpus database..

As used herein, the corpus database may include a collection of written or spoken texts and be used for language research. The processing engine 112 may obtain the corpus database from a storage device (e.g., the storage 140) disclosed elsewhere in the present disclosure or an external device in communication with the proper noun identification system 100.

[0102] The left entropy may indicate a variety of combinations between a reference segment and a segment prior to the reference segment. The right entropy may indicate a variety of combinations between a reference segment and a segment next to the reference segment. In some embodiments, the processing engine 112 may determine the left entropy and the right entropy according to Equation (2) and Equation (3) respectively:

$$L = \sum p(m_i) \times \log(p(m_i)) \quad (2)$$

$$R = \sum p(n_i) \times \log(p(n_i)) \quad (3)$$

wherein L refers to the left entropy, m_i refers to an i^{th} second segment in a set of second segments that are left to a first segment in the corpus database respectively, $P(m_i)$ refers to a probability that the i^{th} second segment is left to the first segment in the corpus database, R refers to the right entropy, n_i refers to a j^{th} third segment in a set of third segments that are right to the first segment in the corpus database respectively, and $P(n_i)$ refers to a probability that the j^{th} third segment is right to the first segment in the corpus database.

[0103] The degree of freedom may indicate a variety of combinations between a reference segments and other segments. The degree of freedom may be associated with the left entropy or the right entropy. In some embodiments, the processing engine 112 may determine the degree of freedom according to Equation (4):

$$F = \min(L, R) \quad (4)$$

wherein F refers to the degree of freedom, L refers to the left entropy, and R refers to the right entropy.

[0104] In some embodiments, the processing engine 112 may determine the plurality of training samples based on candidate second historical search records and feature information corresponding to the candidate second historical search records. The processing engine 12 may select the candidate second historical search records based on the feature information associated with the plurality of second historical search records. For a candidate second historical search record, feature information associated with the candidate second historical search record may satisfy a first predetermined condition. In some embodiments, the first predetermined condition may include at least one of the feature information greater than a predetermined threshold respectively. For example, the first predetermined condition may include the first consistent parameter greater than a predetermined threshold. As another example, the first predetermined condition may include the degree of freedom greater than a predetermined

threshold. As a further example, the first predetermined condition may include the first consistent parameter greater than a first predetermined threshold, and the degree of freedom greater than a second predetermined threshold.

[0105] In some embodiments, the processing engine 112 may determine the plurality of training samples based on filtered (candidate) second historical search records. The (candidate) second historical search records may be filtered manually. A provider of the proper noun identification system 100 may search data e.g., via Internet, and determine whether the corresponding (candidate) second TOI may factually include the corresponding at least one proper noun.

[0106] In some embodiments, the processing engine 112 may determine a plurality of positive training sample and a plurality of negative sample based on the second historical search records, the candidate second historical search records, or the filtered (candidate) second historical records. As used herein, each of the plurality of positive training samples may include a second historical query and a second TOI including at least one proper noun. Each of the plurality of negative training sample may include a second historical query and a second TOI excluding at least one proper noun.

[0107] In 730, the processing engine 112 (e.g., the model training module 440) (e.g., the processing circuits of the processor 220) may determine a preliminary identification model. The preliminary identification model may include a preliminary decision tree model (e.g., a preliminary binary classification tree model), a preliminary naive Bayes model, a preliminary boosted tree model, a preliminary nearest neighbor model, a preliminary support vector machine model, etc.

[0108] In 740, the processing engine 112 (e.g., the model training module 440) (e.g., the processing circuits of the processor 220) may determine a plurality of sample probabilities for the plurality of training samples based on

the preliminary identification model and the feature information for each of the plurality of training samples. As described elsewhere in the present disclosure, similar to the first probability, the sample probability may refer to a probability that a term (e.g., a second TOI) includes at least one proper noun.

[0109] In some embodiments, the processing engine 112 may input the feature information of the plurality of training samples into the trained identification model. The trained identification model may output the sample probabilities. In some embodiments, the processing engine 112 may input the plurality of training samples into the trained identification model. The trained identification model may output the sample probabilities.

[0110] In 750, the processing engine 112 (e.g., the model training module 440) (e.g., the processing circuits of the processor 220) may determine whether the plurality of sample probabilities satisfy a predetermined condition. For example, the processing engine 112 may determine a loss function of the preliminary identification model and determine a value of the loss function based on the plurality of sample probabilities. Further, the processing engine 112 may determine whether the value of the loss function is less than a threshold. The threshold may be default settings of the proper noun identification system 100, or may be adjustable under different situations.

[0111] In response to a determination that the plurality of sample probabilities satisfy the predetermined condition, the processing engine 112 (e.g., the model training module 440) (e.g., the processing circuits of the processor 220) may designate the preliminary identification model as the trained identification model in 760. On the other hand, in response to a determination that the plurality of sample probabilities do not satisfy the predetermined condition, the processing engine 112 may execute process 700 to return to operation 730 to update the preliminary identification model. For example, the processing engine 112 may update one or more preliminary parameters (e.g., a weight matrix, a bias vector) of the preliminary identification model to produce an

updated identification model.

[0112] Further, the processing engine 112 may determine whether a plurality of updated sample probabilities under the updated identification model satisfy the predetermined condition. In response to a determination that the plurality of updated sample probabilities satisfy the predetermined condition, the processing engine 112 may designate the updated identification model as the trained identification model in 760. On the other hand, in response to a determination that the plurality of updated sample probabilities still does not satisfy the predetermined condition, the processing engine 112 may still execute process 700 to return to 730 to update the updated identification model until the plurality of updated sample probabilities satisfy the predetermined condition.

[0113] It should be noted that the above description is provided for the purpose of illustration, and is not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations and modifications may be made under the teaching of the present disclosure. However, those variations and modifications do not depart from the scope of the present disclosure.

[0114] FIG. 8 is a flowchart illustrating an exemplary process for identifying at least one proper noun included in a target query according to some embodiments of the present disclosure. In some embodiments, process 800 may be implemented as a set of instructions (e.g., an application) stored in the storage ROM 230 or RAM 240. The processor 220 and/or the modules in FIG. 4 may execute the set of instructions, and when executing the instructions, the processor 220 and/or the modules may be configured to perform process 800. The operations of the illustrated process presented below are intended to be illustrative. In some embodiments, process 800 may be accomplished with one or more additional operations not described and/or without one or more of the operations herein discussed. Additionally,

the order in which the operations of the process as illustrated in FIG. 8 and described below is not intended to be limiting.

[0115] In 810, the processing engine 112 (e.g., the proper noun identification module 450) (e.g., the processing circuits of the processor 220) may obtain a search request including a target query. As described in connection with 510, the search request may refer to a search request for searching a target query. The target query may refer to a term that the user intends to search. Merely by way of example, types of the target query may include a term (e.g., an address) associated with a transportation search request or a map (e.g., GOOGLE Map, BAIDU Map, TENCENT Map) navigation search request, a term (e.g., goods) associated with an online shopping search request, a term (e.g., food) associated with a meal ordering search request, etc.

[0116] In 820, the processing engine 112 (e.g., the proper noun identification module 450) (e.g., the processing circuits of the processor 220) may determine whether the target query includes at least one proper noun based on a trained identification model. As described in connection with FIG. 7, the trained identification model may be configured to provide a probability that a term (e.g., the target query) includes at least one proper noun. The trained identification model may be trained based on a plurality of second historical search records and a preliminary identification model. More detailed descriptions can be found in FIG. 7 and not be repeated here.

[0117] In 830, the processing engine 112 (e.g., the proper noun identification module 450) (e.g., the processing circuits of the processor 220) may determine that the target query includes at least one proper noun in response to a determination that the probability is greater than the predetermined probability threshold. In some embodiments, the processing engine 112 may compare the probability with a predetermined probability threshold. In response to a determination that the probability is greater than the predetermined probability threshold, the processing engine 112 may

determine that the target query includes at least one proper noun. As used herein, the predetermined probability threshold may be default settings of the proper noun identification system 100, or may be adjustable under different situations.

[0118] It should be noted that the above description is provided for the purpose of illustration, and is not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations and modifications may be made under the teaching of the present disclosure. However, those variations and modifications do not depart from the scope of the present disclosure.

[0119] FIG. 9 is a flowchart illustrating an exemplary process for ranking TOIs based on a proper noun list including a plurality of proper nouns according to some embodiments of the present disclosure. In some embodiments, process 900 may be implemented as a set of instructions (e.g., an application) stored in the storage ROM 230 or RAM 240. The processor 220 and/or the modules in FIG. 4 may execute the set of instructions, and when executing the instructions, the processor 220 and/or the modules may be configured to perform process 900. The operations of the illustrated process presented below are intended to be illustrative. In some embodiments, process 900 may be accomplished with one or more additional operations not described and/or without one or more of the operations herein discussed. Additionally, the order in which the operations of the process as illustrated in FIG. 9 and described below is not intended to be limiting.

[0120] In 910, the processing engine 112 (e.g., the TOI determination module 460) (e.g., the processing circuits of the processor 220) may obtain a search request including a target query. As described in connection with 510 or 810, the search request may refer to a search request for searching a target query. The target query may refer to a term that the user intends to search. Merely by way of example, types of the target query may include a term (e.g., an

address) associated with a transportation search request or a map (e.g., GOOGLE Map, BAIDU Map, TENCENT Map) navigation search request, a term (e.g., goods) associated with an online shopping search request, a term (e.g., food) associated with a meal ordering search request, etc.

[0121] In 920, the processing engine 112 (e.g., the TOI determination module 460) (e.g., the processing circuits of the processor 220) may determine one or more terms of interest (TOIs) associated with the target query. As used herein, the TOI may refer to a term associated with a target query that the user may be interested in.

[0122] In some embodiments, the processing engine 112 may first determine at least one of a prefix, a key term, or a phrase in the target query and determine the one or more TOIs based on the prefix, the key term, or the phrase.

[0123] In some embodiments, the processing engine 112 may determine correlation coefficients for the one or more TOIs. As used herein, the correlation coefficient may indicate a similarity between the target query and the TOI. The larger the correlation coefficient is, the higher the similarity between the TOI and the target query may be, and the greater a probability that a user selects the TOI as a search term associated with the target query may be.

[0124] In 930, the processing engine 112 (e.g., the TOI determination module 460) (e.g., the processing circuits of the processor 220) may rank at least a portion of the one or more TOIs based on a proper noun list including a plurality of proper nouns. In some embodiments, the processing engine 112 may compare the target query with proper nouns on the proper noun list. In response to a determination that the target query includes at least one proper noun on the proper noun list, the processing engine 112 may rank TOIs that include the at least one proper noun in front. In some embodiments, the processing engine 112 may allocate at least one weighting coefficient to the at

least one proper noun. The processing engine 112 may rank the at least one of the one or more TOIs based on the at least one weighting coefficient and the correlation coefficients for the TOIs. In response to a determination that the target query does not include at least one proper noun on the proper noun list, the processing engine 112 may rank the at least a portion of the one or more TOIs based on the correlation coefficients for the TOIs (e.g., from large to small).

[0125] It should be noted that the above description is provided for the purpose of illustration, and is not intended to limit the scope of the present disclosure. For persons having ordinary skills in the art, multiple variations and modifications may be made under the teaching of the present disclosure. However, those variations and modifications do not depart from the scope of the present disclosure. For example, the processing engine 112 may update the trained identification model at a certain time interval (e.g., per month, per two months) based on a plurality of newly obtained second historical search records.

[0126] Having thus described the basic concepts, it may be rather apparent to those skilled in the art after reading this detailed disclosure that the foregoing detailed disclosure is intended to be presented by way of example only and is not limiting. Various alterations, improvements, and modifications may occur and are intended to those skilled in the art, though not expressly stated herein. These alterations, improvements, and modifications are intended to be suggested by this disclosure, and are within the spirit and scope of the exemplary embodiments of this disclosure.

[0127] Moreover, certain terminology has been used to describe embodiments of the present disclosure. For example, the terms “one embodiment,” “an embodiment,” and/or “some embodiments” mean that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present disclosure.

Therefore, it is emphasized and should be appreciated that two or more references to “an embodiment” or “one embodiment” or “an alternative embodiment” in various portions of this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures or characteristics may be combined as suitable in one or more embodiments of the present disclosure.

[0128] Further, it will be appreciated by one skilled in the art, aspects of the present disclosure may be illustrated and described herein in any of a number of patentable classes or context including any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof. Accordingly, aspects of the present disclosure may be implemented entirely hardware, entirely software (including firmware, resident software, micro-code, etc.) or combining software and hardware implementation that may all generally be referred to herein as a “unit,” “module,” or “system.” Furthermore, aspects of the present disclosure may take the form of a computer program product embodied in one or more computer readable media having computer readable program code embodied thereon.

[0129] A computer readable signal medium may include a propagated data signal with computer readable program code embodied therein, for example, in baseband or as part of a carrier wave. Such a propagated signal may take any of a variety of forms, including electro-magnetic, optical, or the like, or any suitable combination thereof. A computer readable signal medium may be any computer readable medium that is not a computer readable storage medium and that may communicate, propagate, or transport a program for use by or in connection with an instruction execution system, apparatus, or device. Program code embodied on a computer readable signal medium may be transmitted using any appropriate medium, including wireless, wireline, optical fiber cable, RF, or the like, or any suitable combination of the

foregoing.

[0130] Computer program code for carrying out operations for aspects of the present disclosure may be written in any combination of one or more programming languages, including an object oriented programming language such as Java, Scala, Smalltalk, Eiffel, JADE, Emerald, C++, C#, VB. NET, Python or the like, conventional procedural programming languages, such as the "C" programming language, Visual Basic, Fortran 2003, Perl, COBOL 2002, PHP, ABAP, dynamic programming languages such as Python, Ruby and Groovy, or other programming languages. The program code may execute entirely on the user's computer, partly on the user's computer, as a stand-alone software package, partly on the user's computer and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through any type of network, including a local area network (LAN) or a wide area network (WAN), or the connection may be made to an external computer (for example, through the Internet using an Internet Service Provider) or in a cloud computing environment or offered as a service such as a Software as a Service (SaaS).

[0131] Furthermore, the recited order of processing elements or sequences, or the use of numbers, letters, or other designations therefore, is not intended to limit the claimed processes and methods to any order except as may be specified in the claims. Although the above disclosure discusses through various examples what is currently considered to be a variety of useful embodiments of the disclosure, it is to be understood that such detail is solely for that purpose, and that the appended claims are not limited to the disclosed embodiments, but, on the contrary, are intended to cover modifications and equivalent arrangements that are within the spirit and scope of the disclosed embodiments. For example, although the implementation of various components described above may be embodied in a hardware device, it may

also be implemented as a software only solution, e.g., an installation on an existing server or mobile device.

[0132] Similarly, it should be appreciated that in the foregoing description of embodiments of the present disclosure, various features are sometimes grouped together in a single embodiment, figure, or description thereof for the purpose of streamlining the disclosure aiding in the understanding of one or more of the various embodiments. This method of disclosure, however, is not to be interpreted as reflecting an intention that the claimed subject matter requires more features than are expressly recited in each claim. Rather, claimed subject matter may lie in less than all features of a single foregoing disclosed embodiment.

WE CLAIM:

1. A system for segmenting target queries, comprising:
 - at least one storage medium including a set of instructions; and
 - at least one processor in communication with the at least one storage medium, wherein when executing the set of instructions, the at least one processor is directed to:
 - obtain a search request including a target query; and
 - segment the target query at least based on a proper noun list including a plurality of proper nouns, wherein the proper noun list is provided by:
 - obtaining a plurality of first historical search records, wherein each of the plurality of first historical search records includes a first query from a first user, and a first Term of Interest (TOI) selected by the first user as a first search term of the first historical search record, and at least a portion of the first TOIs include at least one proper noun;
 - obtaining a trained identification model, wherein the trained identification model is configured to provide a probability that the first TOI includes at least one proper noun; and
 - determining the proper noun list at least based on the trained identification model and the plurality of first historical search records.
2. The system of claim 1, wherein to segment the target query at least based on a proper noun list including a plurality of proper nouns, the at least one processor is directed to:
 - segment the target query into a plurality of segments using a segmenting technique, wherein the segmenting technique includes an N-gram technique;
 - compare the plurality of segments with the proper noun list; and
 - combine at least two segments of the plurality of segments in response to

a determination that the at least two combined segments is a proper noun on the proper noun list.

3. The system of any one of claims 1-2, wherein to segment the target query at least based on a proper noun list including a plurality of proper nouns, the at least one processor is directed to:

- obtain a segmentation model, wherein the segmentation model is configured to segment the target query;

- merge the proper noun list into the segmentation model by allocating a weighting coefficient to the proper noun list; and

- segment the target query based on the merged segmentation model.

4. The system of any one of claims 1-3, wherein the determining the proper noun list at least based on the trained identification model and the plurality of first historical search records includes:

- for each of the plurality of first historical search records;

- determining a first probability that the first TOI includes at least one proper noun; and

- determining whether the first probability is greater than a first predetermined probability threshold; and

- adding the proper nouns into the proper noun list if the first probability is greater than the first predetermined probability threshold.

5. The system any one of claims 1-4, wherein the trained identification model is determined with a training process, the training process comprising:

- obtaining a plurality of second historical search records, wherein each of the plurality of second historical search records includes a second query from a second user, and a second TOI selected by the second user as a second search term of the second historical search record, and at least a portion of

the second TOIs include at least one proper noun;

obtaining a plurality of training samples based on the plurality of second historical search records; and

training a preliminary identification model based on the plurality of training samples to generate the trained identification model.

6. The system of claim 5, wherein the obtaining a plurality of training samples based on the plurality of second historical search records includes:

for each of the plurality of second historical search records,

segmenting the second TOI into a plurality of second segments in the second historical search record;

extracting feature information associated with the plurality of second segments;

determining whether the feature information satisfies a first predetermined condition; and

designating the second historical search record as a candidate search record if the feature information satisfies the first predetermined condition; and

determining the plurality of training samples based on the candidate search records.

7. The system of claim 6, wherein the feature information of each of the plurality of second segments includes at least one of a first consistent parameter associated with the second query and the second segment, a second consistent parameter associated with the second TOI and the second segment, a cohesive parameter associated with the second segment, a left entropy associated with the second segment, a right entropy associated with the second segment, a degree of freedom associated with the second segment, a probability that the second segment is at a start of the second TOI

or an end of the second TOI, and a frequency associated the second segment.

8. The system of claim 6, further comprising:

manually filtering out at least one candidate search record among the candidate search records; and

determining the plurality of training samples based on the filtered candidate search records.

9. The system of any of claims 6-8, wherein the training a preliminary identification model based on the plurality of training samples to generate the trained identification model includes:

determining a plurality of sample probabilities corresponding to the plurality of training samples based on the preliminary identification model and the feature information for each of the plurality of training samples;

determining whether the plurality of sample probabilities satisfy a second predetermined condition; and

designating the preliminary identification model as the trained identification model in response to the determination that the plurality of sample probabilities satisfy the second predetermined condition.

10. The system of any one of claims 1-9, wherein the trained identification model includes a support vector machine model.

11. A system for identifying proper noun included in target queries, comprising:

at least one storage medium including a set of instructions; and

at least one processor in communication with the at least one storage medium, wherein when executing the set of instructions, the at least one

processor is directed to:

- obtain a search request including a target query; and
- determine whether the target query includes at least one proper noun based on a trained identification model, wherein the trained identification model is configured to provide a probability that the target query includes at least one proper noun; and
- in response to a determination that the probability is greater than a predetermined probability threshold, determine that the target query includes at least one proper noun.

12. The system of claim 11, wherein the trained identification model is determined with a training process, the training process comprising:

- obtaining a plurality of first historical search records, wherein each of the plurality of first historical search records includes a first query from a first user, and a first TOI selected by the first user as a first search term of the first historical search record, and at least a portion of the first TOIs includes at least one proper noun;
- obtaining a plurality of training samples based on the plurality of first historical search records; and
- training a preliminary identification model based on the plurality of training samples to generate the trained identification model.

13. The system of claim 12, wherein the obtaining a plurality of training samples based on the plurality of first historical search records includes:

- for each of the plurality of first historical search records,
- segmenting the first TOI into a plurality of segments in the first historical search record;
- extracting feature information associated with the plurality of segments;

determining whether the feature information satisfies a first predetermined condition; and

designating the first historical search records as a candidate search record if the feature information satisfies a first predetermined condition; and

determining the plurality of training samples based on the candidate search records.

14. The system of claim 13, wherein the feature information of each of the plurality of segments includes at least one of a first consistent parameter associated with the first query and the segment, a second consistent parameter associated with the first TOI and the segment, a cohesive parameter associated with the segment, a left entropy associated with the segment, a right entropy associated with the segment, a degree of freedom associated with the segment, a probability that the segment is at a start of the first TOI or an end of the first TOI, or a frequency associated the segment.

15. The system of claim 13, further comprising:

manually filtering out at least one candidate search record among the candidate search records; and

determining the plurality of training samples based on the filtered candidate search records.

16. The system of any of claims 13-15, wherein the training a preliminary identification model based on the plurality of training samples to generate the trained identification model includes:

determining a plurality of sample probabilities corresponding to the plurality of training samples based on the preliminary identification model and the feature information for each of the plurality of training samples;

determining whether the plurality of sample probabilities satisfy a second predetermined first condition; and

designating the preliminary identification model as the trained identification model in response to the determination that the plurality of sample probabilities satisfy the second predetermined condition.

17. The system of any one of claims 11-16, wherein the trained identification model includes a support vector machine model.

18. The system of any one of claims 11-17, wherein the at least one processor is directed further to:

segment the target query based on the at least one proper noun in response to the determination that the probability is greater than a predetermined probability threshold.

19. A system for segmenting target queries, comprising:

at least one storage medium including a set of instructions; and

at least one processor in communication with the at least one storage medium, wherein when executing the set of instructions, the at least one processor is directed to:

obtain a search query including a target query;

determine one or more terms of interest (TOIs) associated with the target query; and

rank at least a portion of the one or more TOIs based on a proper noun list including a plurality of proper nouns.

20. A method implemented on a computing device having at least one processor, at least one storage medium, and a communication platform connected to a network, the method comprising:

- obtaining a search request including a target query; and

- segmenting the target query at least based on a proper noun list including a plurality of proper nouns, wherein the proper noun list is provided by:

 - obtaining a plurality of first historical search records, wherein each of the plurality of first historical search records includes a first query from a first user, and a first TOI selected by the user as a first search term of the first historical search record, and at least a portion of the first TOIs include at least one proper noun;

 - obtaining a trained identification model, wherein the trained identification model is configured to provide a probability that the first TOI includes at least one proper noun; and

 - determining the proper noun list at least based on the trained identification model and the plurality of first historical search records.

21. The method of claim 20, wherein the segmenting the target query at least based on a proper noun list including a plurality of proper nouns includes:

- segmenting the target query into a plurality of segments based on a segmenting technique, wherein the segmenting technique includes an N-gram technique;

- comparing the plurality of segments with the proper noun list; and

- combining at least two segments in response to a determination that the at least two combined segments is a proper noun on the proper noun list.

22. The method of any one of claims 20-21, wherein the segmenting the target query at least based on a proper noun list including a plurality of proper

nouns includes:

- obtaining a segmentation model, wherein the segmentation model is configured to segment the target query;

- merging the proper noun list into the segmentation model by allocating a weighting coefficient to the proper noun list; and

- segmenting the target query based on the merged segmentation model.

23. The method of any one of claims 20-22, wherein the determining the proper noun list at least based on the trained identification model and the plurality of first historical search records includes:

- for each of the plurality of first historical search records;

- determining a first probability that the first TOI includes at least one proper noun; and

- determining whether the first probability is greater than a first predetermined probability threshold; and

- adding the proper nouns into the proper noun list if the first probability is greater than the first predetermined probability threshold.

24. The method of any one of claims 20-23, wherein the trained identification model is determined with a training process, the training process comprising:

- obtaining a plurality of second historical search records, wherein each of the plurality of second historical search records includes a second query from a second user, and a second TOI selected by the user as a second search term of the second historical search record, and at least a portion of the second TOIs include at least one proper noun;

- obtaining a plurality of training samples based on the plurality of second historical search records; and

- training a preliminary identification model based on the plurality of training

samples to generate the trained identification model.

25. The method of claim 24, wherein the obtaining a plurality of training samples based on the plurality of second historical search records includes:

- for each of the plurality of second historical search records,
 - segmenting the second TOI into a plurality of second segments in the second historical search record;
 - extracting feature information associated with the plurality of second segments;
 - determining whether the feature information satisfies a first predetermined condition; and
 - designating the second historical search record as a candidate search record if the feature information satisfies the first predetermined condition; and
 - determining the plurality of training samples based on the candidate search records.

26. The method of claim 25, wherein the feature information of each of the plurality of segments includes at least one of a first consistent parameter associated with the second query and the segment, a second consistent parameter associated with the second TOI and the segment, a cohesive parameter associated with the segment, a left entropy associated with the segment, a right entropy associated with the segmentation, a degree of freedom associated with the segment, a probability that the segment is at a start of the query or an end of the second TOI, and a frequency associated with the segment.

27. The method of claim 25, wherein the determining the plurality of training samples based on the candidate search records includes:

filtering out at least one candidate search record among the second historical search records manually; and

determining the plurality of training samples based on the filtered candidate search records.

28. The method of any of claims 25-27, wherein the training a preliminary identification model based on the plurality of training samples to generate the trained identification model includes:

determining a plurality of sample probabilities corresponding to the plurality of training samples based on the preliminary identification model and the feature information for each of the plurality of training samples;

determining whether the plurality of sample probabilities satisfy a predetermined second condition; and

designating the preliminary identification model as the trained identification model in response to the determination that the plurality of sample probabilities satisfy the predetermined condition.

29. The method of any one of claims 20-28, wherein the trained identification model includes a support vector machine model.

30. A method implemented on a computing device having at least one processor, at least one storage medium, and a communication platform connected to a network, the method comprising:

obtaining a search request including a target query; and

determining whether the target query includes at least one proper noun based on a trained identification model, wherein the trained identification model is configured to provide a probability that the target query includes at least one proper noun; and

in response to a determination that the probability is greater than a

predetermined probability threshold, determining that the target query includes at least one proper noun.

31. The method of claim 30, wherein the trained identification model is determined with a training process, the training process comprising:

- obtaining a plurality of first historical search records, wherein each of the plurality of first historical search records includes a first query from a first user, and a first TOI selected by the user as a first search term of the first historical search record, and at least a portion of the first TOIs includes at least one first proper noun;

- obtaining a plurality of training samples based on the plurality of first historical search records; and

- training a preliminary identification model based on the plurality of training samples to generate the trained identification model.

32. The method of claim 31, wherein the obtaining a plurality of training samples based on the plurality of first historical search records includes:

- for each of the plurality of first historical search records,

- segmenting the first TOI into a plurality of segments in the first historical search record;

- extracting feature information associated with the plurality of segments;

- determining whether the feature information satisfies a first predetermined condition; and

- designating the first historical search records as a candidate search record if the feature information satisfies a first predetermined condition; and

- determining the plurality of training samples based on the candidate search records.

33. The method of claim 32, wherein the feature information of each of the plurality of segments includes at least one of a first consistent parameter associated with the first query and the segment, a second consistent parameter associated with the first TOI and the segment, a cohesive parameter associated with the segment, a left entropy associated with the segment, a right entropy associated with the segment, a degree of freedom associated with the segment, a probability that the segment is at a start of the query or an end of the first TOI, or a frequency associated the segment.

34. The method of claim 32, wherein the determining the plurality of training samples based on the candidate search records includes:

- filtering out at least one candidate search record among the first historical search records manually; and

- determining the plurality of training samples based on the filtered candidate search records.

35. The method of any of claims 32-34, wherein the training a preliminary identification model based on the plurality of training samples to generate the trained identification model includes:

- determining a plurality of sample probabilities corresponding to the plurality of training samples based on the preliminary identification model and the feature information for each of the plurality of training samples;

- determining whether the plurality of sample probabilities satisfy a predetermined first condition; and

- designating the preliminary identification model as the trained identification model in response to the determination that the plurality of sample probabilities satisfy the predetermined condition.

36. The method of any one of claims 30-35, wherein the trained identification model includes a support vector machine model.

37. The method of any one of claims 30-36, further comprising:

segmenting the target query based on the at least one proper noun in response to the determination that the probability is greater than a predetermined probability threshold.

38. A method implemented on a computing device having at least one processor, at least one storage medium, and a communication platform connected to a network, the method comprising:

obtaining a search request including a target query;

determining one or more terms of interest (TOIs) associated with the target query; and

ranking at least a portion of the one or more TOIs based on a proper noun list including a plurality of proper nouns.

39. A non-transitory computer readable medium, comprising executable instructions that, when executed by at least one processor, directs the at least one processor to perform a method, the method comprising:

obtaining a search request including a target query; and

segmenting the target query at least based on a proper noun list including a plurality of proper nouns, wherein the proper noun list is provided by:

obtaining a plurality of first historical search records, wherein each of the plurality of first historical search records includes a first query from a first user, and a first TOI selected by the user as a first search term of the first historical search record, and at least a portion of the first TOIs include at least one proper noun;

obtaining a trained identification model, wherein the trained

identification model is configured to provide a probability that the first TOI includes at least one proper noun; and

determining the proper noun list at least based on the trained identification model and the plurality of first historical search records.

40. A non-transitory computer readable medium, comprising executable instructions that, when executed by at least one processor, directs the at least one processor to perform a method, the method comprising:

obtaining a search request including a target query; and

determining whether the target query includes at least one proper noun based on a trained identification model, wherein the trained identification model is configured to provide a probability that the target query includes at least one proper noun; and

in response to a determination that the probability is greater than a predetermined probability threshold, determining that the target query includes at least one proper noun.

41. A non-transitory computer readable medium, comprising executable instructions that, when executed by at least one processor, directs the at least one processor to perform a method, the method comprising:

obtaining a search query including a target query;

determining one or more terms of interest (TOIs) associated with the target query; and

ranking at least a portion of the one or more TOIs based on a proper noun list including a plurality of proper nouns.

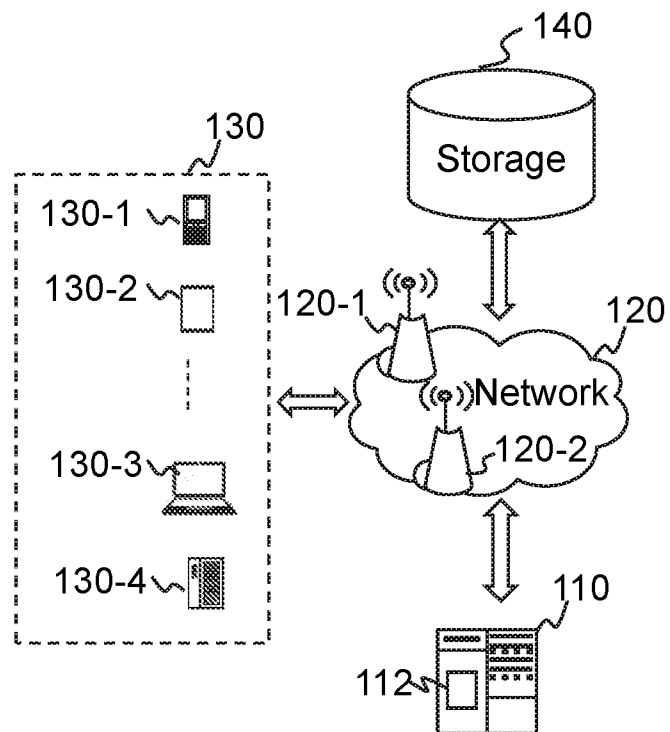
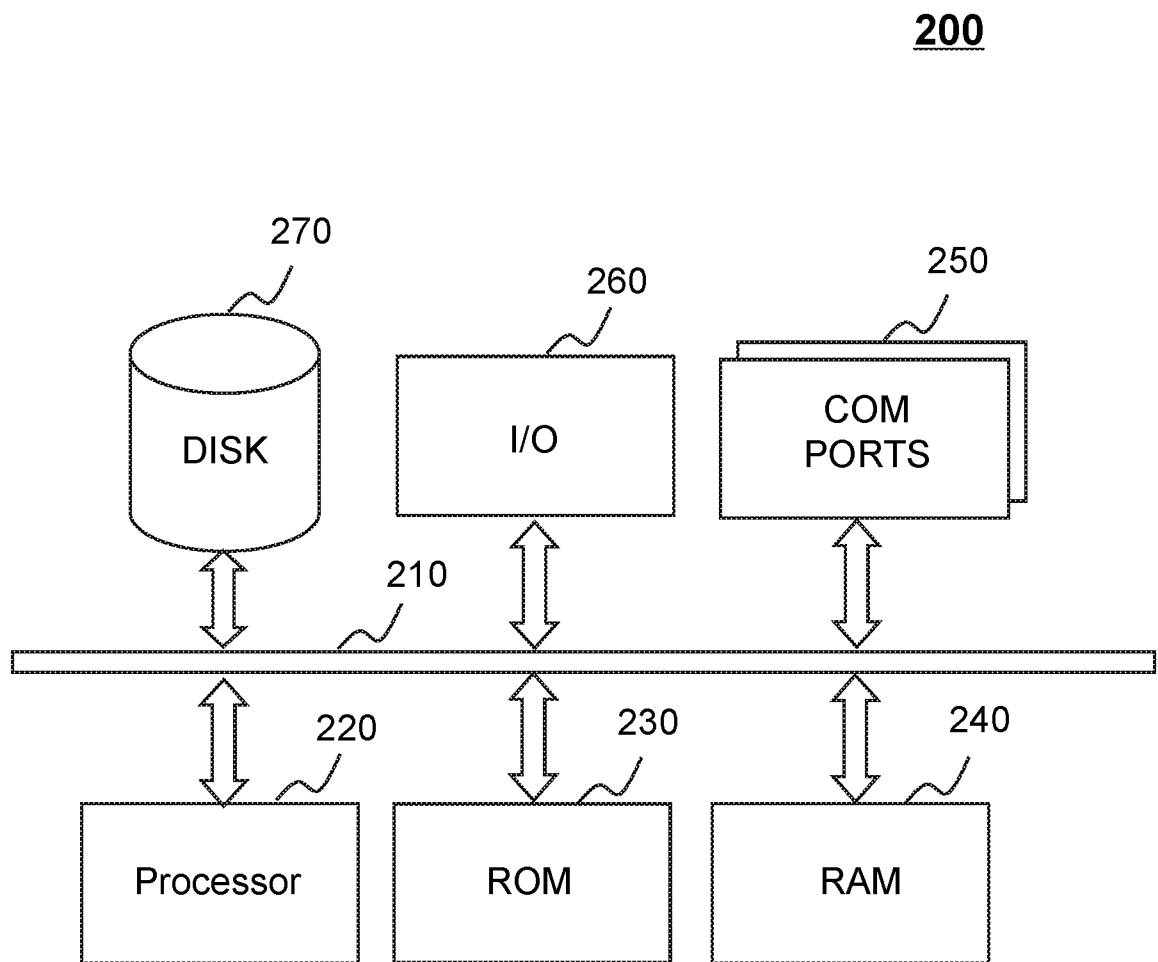
100

FIG. 1

**FIG. 2**

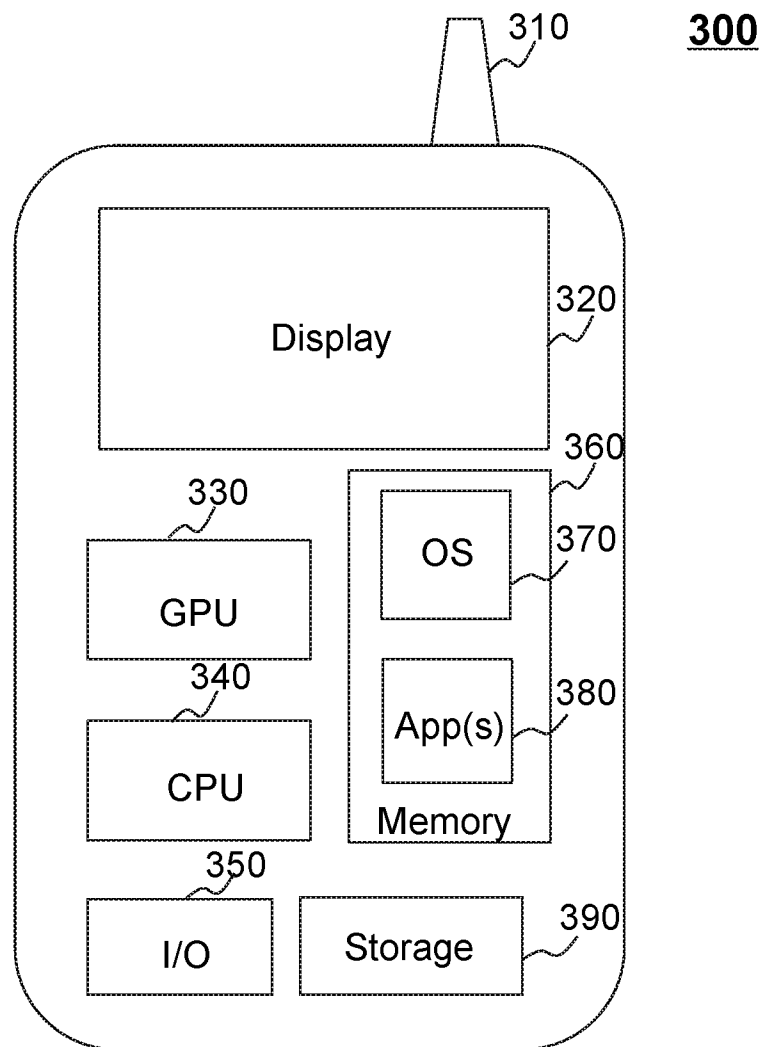
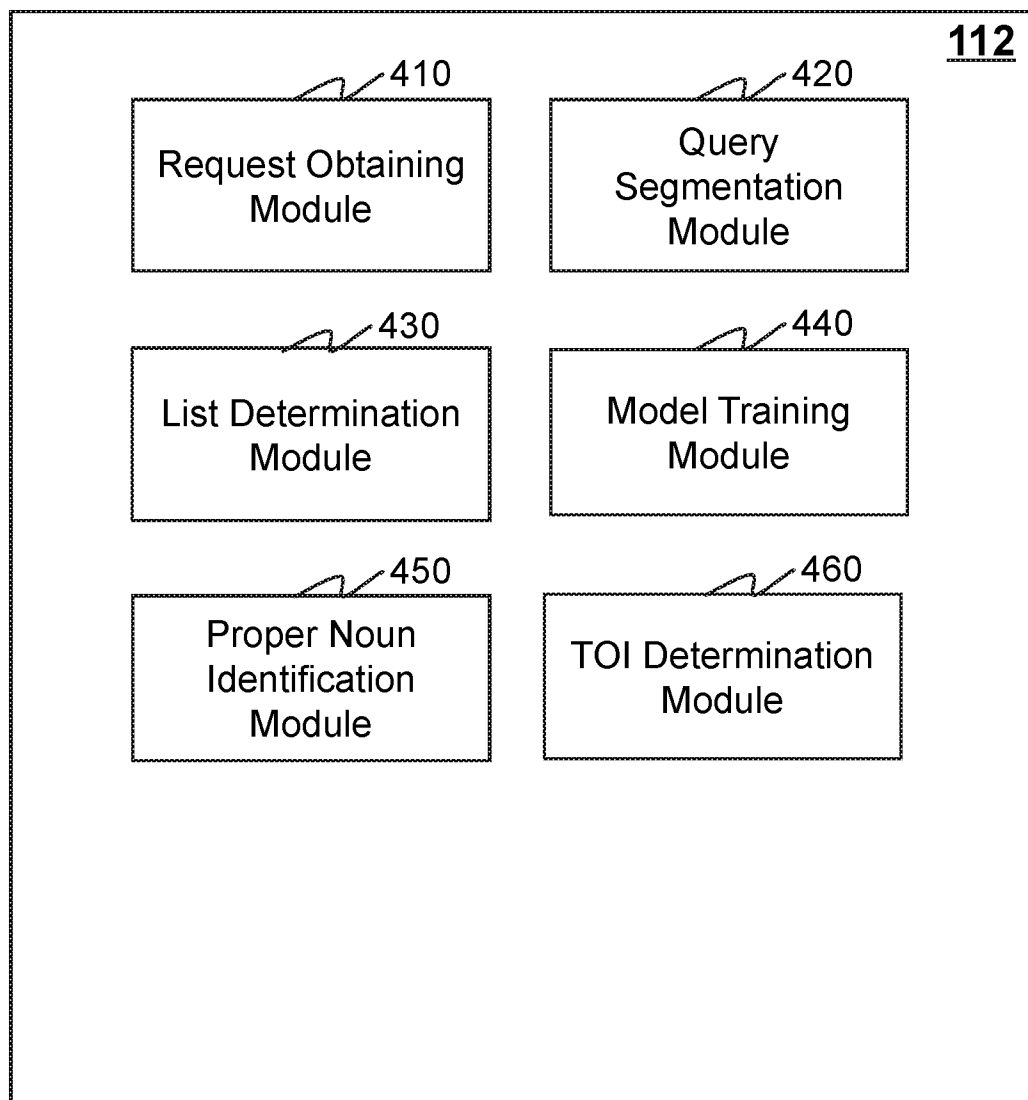
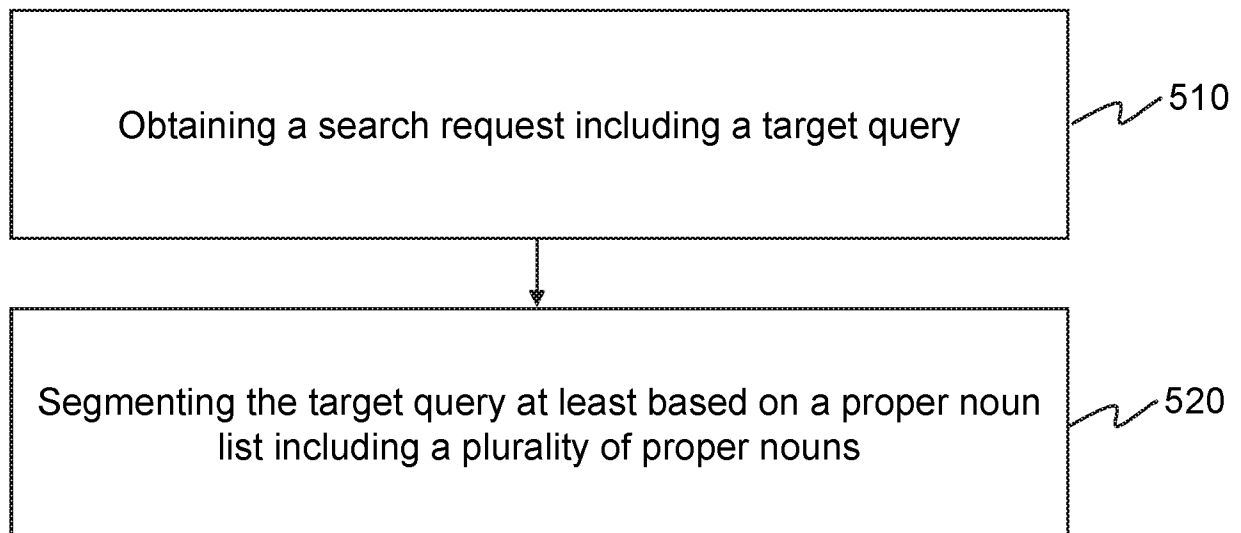
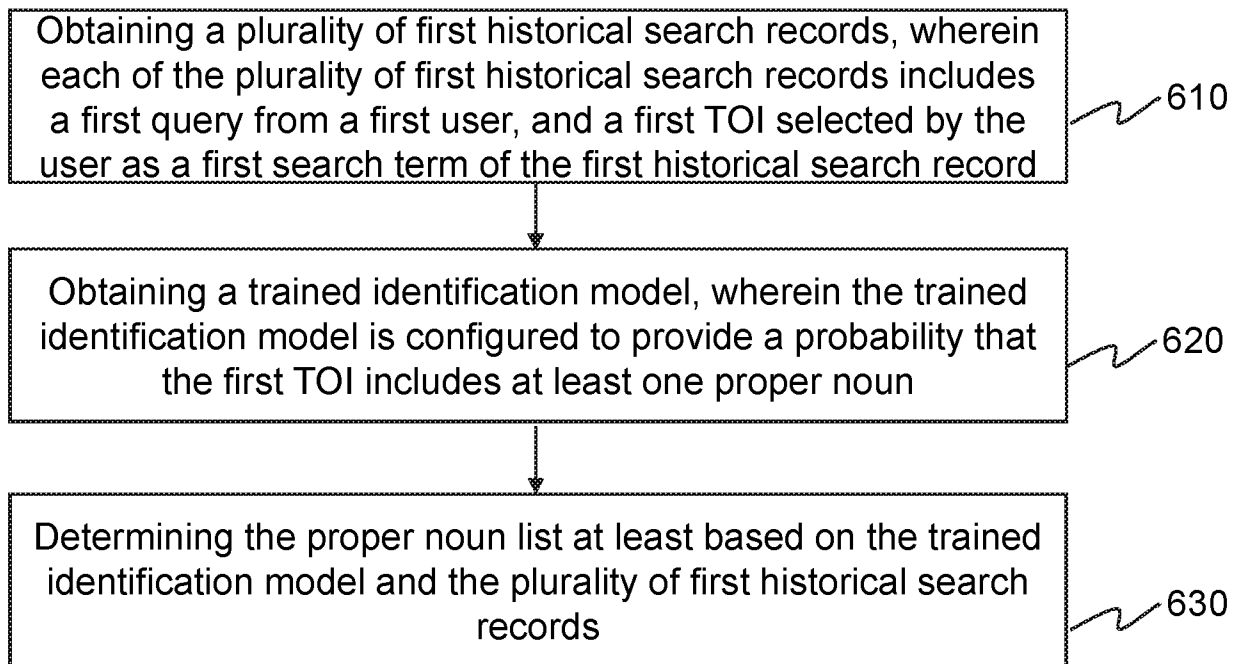
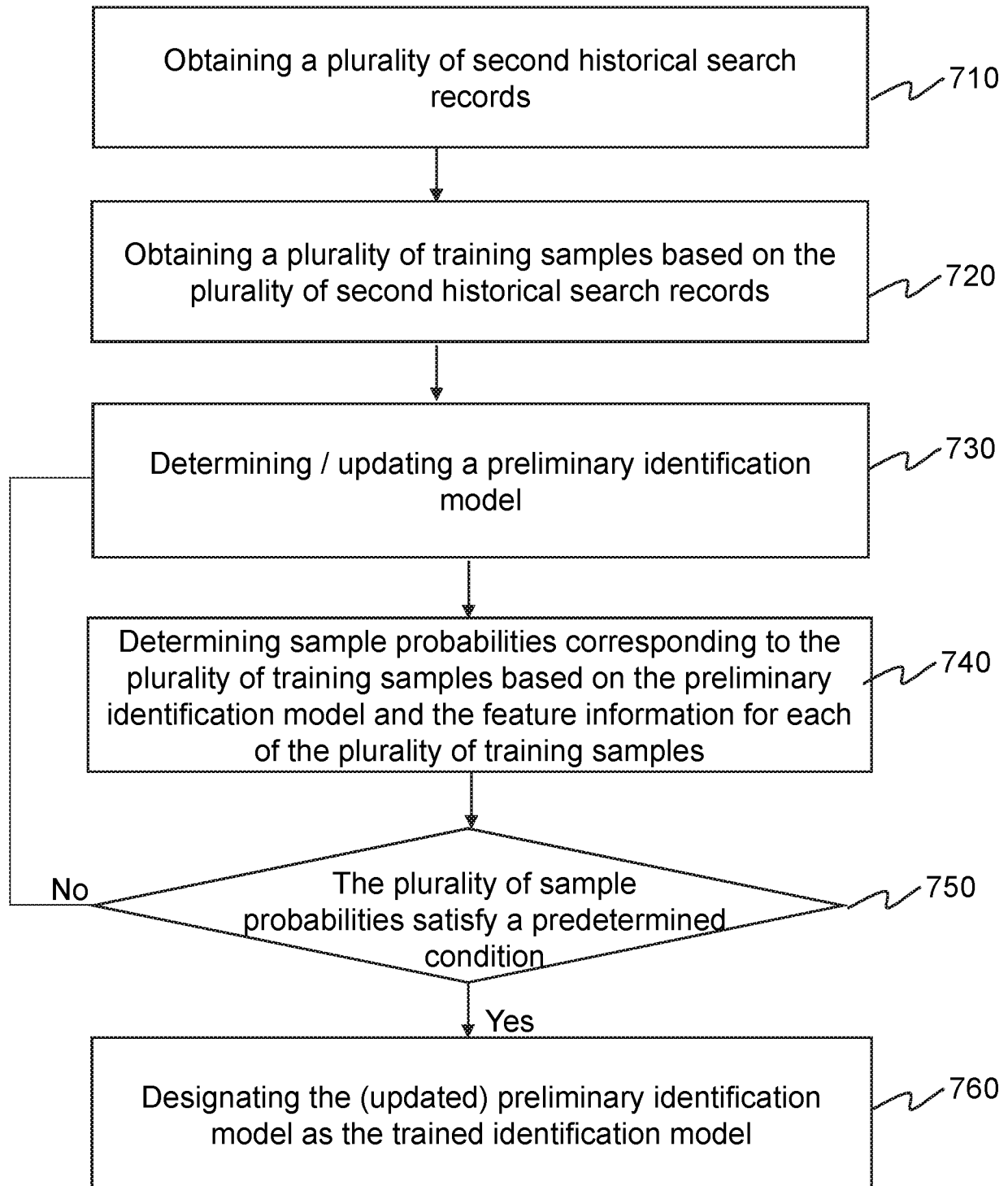


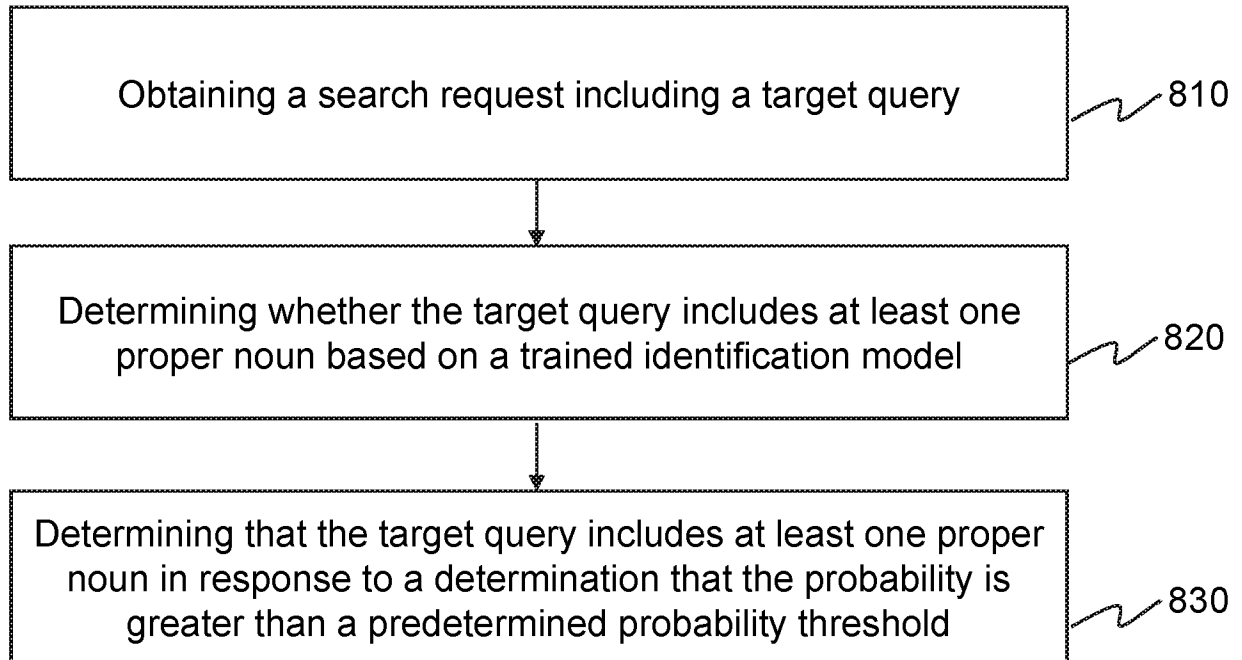
FIG. 3

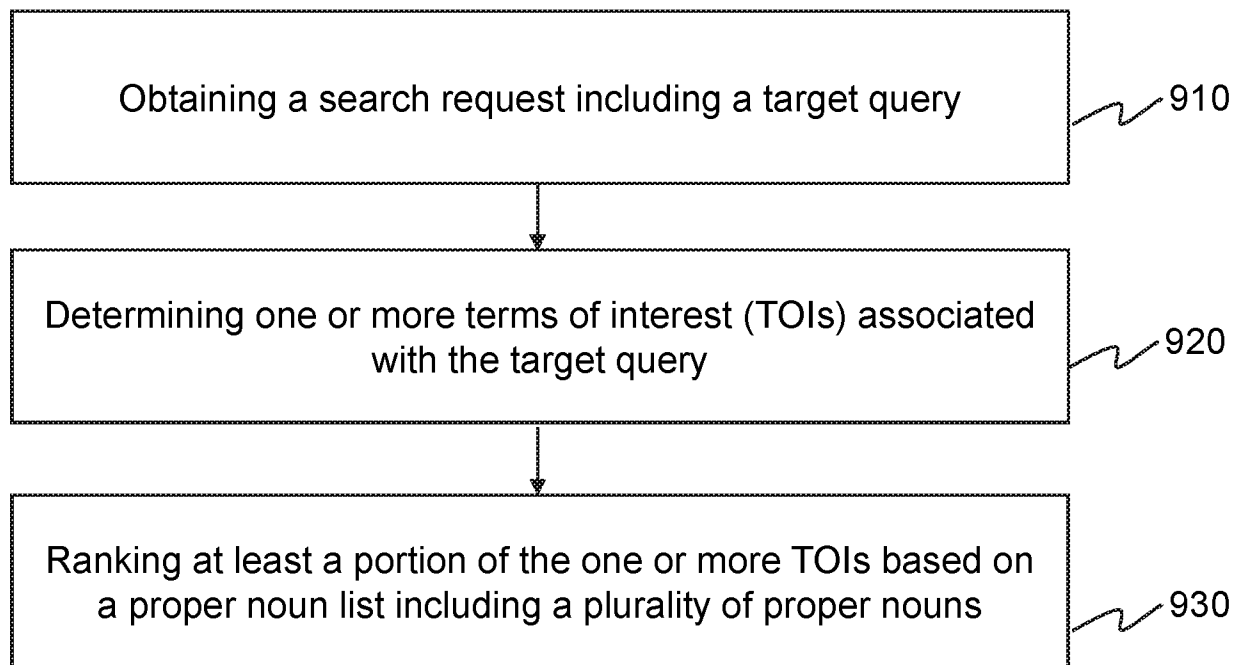
**FIG. 4**

500**FIG. 5**

600**FIG. 6**

700**FIG. 7**

800**FIG. 8**

900**FIG. 9**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/083723

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/27(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT,CNKI,WPI,EPODOC,IEEE: Named Entity, NER, recognition, identify, proper noun, segment, query, search, model, list, dictionary

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HU, Xueying et al. "Recognition of complex named-entities in user queries of search engine," <i>Computer Engineering and Applications</i> , Vol. 44, No. 19, 31 December 2008 (2008-12-31), pages 153-154	1, 2, 4, 5, 10-12, 17-21, 23, 24, 29-31, 36-41
A	CN 102799586 A (BEIJING BAIDU NETCOM SCI & TECHNOLOGY CO.) 28 November 2012 (2012-11-28) the whole document	1-41
A	CN 105653519 A (HE, Huixin) 08 June 2016 (2016-06-08) the whole document	1-41
A	CN 106528796 A (SUZHOU GONGXUN TECHNOLOGY CO., LTD.) 22 March 2017 (2017-03-22) the whole document	1-41
A	Liu, Hui et al. "Identification of Complex Named-Entities in Chinese Queries Using WWW," <i>Fifth International Conference on Fuzzy Systems and Knowledge Discovery</i> , 31 December 2008 (2008-12-31), pages 180-185	1-41



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

20 December 2019

Date of mailing of the international search report

27 December 2019

Name and mailing address of the ISA/CN

National Intellectual Property Administration, PRC
6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing
100088
China

Authorized officer

SUN,Guohui

Facsimile No. (86-10)62019451

Telephone No. 86-(10)-53961538

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/083723

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	102799586	A	28 November 2012	None	
CN	105653519	A	08 June 2016	None	
CN	106528796	A	22 March 2017	None	