(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 May 2020 (22.05.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/101477 A1

(51) International Patent Classification:
G06F 40/30 (2020.01) *G06F 40/279* (2020.01)
G06F 40/211 (2020.01)

(21) International Application Number:
PCT/MY2019/050092

(22) International Filing Date:
14 November 2019 (14.11.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PI 2018001919 14 November 2018 (14.11.2018) MY

(71) Applicant: **MIMOS BERHAD** [MY/MY]; Technology Park Malaysia, Kuala Lumpur, 57000 (MY).

(72) Inventors: **NOOR BATCHA, Mohamed Farid**; Mimos Berhad, Technology Park Malaysia, Kuala Lumpur, 57000 (MY). **DUC, Nghia Pham**; Mimos Berhad, Technology Park Malaysia, Kuala Lumpur, 57000 (MY). **AHAMAD, Muhammad Khairuddin**; Mimos Berhad, Technology Park Malaysia, Kuala Lumpur, 57000 (MY). **ULAGANATHAN, Thenmalar**; Mimos Berhad, Technology Park Malaysia, Kuala Lumpur, 57000 (MY).

(74) Agent: **PYPRUS SDN BHD**; Unit 1608, 16th Floor, Block A, Damansara Intan, No. 1, Jalan SS20/27, Selangor, Petaling Jaya, 47400 (MY).

(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,

(54) Title: SYSTEM AND METHOD FOR DYNAMIC ENTITY SENTIMENT ANALYSIS

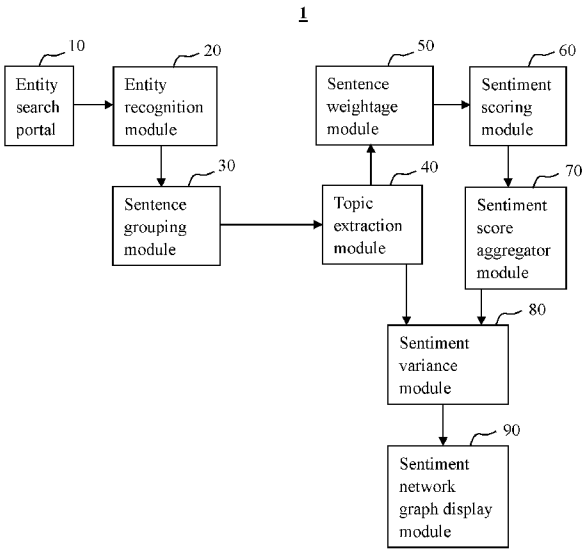

FIG 1

(57) Abstract: The present invention provides a dynamic entity sentiment analysis (DESA) system including: an entity search portal (10); an entity recognition module (20); a sentence grouping module (30); a topic extraction module (40); a sentence weightage module (50); a sentiment scoring module (60); a sentiment score aggregator module (70); a sentiment variance module (80); and a sentiment network graph display module (90). A method for DESA is also provided.

[Continued on next page]

WO 2020/101477 A1

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).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

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

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

SYSTEM AND METHOD FOR DYNAMIC ENTITY SENTIMENT ANALYSIS

Field of the Invention

[0001] The present invention generally relates to the network technologies, and more particularly to a system and method for dynamic entity sentiment analysis.

Background of the Invention

[0002] With the availability of a plethora of data over the internet, it has become difficult to look through an enormous amount of blogs, forums or social sites regarding a perception of an entity (for example, a person or an organization). It is challenging to create intelligence from these data and make sense of them effectively to visualize an overall sentiment regarding an entity due to the vast amount of views from various entities and viewpoints.

[0003] US Patent Application Publication US 2017/0039185 A1 discloses a method and a system for identifying indication for activity in a topic or a sentiment associated to, of an entity in a textual document. The method applies role based association to entities in textual documents. This reference discloses the use of the level of activity of the topic to determine a probabilistic model of the sentiment level, the use of topic occurrence to determine the level of activity of each entity in each topic, and the fine tuning to create a topic tag for each verb in text. However, it does not disclose the consideration of the author credibility in the weighting the sentiment or the alternative description of an entity.

[0004] US Patent Application Publication US 2017/0052971 A1 discloses a mechanism in a data processing system for aggregating sentiment about an entity from a corpus of documents. The mechanism determines a plurality of passage sentiment scores for the plurality of sentiment passages and an actual aggregate sentiment score from the plurality of passage sentiment scores based on a k-valued model. This reference discloses the use of a sentiment confidence score to measure the actual calculated sentiment value and the user input raw sentiment. However, it does not disclose the consideration of the author credibility in the weighting the sentiment, the topic & document grouping, the scaling of sentiment value based on any criteria, or the alternative description of an entity.

[0005] US Patent No. 8,725,494 B2 discloses a signal processing approach to sentiment analysis for entities in documents. The approach provides sentiment values for phrases in the document. The reference discloses the use of a filtering operation to the sequence of sentiment values around each entity to determine a sentiment value for the entity, and the use of signal processing mechanism to determine a weightage value for the sentiment based on surrounding sentiments. However, it does not disclose the consideration of the author credibility in the weighting the sentiment, the topic & document grouping, or the alternative description of an entity.

[0006] For hot topics, the sentiment of entitles is usually on extreme ends from both perspectives, but as an issue cools down, the general sentiment from the public regarding the pertinent entity will approach neutral over time depending on age of the topic. Therefore, there is a need for a dynamic sentiment to mimic the viewpoints of a general person.

Summary

[0007] One aspect of the present invention provides a dynamic entity sentiment analysis (DESA) system. In one embodiment, the DESA system comprises an entity search portal including a graphical user interface (GUI) to allow a user to input a keyword of an entity of interest; an entity recognition module electronically coupled with the entity search portal to receives the input keyword, and perform a query search of the keyword to find alternative matches to provide list of alternative keywords that match the user keyword; a sentence grouping module electronically coupled with the entity recognition module to use the list of keywords from the entity recognition module to do an article search, and then crawls a set of matching articles; a topic extraction module electronically coupled with the sentence grouping module to receive the blocks of sentences, and process each and every block of sentences to extract the keywords so as to provide identified topics, identified subjects and topic popularity; a sentence weightage module electronically coupled with the topic extraction module to receive the blocks of sentences and topic popularities, and extract adjectives from the input sentences to provide a weighted sentiment score of each block of sentences; a sentiment scoring module electronically coupled with the sentence weightage module to receive the blocks of sentences, weighted sentiment scores for the respective block of sentences, and perform sentiment analysis by searching for verbs and providing a

sentiment score value within a range of value; a sentiment score aggregator module electronically coupled with the sentiment scoring module to receives the sentiment scores of each block of sentences relating to the entity, and aggregate the scores from each sentiment per entity/per document/ per topic to provide list of aggregated sentiment scores of entity, document and topic; a sentiment variance module electronically coupled with the topic extraction module and sentiment score aggregator module to receive the topic, popularity, document metadata, sentiment per entity, sentiment per document, and sentiment per topic, and scales the sentiment values according to the inputs using an algorithm to provide final sentiment score on each entity based on sentences, topics and documents, and document type being extracted from the metadata; and

a sentiment network graph display module electronically coupled with the sentiment variance module to display the sentiment score values on the GUI.

[0008] In another embodiment of the DESA system, the entity recognition module comprises a related entity keyword module, a new alternative module and an entity affiliations database; when the entity recognition module receives the keyword of the entity of interest input from the entity search portal, the related entity keyword module performs a query search of the input keyword on the entity affiliation database to extract all alternative keywords, and the new alternative module continuously updates list of alternative keywords that match the input keyword in the entity affiliation database with new alternative keywords from predefined articles.

[0009] In another embodiment of the DESA system, the topic extraction module comprises a keyword extraction module with a subject matter corpus of keywords including predefined topic keywords and subcategories, a key categorization module, a topic grouping module, and a topic popularity module; where the keyword extraction module uses the subject matter corpus of keywords to extract topic related keywords from the blocks of sentences from the sentence grouping module; the key categorization module categorizes the extracted topic related keywords respectively into their subject, the topic grouping module identifies topics by using the maximum number of keywords in a subject to determine the subject and determines the subject matter of the article by further aggregation of the identified topics; and the topic popularity module determines the topic popularity using the maximum occurrence of a keyword and its affiliations over multiple articles.

[0010] In another embodiment of the DESA system, the sentence weightage module comprises an adjective extraction module, an adjective weightage module including a corpus of adjectives with predefined relevant weightage, and an adjective weightage aggregator module; wherein the adjective extraction module extracts adjectives from the block sentences; wherein for each block of sentences, together with the identified topic, the adjective weightage module scales the extracted adjectives according to the popularity of the topic to provide a sentiment score for each adjective, and wherein the adjective weightage aggregator module aggregates the sentiment scores for each block of sentences to provide a weighted sentiment score of each block sentences.

[0011] In another embodiment of the DESA system, the sentiment variance module comprises a relevancy metric module, an author credibility reporting module including a database with the authenticity of authors being predefined and updated continuously, a document metadata extraction module, and a document grouping module.

[0012] Another aspect of the present invention provides a method for dynamic entity sentiment analysis (DESA). In one embodiment, the method comprises the steps of taking in a user input of a keyword of an entity; searching based on the input keyword through a network for matching keywords to the entity, and retrieving from the network the documents related to the entity and the related keywords of the input keyword, where the list of documents is continuously updated if a new document is detected; storing the retrieved documents containing the matching keywords and storing sentiment value per entity per topic per search results; analyzing each of the retrieved documents based on its type and extracting metadata of each document such as author and timestamp; extracting other entities from each document, collecting and grouping organizations, persons or places, and matching or relating new extracted entities to similar input keywords due to acronym or spelling differences; grouping together all sentences containing matching keywords and related keywords relating to the entity (inclusive of different affiliations), regardless of sentence positions, and generating duplicate sentences if a sentence has more than one entity; for each entity, grouping the sentences adjacent to the sentences containing the keywords into blocks for keyword extraction and adjective extraction per block; sentiment per block per topic per entity, and scaling with aggregated adjective weightage block; aggregating the sentiment scores per topic basis in a document; aggregating the sentiment scores throughout multiple documents keeping the topic grouping throughout; and displaying the results of the

dynamic entity sentiment and allowing the user to breakdown the sentiment to its exact score prior to aggregating any sentiment.

[0013] In another embodiment of the method, further comprises a scaled value is associated to each adjective defined in a database; the adjectives in a block of sentence are then aggregated; the keyword extraction is performed from which the keywords are categorized to determine the topic and sub-topics; and the count of topic keywords determines the subject of the document and its popularity.

[0014] In another embodiment of the method, wherein the sentiment value is further fine-tuned to take into account the author credibility rating and topic relevancy (timestamp), and as new documents are collected, the sentiment value will scale depending on topic popularity over time.

[0015] In another embodiment, the method further comprises obtaining the blocks of grouped adjacent sentences, and extracting the adjectives from the blocks of grouped adjacent sentences to provide a list of adjectives; matching each adjective of the list of adjectives to a predefined value from a database to generate a list of adjectives with a value appended; and aggregating the values of the adjectives in a block of sentences based on the list of adjectives with a value appended to provide a single aggregated value per block of sentences.

[0016] In another embodiment, the method further comprises obtaining the blocks of grouped adjacent sentences, and extracting keywords such as proper nouns from each block of sentences and each document to provide a set of keywords grouped in terms of block of sentence and entire document; grouping the set of keywords into their respective categories according to a corpus of words to provide the categories involved in the block of sentence and throughout the document; for the categories involved in the block of sentence and entire document, counting the number of keywords belonging to each category, and determining from the corpus the main topic and sub-topics of the document, where the determined main topic and sub-topics of the document are forwarded for processing; obtaining a historical list of multiple documents, and keeping track of the popularity of topics to provide a list of popular occurring topics; passing the topic group field for the front end GUI to the aforesaid steps; and passing to the front end GUI the document type grouping, and the GUI is from the step of the aforesaid step.

[0017] In another embodiment, the method further comprises obtaining the other entities extracted from the document, and matching the user input keyword to the extracted entities to find other similar relationships to extract new alternatives to the user keywords, where the new alternatives will be added to the database; and finding from a defined database a set of alternatives of the user input keyword to generate a list of alternatives from the database and also a new set of alternatives from the aforesaid steps.

[0018] In another embodiment, the method further comprises extracting the metadata from each document to generate a list of information regarding the document such as the author and timestamp; updating the credibility of the author over time into a database based on the authors background information such as education or field of work to generate a set of ratings that are dynamically updated over time to the database; storing the author's credibility rating and information according to the subject into a database that is a predefined source and is updated with new authors; rating the author according to the timestamp of the document based on the database and the metadata of the document; and based on the ranking of the author's credibility, dynamically scaling the authors ranking according to the topic of the document and timestamp to generate a scaled authors ranking, ensuring that the authors ranking is a variable in terms of time and document topic.

[0019] The objectives and advantages of the invention will become apparent from the following detailed description of preferred embodiments thereof in connection with the accompanying drawings.

Brief Description of the Drawings

[0020] Preferred embodiments according to the present invention will now be described with reference to the Figures, in which like reference numerals denote like elements.

[0021] FIG 1 shows a block architecture of the dynamic entity sentiment analysis (DESA) system in accordance with one embodiment of the present invention.

[0022] FIG 2 shows a block architecture of the entity recognition module of the dynamic entity sentiment analysis (DESA) system in accordance with one embodiment of the present invention.

[0023] FIG 3 shows a block architecture of the topic extraction module of the dynamic entity sentiment analysis (DESA) system in accordance with one embodiment of the present invention.

[0024] FIG 4 shows a block architecture of the sentence weightage module of the dynamic entity sentiment analysis (DESA) system in accordance with one embodiment of the present invention.

[0025] FIG 5 shows a block architecture of the sentiment variance module of the dynamic entity sentiment analysis (DESA) system in accordance with one embodiment of the present invention.

[0026] FIG 6 shows one exemplary format of a network graph for displaying the sentiment score values.

[0027] FIG 7 shows an exemplary scenario of zooming into various groupings to check the sentiment score values.

[0028] FIG 8 shows a flow chart of overall process workflow for the method for dynamic entity sentiment analysis in accordance with one embodiment of the present invention.

[0029] FIG 9 shows a flow chart of the main process flow of the method of DESA in accordance with one embodiment of the present invention.

[0030] FIG 10 shows a flow chart of the sentence weightage flow of the method of DESA in accordance with one embodiment of the present invention.

[0031] FIG 11 shows a flow chart of topic extraction and grouping flows of the method of DESA in accordance with one embodiment of the present invention.

[0032] FIG 12 shows a flow chart of the entity recognition flow of the method of DESA in accordance with one embodiment of the present invention.

[0033] FIG 13 shows a flow chart of the sentiment variance flow of the method of DESA in accordance with one embodiment of the present invention.

Detailed Description of the Invention

[0034] The present invention may be understood more readily by reference to the following detailed description of certain embodiments of the invention.

[0035] Throughout this application, where publications are referenced, the disclosures of these publications are hereby incorporated by reference, in their entireties, into this application in order to more fully describe the state of art to which this invention pertains.

[0036] The present invention provides a dynamic entity sentiment analysis system for assisting a user to visualize the sentiments regarding an entity. In the dynamic entity sentiment analysis system, topics are grouped to help the user view different sentiments regarding each topic, and the degree of topics can be zoomed in to view the sub-topic categories. The degree of control can be zoomed to point to a specific document.

[0037] The sentiment values are dynamic depending on popularity of topic, authenticity of author, and timestamp of articles.

[0038] The main objective of the invention is to effectively evaluate the sentiment of an entity by doing a search of the entity on the internet. The search would continuously retrieve blogs, Facebook post, tweets, forums, news portals and other relevant sites. The information retrieved will be categorized into the relevant topics and subtopics. By searching a single entity, all the relevant keywords of the entity are also searched which are stored in a database of similar entities reference. This will include titles / salutation or other names of the entity. This database of similar entity reference will be updated if new titles / salutations are obtained. Updating will either require user input, or automated with new alternative found in authenticated portals.

[0039] The popularity of the topic would be measured according to number of counts of the topic search. This popularity of a topic and the topic itself is taken into account when providing a sentiment value of the entity regarding the topic.

[0040] The authenticity of an author of the statement (blogs, forums...) would also be taken into account in measuring the sentiment value. Sentiments from official sites such as news forums would hold more value than blogs where the blogs activity is low. The sentiments from news portal would also be grouped, while those of blogs would be grouped. This allows the user to view the sentiment from each group.

[0041] The final outcome would be an aggregated value for the sentiment of an entity. The aggregated value will be able to expand to show the breakdown of the sentiments to each groupings until to the level of the specific blog/forum/portal/social media.

[0042] Referring now to FIG 1, there is provided a block architecture of the dynamic entity sentiment analysis (DESA) system in accordance with one embodiment of the present

invention. The DESA system **1** comprises an entity search portal **10**, an entity recognition module **20**, a sentence grouping module **30**, a topic extraction module **40**, a sentence weightage module **50**, a sentiment scoring module **60**, a sentiment score aggregator module **70**, a sentiment variance module **80**, and a sentiment network graph display module **90**.

[0043] The entity search portal **10** includes a graphical user interface (GUI) to allow a user to input a keyword of an entity of interest.

[0044] The entity recognition module **20** electronically coupled with the entity search portal **10** receives the input keyword, and performs a query search of the keyword to find alternative matches to provide list of alternative keywords that match the user keyword. As shown in FIG 2, the entity recognition module **20** comprises a related entity keyword module **21**, a new alternative module **22** and an entity affiliations database **23.** When the entity recognition module **20** receives the keyword of the entity of interest input from the entity search portal **10**, the related entity keyword module **21** performs a query search of the input keyword on the entity affiliation database **23** to extract all alternative keywords, and the new alternative module **22** continuously updates list of alternative keywords that match the input keyword in the entity affiliation database **23** with new alternative keywords from predefined articles. The related keywords (such as titles, salutations, other names) of the entity is used to find various combination in order to allow the crawlers a better return of documents. New alternatives can be discovered through Natural Language Analysis.

[0045] The sentence grouping module **30** electronically coupled with the entity recognition module **20** uses the list of keywords from the entity recognition module **20** to do an article search, and then crawls a set of matching articles; for each article, sentences that relate to a specific entity using anophora detection are grouped to provide blocks of sentences relating to the entity of interest.

[0046] The topic extraction module **40** electronically coupled with the sentence grouping module **30** receives the blocks of sentences, and process each and every block of sentences to extract the keywords so as to provide identified topics, identified subjects and topic popularity. As shown in FIG 3, the topic extraction module **40** comprises a keyword extraction module **41** with a subject matter corpus of keywords including predefined topic keywords and subcategories, a key categorization module **42**, a topic grouping module **43**, and a topic popularity module **44**. The keyword extraction module **41** uses the subject matter corpus of keywords to extract topic related keywords from the blocks of sentences from the

sentence grouping module **30**. The key categorization module **42** categorizes the extracted topic related keywords respectively into their subject. The topic grouping module **43** identifies topics by using the maximum number of keywords in a subject to determine the subject and determines the subject matter of the article by further aggregation of the identified topics. The topic popularity module **44** determines the topic popularity using the maximum occurrence of a keyword and its affiliations over multiple articles.

[0047] The sentence weightage module **50** electronically coupled with the topic extraction module **40** receives the blocks of sentences and topic popularities, and extracts adjectives from the input sentences to provide a weighted sentiment score of each block of sentences. As shown in FIG 4, the sentence weightage module **50** comprises an adjective extraction module **51**, an adjective weightage module **52** including a corpus of adjectives with predefined relevant weightage, and an adjective weightage aggregator module **53**. The adjectives are the ones that describe the noun, and adjectives have the ability to describe a noun in different levels. Therefore, the extracted adjectives are weighed accordingly in accordance to the topic. A topic that is not a current issue, but has a strong adjective will have a lower scaled weightage than an adjective describing a current topic. For example "the devastating hurricane in 1930" vs "the devastating hurricane in 2017". The adjective "devastating" in the second occurrence will carry more weight due to its immediate impact to human phycology. The adjective extraction module **51** extracts adjectives from the block sentences. For each block of sentences, together with the identified topic, the adjective weightage module **52** scales the extracted adjectives according to the popularity of the topic to provide a sentiment score for each adjective. The adjective weightage aggregator module **53** aggregates the sentiment scores for each block of sentences to provide a weighted sentiment score of each block sentences.

[0048] The sentiment scoring module **60** electronically coupled with the sentence weightage module **50** receives the blocks of sentences, weighted sentiment scores for the respective block of sentences, and performs sentiment analysis by using existing techniques in searching for verbs and providing a sentiment score value within a range of value from 1-10, where 1 is low positive and 10 is highly positive.

[0049] The sentiment score aggregator module **70** electronically coupled with the sentiment scoring module **60** receives the sentiment scores of each block of sentences relating to the entity, and aggregates the scores from each sentiments per entity/per

document/ per topic to provide list of aggregated sentiment scores of entity, document and topic. This grouping of aggregated sentiments help in expanding the network graph to display the origin of sentiments before being aggregating.

[0050] The sentiment variance module **80** electronically coupled with the topic extraction module **40** and sentiment score aggregator module **70** receives the topic, popularity, document metadata, sentiment per entity, sentiment per document, and sentiment per topic, and scales the sentiment values according to the inputs using an algorithm to provide final sentiment score on each entity based on sentences, topics and documents, and document type being extracted from the metadata. The algorithm here defines a mathematical model in which the sentiment is varied according to the respective inputs. This algorithm could be a as simple as a weighted decision matrix or a machine learning approach. As shown in FIG 5, the sentiment variance module **80** comprises a relevancy metric module **81**, an author credibility reporting module **82** including a database with the authenticity of authors being predefined and updated continuously, a document metadata extraction module **83**, and a document grouping module **84**. Using the information extracted from the document metadata such as authenticity of the author, the topic popularity, the timestamp of the document, the final aggregated sentiment of the entity per topic/ per document will be scaled according to an algorithm.

[0051] The sentiment graph network display module **90** electronically coupled with the sentiment variance module **80** displays the sentiment score values on the GUI; one exemplary format of displaying the sentiment score values is a network graph as shown in FIG 6. The sentiment score values can be zoomed into various groupings as shown in FIG 7. The outputs from the previous modules can be visualized, where a search on an entity will reveal the entities sentiment regarding a topic, and further elaborated to sub topics. The visualization will allow drill down of sentiments to a sentence level.

[0052] The present invention also provides a method for dynamic entity sentiment analysis (DESA).

[0053] Referring now to FIG 8, there is provided a flow chart of overall process workflow for the method for DESA in accordance with one embodiment of the present invention. The details of the operations of the method **800** are explained hereinbelow.

[0054] Referring now to FIG 9, there is provided a flow chart of the main process flow of the method of DESA in accordance with one embodiment of the present invention. The main process flow **900** includes the steps below.

[0055] A step **901**, taking in a user input of a keyword of an entity; this is just an indicator to indicate the Start of the Process.

[0056] At step **902** searching entity based on the input keyword through a network for matching keywords to the entity, and retrieving from the network the documents related to the entity and the related keywords of the input keyword, where the list of documents is continuously updated if a new document is detected. The search would initially start crawling for documents containing the matching keywords. For example, if a user inserts "Dr Mahathir' as a keyword, all documents that contain the matching keyword 'Dr Mahathir' will be collected. Once the related keywords are retrieved, all documents containing the related keywords such as "Tun M", "Mahathir Mohamed" will also be collected.

[0057] At step **903**, searching results based on the retrieved documents containing the matching keywords and storing sentiment value per entity per topic per search results.

[0058] At step **904,** analyzing retrieved document, each of the retrieved documents based on its type and extracting metadata of each document such as author and timestamp.

[0059] At step **905**, Named Entity Recognition (NER) (entity=proper noun), extracting other entities from each document, collecting and grouping organizations, persons or places, and matching or relating new extracted entities to similar input keywords due to acronym or spelling differences.

[0060] At step **906**, grouping of same entities, grouping together all sentences containing matching keywords and related keywords relating to the entity (inclusive of different affiliations), regardless of sentence positions, and generating duplicate sentences if a sentence has more than one entity.

[0061] At step **907**, grouping of adjacent sentences of the same entity into blocks using anaphora detection, for each entity, grouping the sentences adjacent to the sentences containing the keywords into blocks for keyword extraction and adjective extraction per block. A scaled value is associated to each adjective defined in a database. The adjectives in a block of sentence are then aggregated. Keyword extraction is performed from which the keywords are categorized to determine the topic and sub-topics. The count of topic keywords determines the subject of the document and its popularity.

[0062] At step **908,** adjusting sentiment of entity per block by weightage, computing sentiment per block per topic per entity, and scaling with the aggregated adjective weightage block **1003**.

[0063] At step **909**, throughout document per topic, aggregating the sentiment scores per topic basis in a document.

[0064] At step **910**, for sentiment value per entity per topic per search result, aggregating the sentiment scores throughout multiple documents keeping the topic grouping throughout. The sentiment value is further fine-tuned to take into account the author credibility rating and topic relevancy (timestamp). As new documents are collected, the sentiment value will scale depending on topic popularity over time.

[0065] At step **911**, for final network for search results, displaying the results of the dynamic entity sentiment and allowing the user to breakdown the sentiment to its exact score prior to aggregating any sentiment; the exemplary displays are shown in FIG 6 and FIG 7.

[0066] Referring now to FIG 10, there is provided a flow chart of the sentence weightage flow of the method of DESA in accordance with one embodiment of the present invention. The sentence weightage flow provides the adjective weightage to a sentence. Using the identified topic, the adjectives play a different role accordingly. The same adjective could carry different weight in describing a noun depending on the topic. For example, if the topic is on "war" the adjective "bloody" carries a strong weight than a sentence where somebody uses the same word to curse. All the adjectives weightage in a similar topic are aggregated to provide a mean smoothed value throughout the document.

[0067] As shown in FIG 10, the sentence weightage flow **1000** includes the steps below:

[0068] At step **1001**, for adjective extraction per block, obtaining the blocks of grouped adjacent sentences from the step **907**, and extracting the adjectives from the blocks of grouped adjacent sentences to provide a list of adjectives.

[0069] At step **1002**, for adjective weightage depending on topic, matching each adjective of the list of adjectives to a predefined value from a database from the step of Topic identification (subject) (+sub-topic classification) **1103** to generate a list of adjectives with a value appended.

[0070] At step **1003**, for aggregated block weightage, aggregating the values of the adjectives in a block of sentences based on the list of adjectives with a value appended to

provide a single aggregated value per block of sentences for the step of Sentiment of entity per block adjusted by weightage **908**.

[0071] Referring now to FIG 11, there is provided a flow chart of topic extraction and grouping flows of the method of DESA in accordance with one embodiment of the present invention. This method requires a subject corpus where unique words describe a unique topic. By measuring the frequency of usage of these unique keywords, the general topic of the document can be identified. Once a topic is identified, the popularity of the topic is obtained by doing a background search of the topic. The number of hits (count) regarding the topic translates to the popularity of the topic. This popularity has to also take care of the timestamp of the documents. Documents older than a preconfigured date will be omitted in the hit count.

[0072] The topic grouping would also provide information to the display mechanism on how the structure of the topic groups will be displayed in the network graph. The document type will also be grouped so the user is able to segregate sentiment of an entity with regards to the document type such as Facebook, twitter, news, blogs and others.

[0073] The topic extraction and grouping flows **1100** includes the steps below:

[0074] At step **1101**, extracting Keyword which includes obtaining the blocks of grouped adjacent sentences from step **907**, and extracting keywords such as proper nouns from each block of sentences and each document to provide a set of keywords grouped in terms of block of sentence and entire document.

[0075] At step **1102**, categorizing the set of keywords into their respective categories according to a corpus of words to provide the categories involved in the block of sentence and throughout the document.

[0076] At step **1103**, identifying topic (subject) (+sub-topic classification), for the categories involved in the block of sentence and entire document, counting the number of keywords belonging to each category, and determining from the corpus the main topic and sub-topics of the document, where the determined main topic and sub-topics of the document are forwarded to the step of Adjective weightage depending on topic **1002** and the step of Aggregate throughout document per topic **909**;

[0077] At step **1104**, defining popularity of topic depending on topic search hit count, obtaining a historical list of multiple documents, and keeping track of the popularity of topics to provide a list of popular occurring topics to the step of Dynamic relevancy metric (filtering

algorithm) **1305**. During certain time frame certain topics are more frequent and hence more popular;

[0078] At step **1105**, grouping topic by passing the topic group field for the front-end GUI to the step of Final network for search results **911**. This would allow the GUI to allow to expand the entity in terms of the topic grouping.

[0079] At step **1110**, grouping document type by passing to the front-end GUI the document type grouping, where the document is from the step **904**, and the GUI is from the step of the step of Final network for search results **911**. This would allow to group the document in terms of news, blogs, social media and others.

[0080] Referring now to FIG 12, there is provided a flow chart of the entity recognition flow of the method of DESA in accordance with one embodiment of the present invention. FIG 12 describes the method of how other related keywords of an entity is used to enhance the crawler's capability. If a person searches for "trump", the system would automatically include "president trump", "Donald trump", "Donald john trump", "45th president of USA", "Commander in chief" and so on. This corpus of related names will be updated when the NER recognizes an entity with new salutation or titles after receiving confirmation from the user.

[0081] The entity recognition flow **1200** includes the steps below:

[0082] At step **1201**, determining new alternative by obtaining the other entities extracted from the document from the step of NER **905**, and matching the user input keyword to the extracted entities to find other similar relationships to extract new alternatives to the user keywords, where the new alternatives will be added to the database.

[0083] If the step **1201** is in affirmative, at step 1020, finding the related keywords (e.g.: salutations, title) from a defined database a set of alternatives of the user input keyword from the step of Search entity **902** to generate a list of alternatives from the database and also a new set of alternatives from the step of **1201** to the step of Grouping of same entities **906**.

[0084] Referring now to FIG 13, there is provided a flow chart of the sentiment variance flow of the method of DESA in accordance with one embodiment of the present invention. A background search is conducted to find number of articles the author has written, and in what hosting organization the articles have been posted. For example, documents in .org or .net will have higher credibility than in Facebook or twitter. Also authors from news portal would also have substantial credibility due to the type of

profession. The time stamp of the document from the metadata will also provide information in how old the articles is. Older article will carry less weight compared to current issue, as this is also related to human psychology. All of the above will scale the sentiment score from time to time, and constantly provide a dynamic sentiment of an entity.

[0085] The sentiment variance flow **1300** includes the steps below:

[0086] At step **1301**, extracting document metadata from each document from the step **904** to generate a list of information regarding the document such as the author and timestamp.

[0087] At step **1302**, updating credibility rating of the author over time into a database based on the authors background information such as education or field of work to generate a set of ratings that are dynamically updated over time to the database.

[0088] At step **1303**, storing the author's credibility (organization/institute) rating and information according to the subject into a database that is a predefined source and is updated with new authors;

[0089] At step **1304**, ranking of author and timestamp of document according to the timestamp of the document based on the database and the metadata of the document;

[0090] At step **1305**, dynamically scaling the authors ranking according to the topic of the document and timestamp to generate a scaled authors ranking to the step of Sentiment value per entity per topic per search result. The dynamic relevancy metric (filtering algorithm) is based on the ranking of the author's credibility from the step of popularity of topic depending on topic search hit count **1104**. This step ensures the authors ranking is a variable in terms of time and document topic.

[0091] While the present invention has been described with reference to particular embodiments, it will be understood that the embodiments are illustrative and that the invention scope is not so limited. Alternative embodiments of the present invention will become apparent to those having ordinary skill in the art to which the present invention pertains. Such alternate embodiments are considered to be encompassed within the scope of the present invention. Accordingly, the scope of the present invention is defined by the appended claims and is supported by the foregoing description.

CLAIMS

What is claimed is:

1. A dynamic entity sentiment analysis (DESA) system (1), characterized in that said system comprising:

an entity search portal (10) including a graphical user interface (GUI) to allow a user to input a keyword of an entity of interest;

an entity recognition module (20) electronically coupled with the entity search portal (10) to receives the input keyword, and perform a query search of the keyword to find alternative matches to provide list of alternative keywords that match the user keyword;

a sentence grouping module (30) electronically coupled with the entity recognition module (20) to use the list of keywords from the entity recognition module (20) to do an article search, and then crawls a set of matching articles;

a topic extraction module (40) electronically coupled with the sentence grouping module (30) to receive the blocks of sentences, and process each and every block of sentences to extract the keywords so as to provide identified topics, identified subjects and topic popularity;

a sentence weightage module (50) electronically coupled with the topic extraction module (40) to receive the blocks of sentences and topic popularities, and extract adjectives from the input sentences to provide a weighted sentiment score of each block of sentences;

a sentiment scoring module (60) electronically coupled with the sentence weightage module (50) to receive the blocks of sentences, weighted sentiment scores for the respective block of sentences, and perform sentiment analysis by searching for verbs and providing a sentiment score value within a range of value;

a sentiment score aggregator module (70) electronically coupled with the sentiment scoring module (60) to receives the sentiment scores of each block of sentences relating to the entity, and aggregate the scores from each sentiment per entity per document per topic to provide list of aggregated sentiment scores of entity, document and topic;

a sentiment variance module (80) electronically coupled with the topic extraction module (40) and sentiment score aggregator module (70) to receive the topic, popularity, document metadata, sentiment per entity, sentiment per document, and sentiment per topic,

and scales the sentiment values according to the inputs using an algorithm to provide final sentiment score on each entity based on sentences, topics and documents, and document type being extracted from the metadata; and

a sentiment network graph display module (90) electronically coupled with the sentiment variance module (80) to display the sentiment score values on the GUI.

2. The DESA system (1) of claim 1, wherein the entity recognition module (20) comprises a related entity keyword module (21), a new alternative module (22) and an entity affiliations database (23); wherein the entity recognition module (20) receives the keyword of the entity of interest input from the entity search portal (10), the related entity keyword module (21) performs a query search of the input keyword on the entity affiliation database (23) to extract all alternative keywords, and the new alternative module (22) continuously updates list of alternative keywords that match the input keyword in the entity affiliation database (23) with new alternative keywords from predefined articles.

3. The DESA system (1) of claim 1, wherein the topic extraction module (40) comprises:

a keyword extraction module (41) with a subject matter corpus of keywords including predefined topic keywords and subcategories, the keyword extraction module (41) uses the subject matter corpus of keywords to extract topic related keywords from the blocks of sentences from the sentence grouping module (30);

a key categorization module (42) for categorizing the extracted topic related keywords respectively into their subject;

a topic grouping module (43) for identifying topics by using the maximum number of keywords in a subject to determine the subject and determines the subject matter of the article by further aggregation of the identified topics; and

a topic popularity module (44) for determining the topic popularity using the maximum occurrence of a keyword and its affiliations over multiple articles.

4. The DESA system (1) of claim 1, wherein the sentence weightage module (50) comprises:

an adjective extraction module (51) for extracting adjectives from the block sentences, wherein for each block of sentences, together with the identified topic;

an adjective weightage module (52) including a corpus of adjectives with predefined relevant weightage, the adjective weightage module (52) scales the extracted adjectives according to the popularity of the topic to provide a sentiment score for each adjective; and

an adjective weightage aggregator module (53) for aggregating the sentiment scores for each block of sentences to provide a weighted sentiment score of each block sentences.

5. The DESA system (1) of claim 1, wherein the sentiment variance module (80) comprises a relevancy metric module (81), an author credibility reporting module (82) including a database with the authenticity of authors being predefined and updated continuously, a document metadata extraction module (83), and a document grouping module (84).

6. A method for dynamic entity sentiment analysis (DESA), characterized in that said method comprising:

receiving (901) a user input of a keyword of an entity (901);

searching (902) based on the input keyword through a network for matching keywords to the entity, and retrieving from the network documents related to the entity and the related keywords of the input keyword, where a list of documents is continuously updated if a new document is detected;

storing (903) the retrieved documents containing the matching keywords and sentiment value;

analyzing (904) each of the retrieved documents based on its type and extracting metadata of each document;

extracting (905) other entities from each document, collecting and grouping organizations, persons or places, and matching or relating new extracted entities to similar input keywords due to acronym or spelling differences;

grouping (906) all sentences containing matching keywords and related keywords and affiliations relating to the entity, regardless of sentence positions, and generating duplicate sentences if a sentence has more than one entity;

for each entity, grouping (907) the sentences adjacent to the sentences containing the keywords into blocks for keyword extraction and adjective extraction per block (907);

computing (908) sentiment per entity per document per topic, and scaling with aggregated adjective weightage block;

aggregating (909) the sentiment scores per topic basis in a document;

aggregating (910) the sentiment scores throughout multiple documents keeping the topic grouping throughout; and

displaying (911) the results of the dynamic entity sentiment and allowing the user to breakdown the sentiment to its exact score prior to aggregating any sentiment.

7. The method of claim 6, wherein in the step of grouping of adjacent sentences (907) further comprising the steps of associating a scaled value to each adjective defined in a database; the adjectives in a block of sentence are then aggregated; the keyword extraction is performed from which the keywords are categorized to determine the topic and sub-topics; and the count of topic keywords determines the subject of the document and its popularity.

8. The method of claim 6, wherein in the step of aggregating the sentiment scores (910) further comprising fine-tuning the sentiment value taking into account the author credibility rating and topic relevancy with timestamp, and as new documents are collected, the sentiment value will scale depending on topic popularity over time.

9. The method of claim 6, further comprising:

obtaining (1001) the blocks of grouped adjacent sentences from step (907), and extracting the adjectives from the blocks of grouped adjacent sentences to provide a list of adjectives;

matching (1002) each adjective of the list of adjectives to a predefined value from a database to generate a list of adjectives with a value appended; and

aggregating (1003) the values of the adjectives in a block of sentences based on the list of adjectives with a value appended to provide a single aggregated value per block of sentences for the step of (908).

10. The method of claim 9, further comprising:

obtaining (1101) the blocks of grouped adjacent sentences from step (907), and extracting keywords such as proper nouns from each block of sentences and each document to provide a set of keywords grouped in terms of block of sentence and entire document;

grouping (1102) the set of keywords into their respective categories according to a corpus of words to provide the categories involved in the block of sentence and throughout the document;

identifying (1103) topic for the categories involved in the block of sentence and entire document, counting the number of keywords belonging to each category, and determining from the corpus the main topic and sub-topics of the document, where the determined main topic and sub-topics of the document are forwarded to the step of matching each adjective (1002) and the step of aggregating the sentiment scores (909);

defining (1104) popularity of topic depending on topic search hit count to obtain a historical list of multiple documents, and keeping track of the popularity of topics to provide a list of popular occurring topics;

grouping (1105) passing the topic group field for the front-end GUI to the step of displaying the results (911); and

grouping (1110) passing to the front-end GUI the document type grouping (1110).

11. The method of claim 6, further comprising:

obtaining (1201) the other entities extracted from the document from the step of extracting other entities from each document (905), and matching the user input keyword to the extracted entities to find other similar relationships to extract new alternatives to the user keywords, where the new alternatives will be added to the database; and

finding (1202) from a defined database a set of alternatives of the user input keyword from the step of searching entity (902) to generate a list of alternatives from the database and also a new set of alternatives.

12. The method of claim 6, further comprising:

extracting (1301) the metadata from each document from the step of analyzing retrieved document (904) to generate a list of information regarding the document such as the author and timestamp;

updating (1302) dynamically credibility rating of the author over time into a database based on the author background information such as education or field of work to generate a set of ratings;

storing (1303) the author's credibility rating and information according to the subject into the database that is a predefined source and is updated with new authors;

ranking (1304) the author according to the timestamp of the document based on the database and the metadata of the document; and

dynamically scaling (1305) the author ranking according to the topic of the document and timestamp to generate a scaled authors ranking to the step of aggregating the sentiment scores (910), ensuring that the authors ranking is a variable in terms of time and document topic.

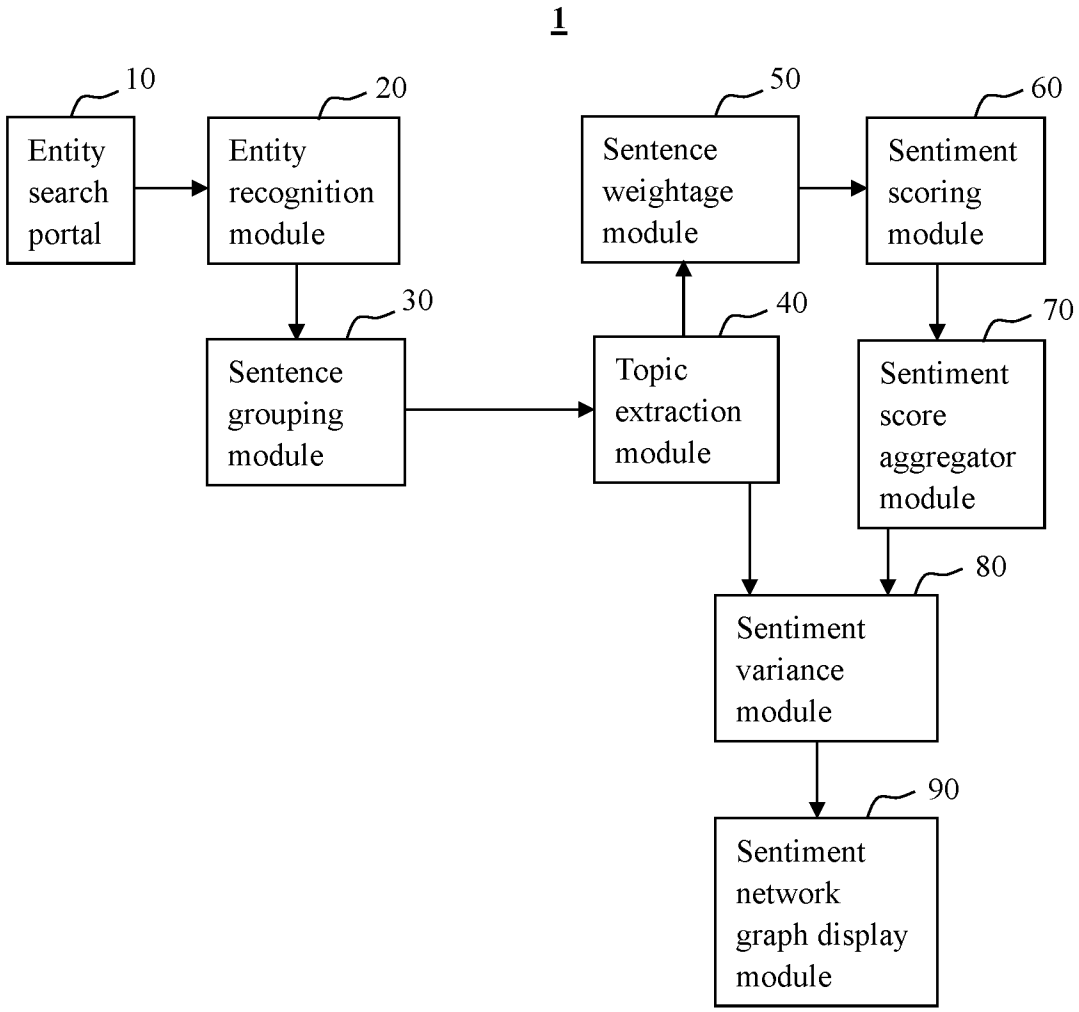
1
10
Entity search portal
20
Entity recognition module
30
Sentence grouping module
40
Topic extraction module
50
Sentence weightage module
60
Sentiment scoring module
70
Sentiment score aggregator module
80
Sentiment variance module
90
Sentiment network graph display module

FIG 1

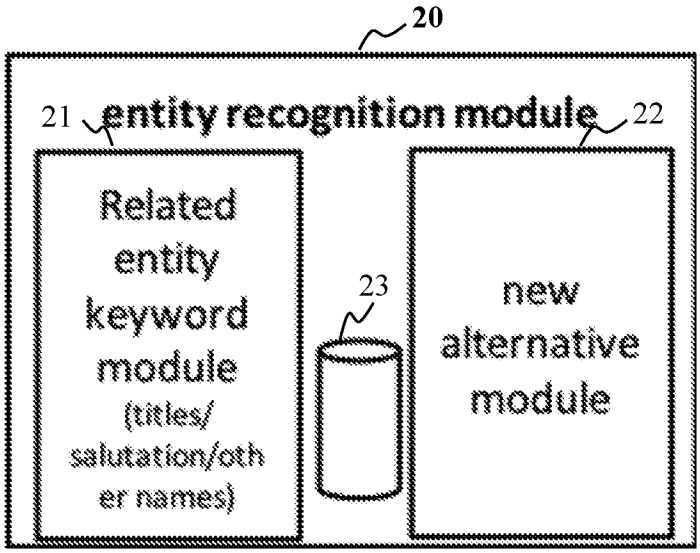
20
21
entity recognition module
22
Related entity keyword module (titles/ salutation/oth er names)
23
new alternative module

FIG 2

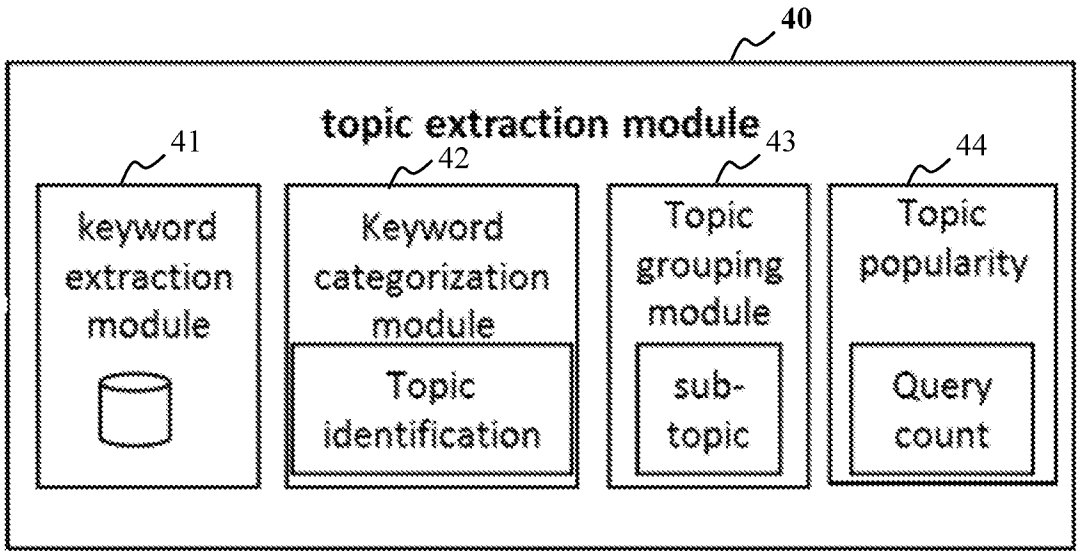
40
topic extraction module
41
keyword extraction module
42
Keyword categorization module
Topic identification
43
Topic grouping module
sub-topic
44
Topic popularity
Query count

FIG 3

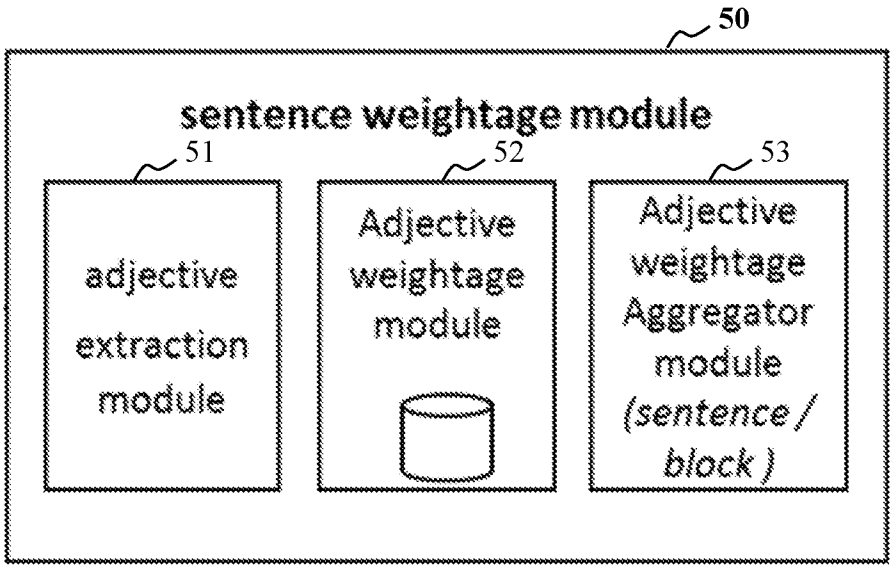
50
sentence weightage module
51
52
53
adjective extraction module
Adjective weightage module
Adjective weightage Aggregator module (sentence / block)

FIG 4

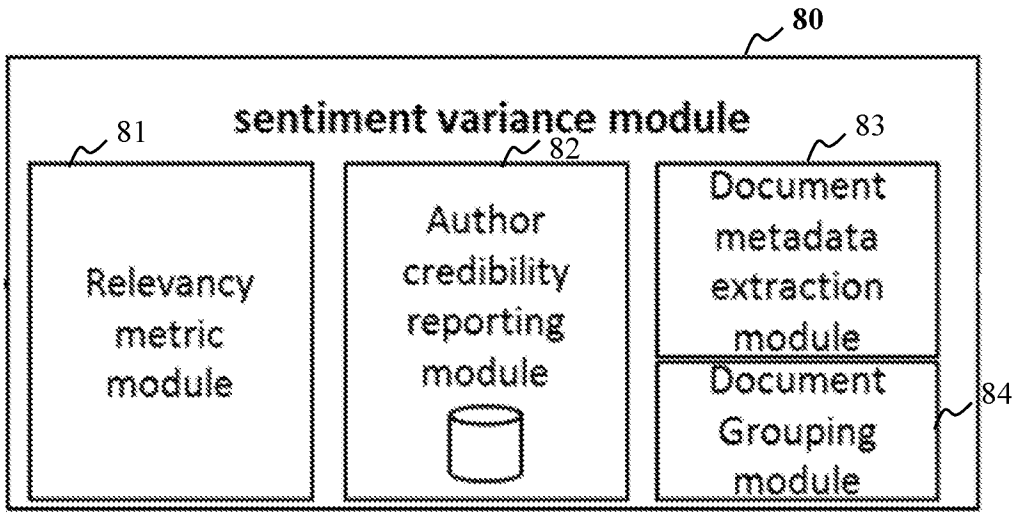
80
sentiment variance module
81
82
83
84
Relevancy metric module
Author credibility reporting module
Document metadata extraction module
Document Grouping module

FIG 5

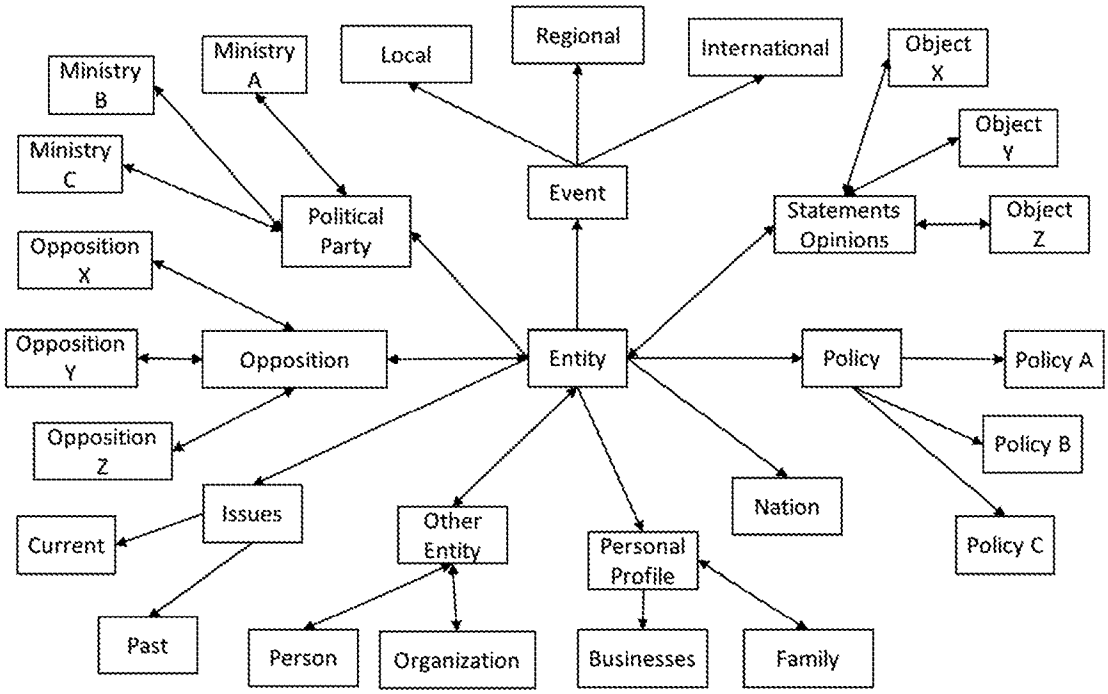
Ministry B
Ministry A
Local
Regional
International
Object X
Ministry C
Object Y
Event
Political Party
Statements Opinions
Object Z
Opposition X
Opposition Y
Opposition
Entity
Policy
Policy A
Opposition Z
Policy B
Issues
Other Entity
Nation
Current
Personal Profile
Policy C
Past
Person
Organization
Businesses
Family

FIG 6

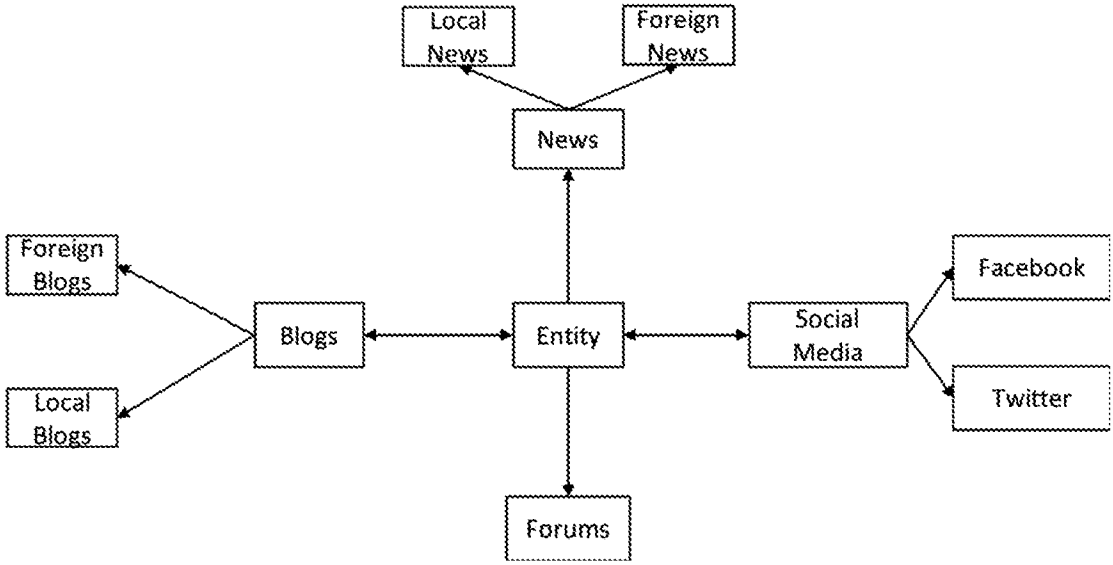
Local News
Foreign News
News
Foreign Blogs
Blogs
Local Blogs
Entity
Social Media
Facebook
Twitter
Forums

FIG 7

800

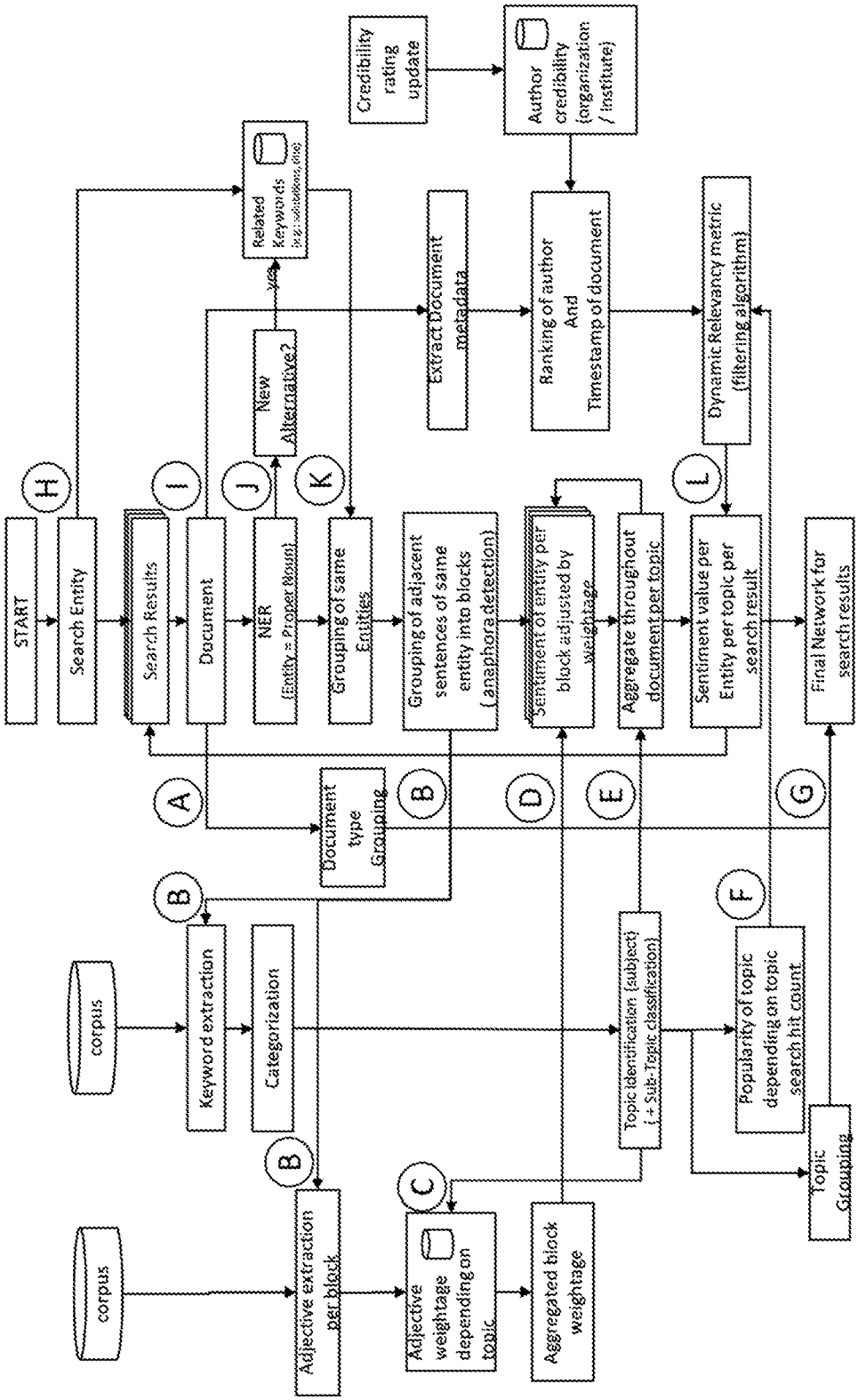
START
Search Entity
Search Results
Document
NER
(Entity = Proper Noun)
Grouping of same Entities
Grouping of adjacent sentences of same entity into blocks (anaphora detection)
Sentiment of entity per block adjusted by weightage
Aggregate throughout document per topic
Sentiment value per Entity per topic per search result
Final Network for search results
Related Keywords
New Alternative?
yes
Extract Document metadata
Ranking of author And Timestamp of document
Dynamic Relevancy metric (filtering algorithm)
Credibility rating update
Author credibility (organization / institute)
Document type Grouping
corpus
Keyword extraction
Categorization
Topic Identification (subject) (+ Sub-Topic classification)
Popularity of topic depending on topic search hit count
Topic Grouping
corpus
Adjective extraction per block
Adjective weightage depending on topic
Aggregated block weightage
A
B
C
D
E
F
G
H
I
J
K
L

FIG 8

900

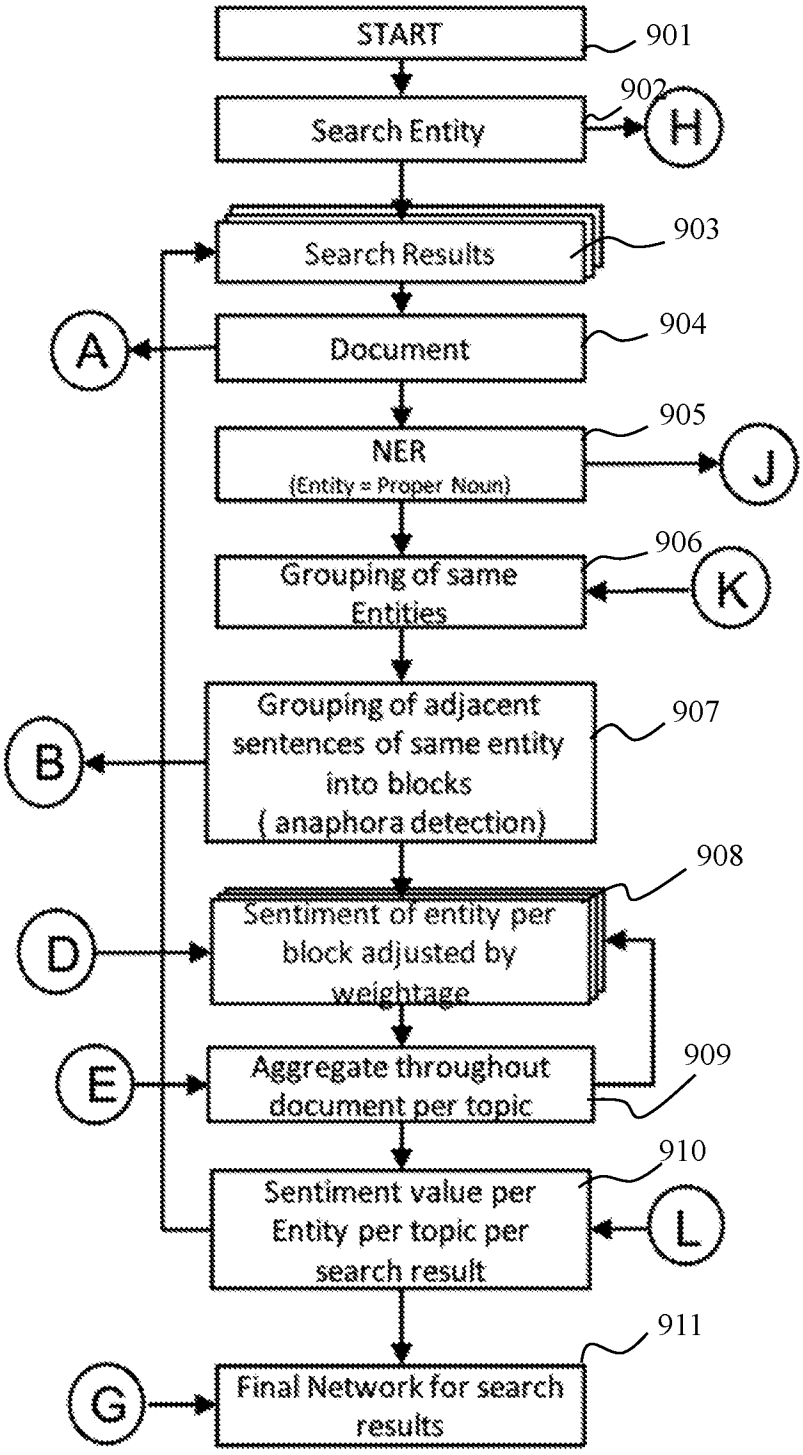
START
901
Search Entity
902
H
Search Results
903
A
Document
904
NER
(Entity = Proper Noun)
905
J
Grouping of same Entities
906
K
Grouping of adjacent sentences of same entity into blocks (anaphora detection)
907
B
Sentiment of entity per block adjusted by weightage
908
D
Aggregate throughout document per topic
909
E
Sentiment value per Entity per topic per search result
910
L
Final Network for search results
911
G

FIG 9

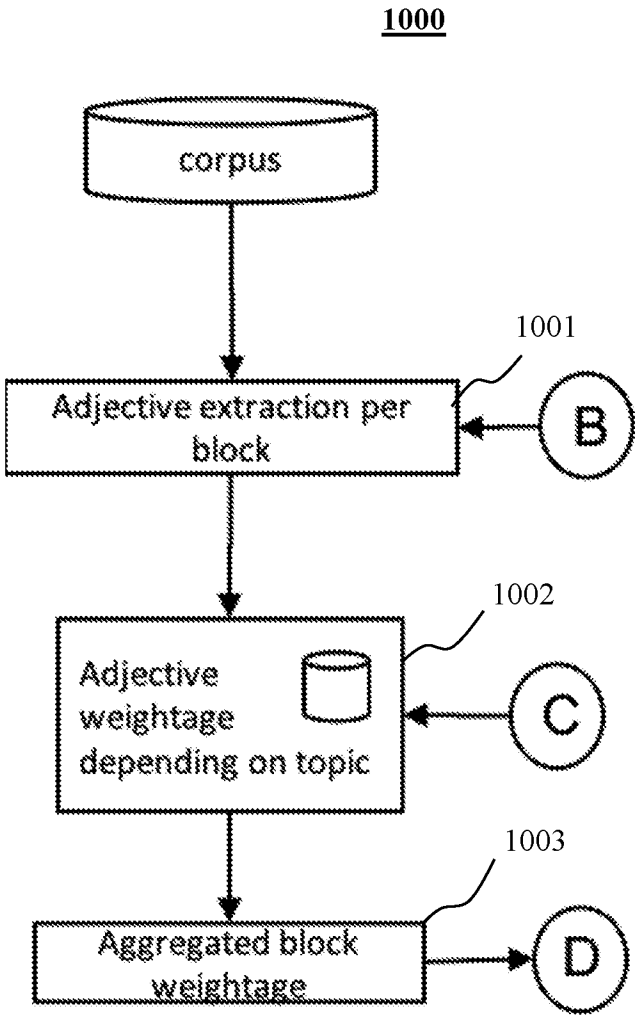
1000
corpus
1001
Adjective extraction per block
B
1002
Adjective weightage depending on topic
C
1003
Aggregated block weightage
D

FIG 10

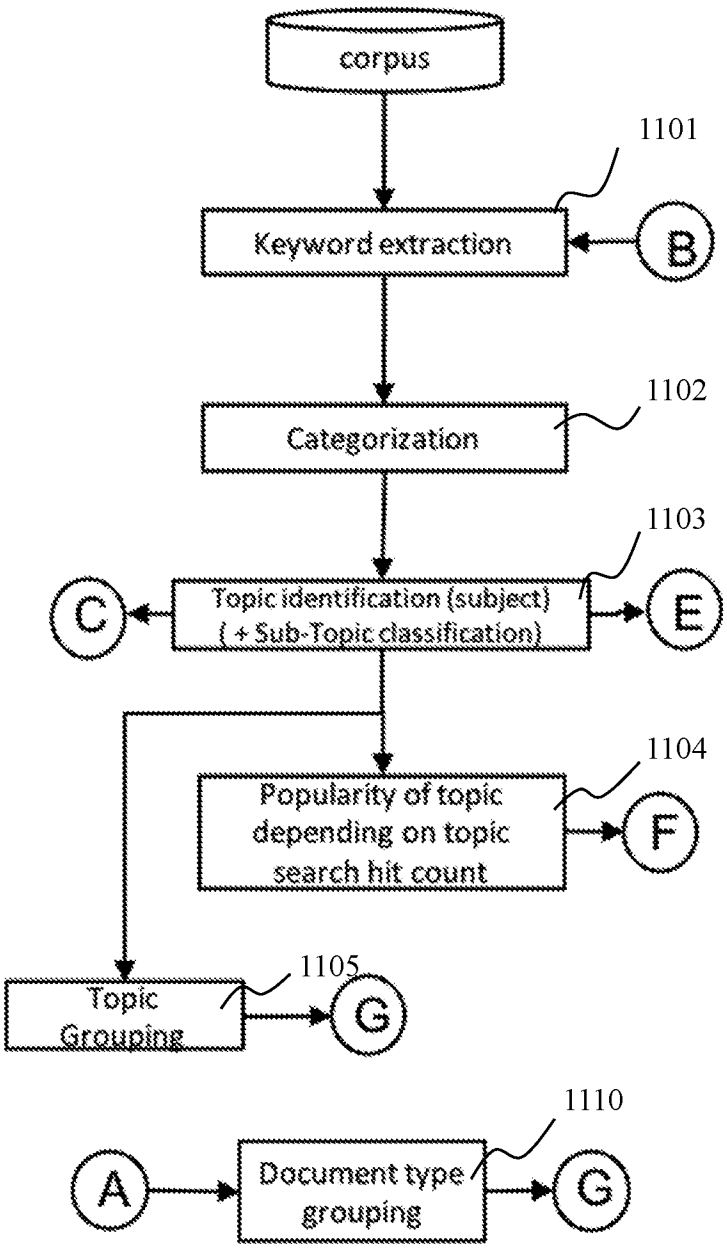
1100
corpus
1101
Keyword extraction
B
1102
Categorization
1103
C
Topic identification (subject)
(+ Sub-Topic classification)
E
1104
Popularity of topic
depending on topic
search hit count
F
1105
Topic
Grouping
G
1110
A
Document type
grouping
G

FIG 11

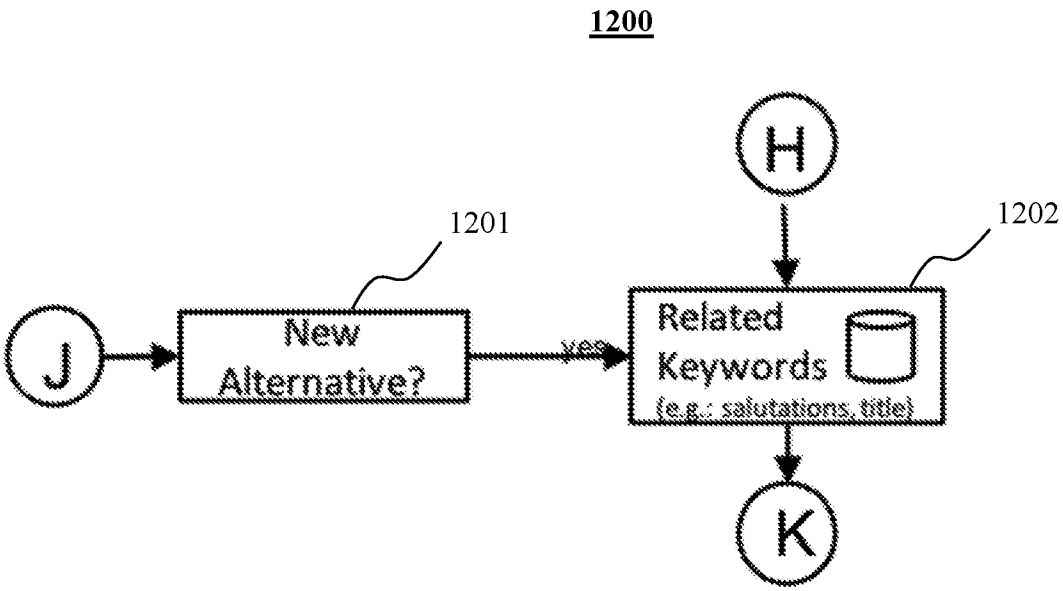
1200
H
1201
1202
J
New Alternative?
yes
Related Keywords
(e.g.: salutations, title)
K

FIG 12

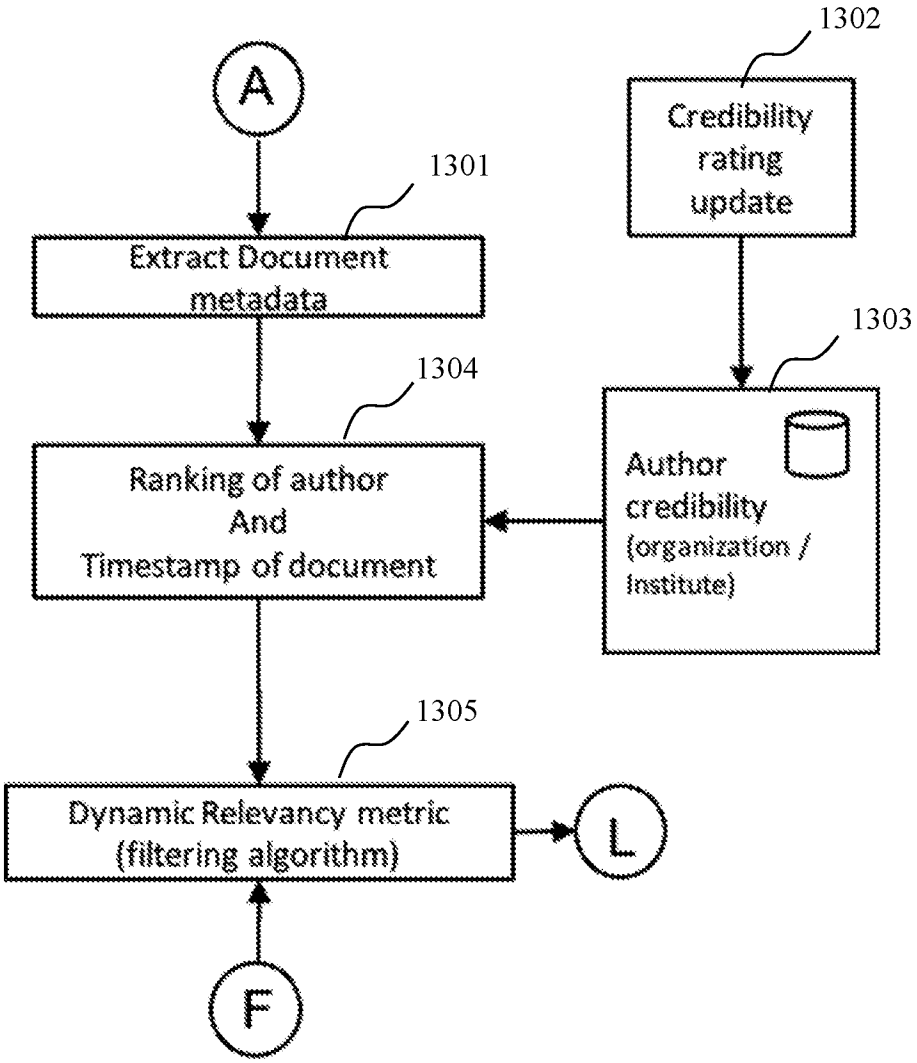
1300
1302
A
Credibility rating update
1301
Extract Document metadata
1303
1304
Ranking of author And Timestamp of document
Author credibility (organization / Institute)
1305
Dynamic Relevancy metric (filtering algorithm)
L
F

FIG 13

INTERNATIONAL SEARCH REPORT

International application No.
PCT/MY2019/050092

A. CLASSIFICATION OF SUBJECT MATTER

G06F 40/30(2020.01)i, G06F 40/211(2020.01)i, G06F 40/279(2020.01)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 40/30; G06F 17/20; G06F 17/27; G06F 17/28; G06F 17/30; G06N 5/02; G06Q 50/18; G06F 40/211; G06F 40/279

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & Keywords: sentiment, keyword, score, entity, document, topic, adjacent sentences, adjective, grouping

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2017-0220652 A1 (FACEBOOK, INC.) 03 August 2017 paragraphs [0001]-[0003], [0007]-[0006], [0043]-[0072]; and figure 3	1-12
A	US 2018-0082389 A1 (INTERNATIONAL BUSINESS MACHINES CORP.) 22 March 2018 paragraphs [0017]-[0056]; and figures 2-3A	1-12
A	KR 10-1423549 B1 (KOREA UNIVERSITY RESEARCH AND BUSINESS FOUNDATION) 01 August 2014 paragraphs [0018]-[0060]; and figures 1-4	1-12
A	US 2015-0286627 A1 (ADOBE SYSTEMS INC.) 08 October 2015 paragraphs [0029]-[0049]; and figures 2-4	1-12
A	JP 2013-525868 A (MINH DUONG-VAN) 20 June 2013 paragraphs [0021]-[0026]; and figures 3-5	1-12

☐ 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
"D" document cited by the applicant in the international application
"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	Date of mailing of the international search report
23 March 2020 (23.03.2020)	**23 March 2020 (23.03.2020)**

Name and mailing address of the ISA/KR
International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea
Facsimile No. +82-42-481-8578

Authorized officer
KWON, Sungho
Telephone No. +82-42-481-3547

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/MY2019/050092

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2017-0220652 A1	03/08/2017	US 10242074 B2	26/03/2019
US 2018-0082389 A1	22/03/2018	None	
KR 10-1423549 B1	01/08/2014	KR 10-2014-0053717 A	08/05/2014
		US 2015-0227528 A1	13/08/2015
		WO 2014-065630 A1	01/05/2014
US 2015-0286627 A1	08/10/2015	None	
JP 2013-525868 A	20/06/2013	CN 102812475 A	05/12/2012
		EP 2517156 A1	31/10/2012
		US 2011-0161071 A1	30/06/2011
		US 2012-0101808 A1	26/04/2012
		US 8849649 B2	30/09/2014
		US 9201863 B2	01/12/2015
		WO 2011-079311 A1	30/06/2011