



(22) Date de dépôt/Filing Date: 2011/07/22

(41) Mise à la disp. pub./Open to Public Insp.: 2013/01/22

(45) Date de délivrance/Issue Date: 2022/08/23

(62) Demande originale/Original Application: 3 010 817

(51) Cl.Int./Int.Cl. *G06F 40/169* (2020.01),
G06F 16/33 (2019.01), *G06F 16/95* (2019.01)

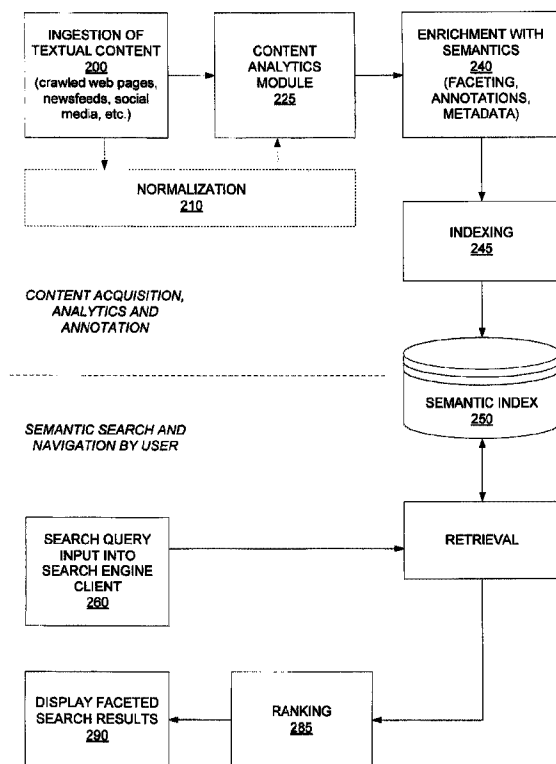
(72) Inventeurs/Inventors:
DIMASSIMO, PASCAL, CA;
PETTIGREW, STEVE, CA;
BROUSSEAU, MARTIN, CA;
SIMARD, CHARLES-OLIVIER, CA;
WILLIAMS, ERIC, CA;
LACROIX, FRANCIS, CA;
...

(73) Propriétaire/Owner:
OPEN TEXT CORPORATION, CA

(74) Agent: ROBIC

(54) Titre : METHODES, SYSTEMES ET SUPPORT INFORMATIQUE POUR L'ENRICHISSEMENT SEMANTIQUE DU
CONTENU ET LA NAVIGATION SEMANTIQUE

(54) Title: METHODS, SYSTEMS, AND COMPUTER-READABLE MEDIA FOR SEMANTICALLY ENRICHING CONTENT
AND FOR SEMANTIC NAVIGATION



(57) Abrégé/Abstract:

Methods, systems and computer-readable media enable various techniques related to semantic navigation. One aspect is a technique for displaying semantically derived facets in the search engine interface. Each of the facets comprises faceted search

(72) **Inventeurs(suite)/Inventors(continued)**: DOWGAILENKO, ALEX, CA; TEXIER, JEAN-MICHEL, CA;
DELIGIA, AGOSTINO, CA

(57) **Abrégé(suite)/Abstract(continued)**:

results. Each of the faceted search results is displayed in association with user interface elements for including or excluding the faceted search result as additional search terms to subsequently refine the search query. Another aspect automatically infers new metadata from the content and from existing metadata and then automatically annotates the content with the new metadata to improve recall and navigation. Another aspect identifies semantic annotations by determining semantic connections between the semantic annotations and then dynamically generating a topic page based on the semantic connections.

ABSTRACT

Methods, systems and computer-readable media enable various techniques related to semantic navigation. One aspect is a technique for displaying semantically derived facets in the search engine interface. Each of the facets comprises faceted search results. Each of the faceted search results is displayed in association with user interface elements for including or excluding the faceted search result as additional search terms to subsequently refine the search query. Another aspect automatically infers new metadata from the content and from existing metadata and then automatically annotates the content with the new metadata to improve recall and navigation. Another aspect identifies semantic annotations by determining semantic connections between the semantic annotations and then dynamically generating a topic page based on the semantic connections.

METHODS, SYSTEMS, AND COMPUTER-READABLE MEDIA FOR
SEMANTICALLY ENRICHING CONTENT AND FOR SEMANTIC
NAVIGATION

TECHNICAL FIELD

[0001] The present technology relates generally to computer systems and information technology and, in particular, to content management and content searching based on semantics.

BACKGROUND

[0002] Conventional techniques for annotating, managing and searching content is generally limited to keyword searching and manually annotated meta tags. Frequently keyword searches yield large numbers of hits with only marginal actual relevance. The problem of locating information remains a daunting challenge.

[0003] Semantics-based technologies represent one promising avenue for addressing this challenge. Here, semantics refers to the explicit or derived (inferred) meaning provided by a fragment of information. At the document level, this meaning can be factual information like "J. Doe is the CEO of Acme Corp." or a piece of information that is inferred from the context like "Acme is a company" or "This article is about mergers & acquisitions". At the inter-document level, the semantic information gathered at the document level can be reused by the computer system to help derive information at a higher level. For example: "In the last month, there has been a trend showing negative criticism of Acme Corp".

[0004] Some examples of semantic analysis techniques are disclosed in US 2011/00822863 (Chang et al.) and US 2010/0312743 (Thompson et al.). Some examples of semantic search engines are disclosed in US 2007/0124263 (Katariya et

al.), US 2010/0036797 (Wong et al.), US 2011/0072021 (Lu et al.) and US 2010/0268702 (Wissner et al.). Some semantic annotation techniques are disclosed in US 2008/0221874 (Cao et al.). An indexing technique is disclosed in US 2009/0063550 (van den Berg et al.). A text categorization method and apparatus is disclosed in US Patent 7,356,461 (Tremblay et al.). Various faceted search and display techniques are disclosed in US Patent 7,774,383 (Acevedo-Aviles et al.), US Patent 7,152,064 (Bourdoncle et al.), US 2010/0198637 (Wu et al.) US 2009/0106244 (Dash et al.), US 2007/0185836 (Handy-Bosma et al.), and US 2009/0234849 (Erera et al.).

[0006] Improvements on these known techniques remain highly desirable. Such improvements are disclosed in the present specification and the appended drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] Further features and advantages of the present technology will become apparent from the following detailed description, taken in combination with the appended drawings, in which:

[0008] FIG. 1 is a schematic depiction of a networked computer system on which embodiments of the present invention may be implemented;

[0009] FIG. 2 is a schematic depiction of a computing device for implementing embodiments of the present invention;

[0010] FIG. 3 is a functional block diagram illustrating a process of creating semantically enriched data by content ingestion, analytics, annotation, and indexing as well a process for semantic search and navigation;

[0011] FIG. 4 is a depiction of a semantic navigation user interface for presenting faceted search results in accordance with one embodiment of the present invention;

[0012] FIG. 5 is a depiction of a further user interface after a user has included a faceted element to refine the search of FIG. 4;

[0013] FIG. 6 is a depiction of yet a further interface after the user has further included a faceted element and excluded another faceted element to further refine the search of FIG. 5;

[0014] FIG. 7 is a flowchart depicting steps of a method of annotating content generating metadata by inferring new metadata from the content and from existing metadata;

[0015] FIG. 8 is a flowchart depicting steps of a method of dynamically generating a topic page based on semantic connections made between annotations; and

[0016] FIG. 9 is a schematic depiction of a semantic navigation system in accordance with one implementation of the present invention.

[0017] It will be noted that throughout the appended drawings, like features are identified by like reference numerals.

SUMMARY

[0018] The present invention generally provides novel methods, systems, and computer-readable media that provide improvements for generating semantically-enriched content and for semantic navigation.

[0019] One aspect of the present invention is a method of displaying search results on a computing device. The method

entails receiving search results in response to a search query, displaying a list of search results in a search engine interface and displaying a plurality of facets in the search engine interface. Each of the facets is derived from explicit editorial metadata or inferred semantic annotations and comprises faceted search results. Each of the faceted search results is displayed in association with user interface elements for including or excluding the faceted search result as additional search terms to subsequently refine the search query.

[0020] Another aspect of the present invention is a method of presenting content on a computing device. The method entails identifying a plurality of semantic annotations for content, determining one or more semantic connections between the semantic annotations, and dynamically generating a topic page based on the one or more semantic connections.

[0021] Yet another aspect of the present invention is a method, performed by a computing device, for annotating content. The method entails automatically inferring new metadata from the content and from existing metadata and automatically annotating the content with the new metadata. New metadata generated using this method can be used to improve the relevancy of the search results and enhance the search and navigation experience for the end-user of the computing device.

[0022] Yet another aspect of the present invention is a method for semantically searching content, the method comprising displaying a first facet and a second facet as part of faceted search results in response to a search query, associating a first facet with a second facet in response to user input, and searching a semantic index based on the first and second facets by cross-referencing metadata for the

content associated with the first facet with metadata for the content associated with the second facet.

[0023] Yet a further aspect of the invention is a semantic navigation system comprising an ingestion broker for ingesting content from client systems, a web crawler operatively connected to the ingestion broker for crawling web content for the ingestion broker, an engine for operating a semantic index, a content analytics module operatively connected to the engine for analyzing content to generate semantic metadata for the semantic index, and a presentation module having presentation logic for presenting search results obtained by querying the semantic index. The presentation module includes a widget server and a template engine that allow widgets to be aggregated into a theme. A theme is designed to address specific business needs. A complete semantic search and navigation application can be made of one or several themes.

[0023a] More particularly, there is provided a semantic navigation system comprising:

- an ingestion broker for ingesting content from client systems;
- a web crawler operatively connected to the ingestion broker for crawling web content for the ingestion broker;
- an engine for operating a semantic index;
- a content analytics module operatively connected to the ingestion broker and to the engine for analyzing content from the ingestion broker and generating semantic metadata for the engine to index into the semantic index; and

- a presentation module having presentation logic for presenting search results obtained by querying the semantic index.

[0024] Various techniques including normalization of entities and categories may be used. Fine tuning of the relevancy and the document ranking can be accomplished by applying boosts to different types of metadata.

[0025] Another aspect of the invention is a computer-readable medium upon which are stored instructions in code that are configured to perform any one of the foregoing methods when the computer-readable medium is loaded into memory and executed on a processor of a computing device.

[0025a] More particularly, there is provided a computer-readable medium comprising programmed instructions in code which, when loaded into a memory and executed by a processor of a computing device, causes the computing device to:

- ingest content from client systems;
- analyze the content;
- generate semantic metadata from an analysis of the content;
- index the semantic metadata into a semantic index; and
- present search results obtained by querying the semantic index.

[0025b] There is further provided a method, comprising:

performing, by a computing device having a processor and a non-transitory computer readable medium, an ingestion process to ingest content of different formats sourced from a plurality of data sources into search indices, the ingestion process including:

inputting into an ingestion broker embodied on the non-transitory computer readable medium, the content of different formats sourced from a plurality of data sources, the ingestion broker supporting chaining of processors such that, during the ingestion process and prior to any indexing process, an indexing engine can call a content analytics module or a semantic annotator that semantically analyzes, annotates, and relates data in the content before the content is indexed;

normalizing, by the ingestion broker, the content of different formats into a uniform representation that is supported by the indexing engine and that is deliverable across multiple digital channels for a variety of applications;

calling the content analytics module or the semantic annotator to enrich the content having the uniform representation from the ingestion broker, the calling performed by the indexing engine; and

analyzing the content having the uniform representation, the analyzing performed by the content analytics module or the semantic annotator during the ingestion process and prior to any indexing process, the analyzing including:

extracting information from a syntactic structure reflecting semantics of the content having the uniform representation;

semantically mapping the information extracted from the syntactic structure with information about meanings of words in the content having the uniform representation in order to determine semantic metadata; and

associating the semantic metadata with the words in the content having the uniform representation such that the content having the uniform representation is semantically enriched with the semantic metadata during the ingestion process; and

subsequent to the ingestion process, indexing the content having the uniform representation, the indexing performed by the indexing engine based on both syntax and the semantic metadata determined by the content analytics module or the semantic annotator during the ingestion process, the indexing including storing the semantic metadata in a semantic index, wherein the content having the uniform representation is searchable via the semantic index by association with the semantic metadata stored in the semantic index.

[0025c] There is also provided a computer program product comprising a non-transitory computer readable medium storing instructions translatable by a data integration server to:

perform an ingestion process to ingest content of different formats sourced from a plurality of data sources into search indices, the data integration server supporting chaining of processors such that, during the ingestion process and prior to any indexing process, an indexing engine can call a content analytics module or a semantic annotator that semantically analyzes, annotates, and relates data in the content before the content is indexed, the ingestion process including:

receiving content of different formats sourced from a plurality of data sources;

normalizing the content of different formats into a uniform representation that is supported by the indexing engine of the data integration server and that is deliverable across multiple digital channels for a variety of applications;

calling the content analytics module or the semantic annotator to enrich the content having the uniform representation; and

analyzing the content having the uniform representation, the analyzing performed by the content analytics module or the semantic annotator during the ingestion process and prior to any indexing process, the analyzing including:

extracting information from a syntactic structure reflecting semantics of the content having the uniform representation;

semantically mapping the information extracted from the syntactic structure with information about meanings of words in the content having the uniform representation in order to determine semantic metadata; and

associating the semantic metadata with the words in the content having the uniform representation such that the content having the uniform representation is semantically enriched with the semantic metadata during the ingestion process; and

subsequent to performing the ingestion process, indexing the content having the uniform representation, the indexing performed based on both syntax and the semantic metadata

determined during the ingestion process, the indexing including storing the semantic metadata in a semantic index, wherein the content having the uniform representation is searchable via the semantic index by association with the semantic metadata stored in the semantic index.

[0025d] There is also provided a system, comprising:

a processor; and

a non-transitory computer readable medium storing instructions translatable by the processor to:

perform an ingestion process to ingest content of different formats sourced from a plurality of data sources into search indices, the system supporting chaining of processors such that, during the ingestion process and prior to any indexing process, an indexing engine can call a content analytics module or a semantic annotator that semantically analyzes, annotates, and relates data in the content before the content is indexed, the ingestion process including:

receiving content of different formats sourced from a plurality of data sources;

normalizing the content of different formats into a uniform representation that is supported by the indexing engine and that is deliverable across multiple digital channels for a variety of applications;

calling the content analytics module or the semantic annotator to enrich the content having the uniform representation; and

analyzing the content having the uniform representation, the analyzing performed by the content analytics module or the semantic annotator during the ingestion process and prior to any indexing process, the analyzing including:

extracting information from a syntactic structure reflecting semantics of the content having the uniform representation;

semantically mapping the information extracted from the syntactic structure with information about meanings of words in the content having the uniform representation in order to determine semantic metadata; and

associating the semantic metadata with the words in the content having the uniform representation such that the content having the uniform representation is semantically enriched with the semantic metadata during the ingestion process; and

subsequent to performing the ingestion process, indexing the content having the uniform representation, the indexing performed based on both syntax and the semantic metadata determined during the ingestion process, the indexing including storing the semantic metadata in a semantic index, wherein the content having the uniform representation is searchable via the semantic index by association with the semantic metadata stored in the semantic index.

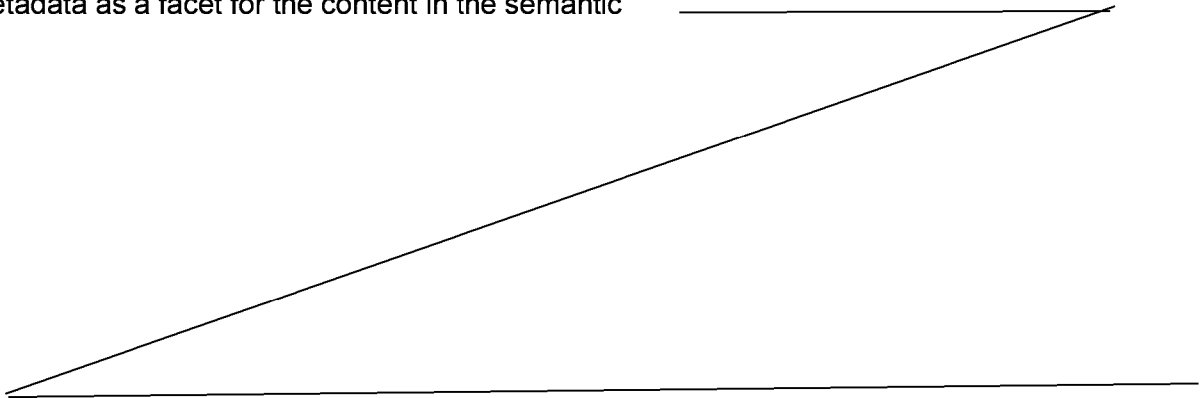
[0025e] In another aspect, there is provided a method, performed by a computing device, for annotating content, the method comprising: automatically inferring new semantic metadata from existing semantic metadata and from content ingested by an ingestion broker of a semantic navigation system operating the computing device and automatically annotating the content with the new semantic metadata.

[0025f] In another aspect, there is provided a computer-readable medium comprising programmed instructions in code which, when loaded into a memory and executed by a processor of a computing device, causes the computing device to: automatically infer new semantic metadata from existing semantic metadata and from content ingested by an ingestion broker of a semantic navigation system operating the computing device and automatically annotate the content with the new metadata.

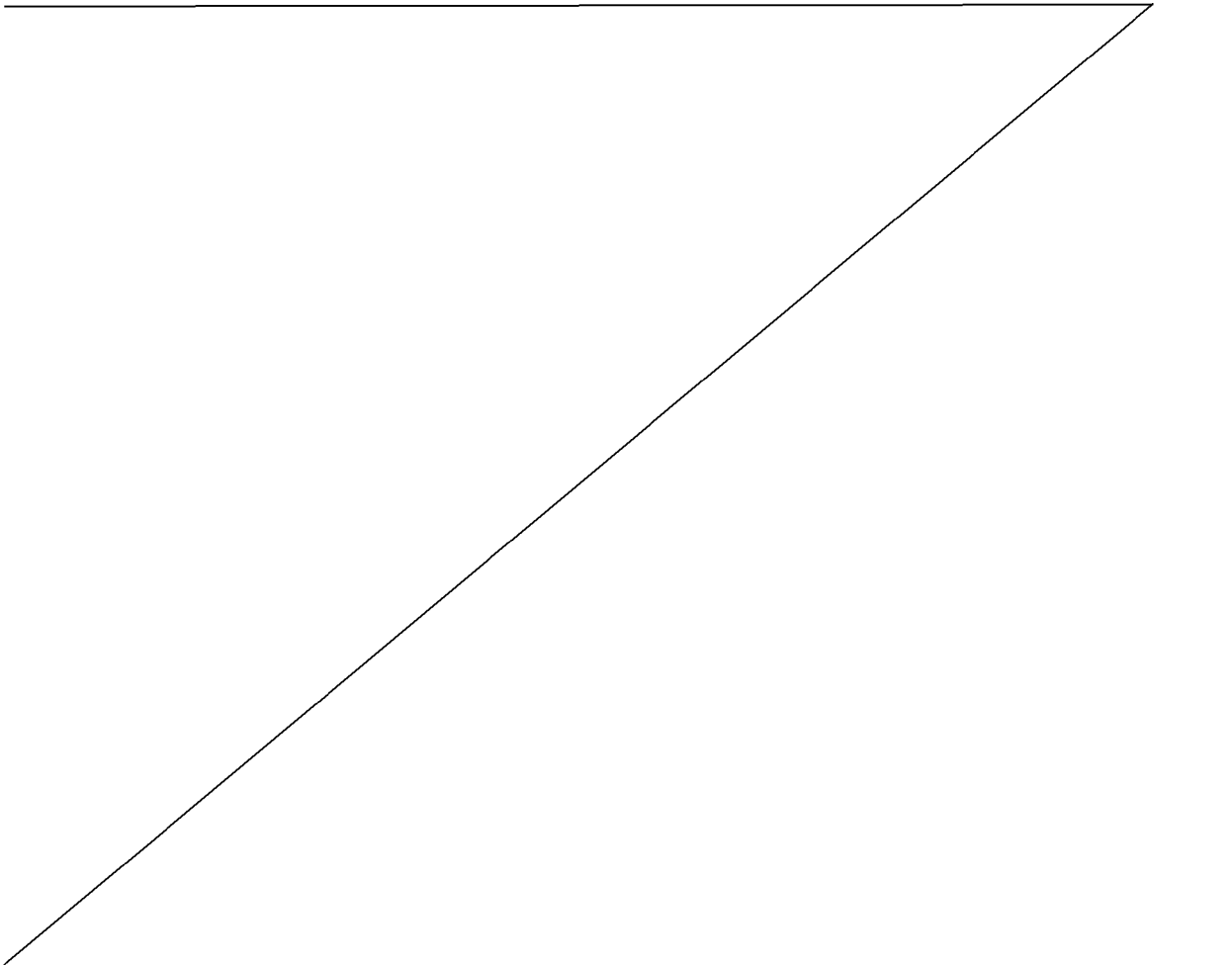
[0025g] In another aspect, there is provided a method of semantically enriching content during content ingestion for semantic navigation during search and discovery, the method comprising: enriching content by creating facets in an enriching stage of a plurality of stages; and enhancing a search and navigation experience with faceting in an enhancing stage of the plurality of stages; wherein the enriching comprises: determining semantic metadata from content ingested by an ingestion broker of a semantic navigation system operating on a server machine, the ingestion broker including an ingestion pipeline for processing the content from a plurality of data sources, the determining performed by a content analytics module or a semantic annotator operatively connected to an engine operating on the server machine, the determining including automatically generating a piece of derived or inferred metadata from the content; adding the piece of derived or inferred metadata to the content by storing the piece of derived or inferred metadata as a facet for the content in the semantic index, the storing performed by the engine; and wherein the enhancing comprises: receiving, by the engine through a search

engine interface of the engine, a search query from a user; searching, by the engine, the semantic index for documents or portions of documents that are relevant to a meaning of the search query, the searching including analyzing the search query, comparing query terms in the search query with information stored in the semantic index, the information stored in the semantic index including the facet for the content, and instantiating metadata associated with the documents of the portions of documents as facets; and responsive to the search query, presenting semantic facets to the user through the search engine interface, the semantic facets including a first semantic facet instantiated by the engine in response to the search query and a second semantic facet retrieved from the semantic index, each of the semantic facets capable of triggering its own faceted search results responsive to navigation or refinement of a search criterion by the user through the search engine interface, the presenting performed by the search engine interface on a client computer communicatively connected to the server machine.

[0025h] In another aspect, there is provided a system for semantically enriching content during content ingestion for semantic navigation during search and discovery, the system comprising a hardware processor; a non-transitory computer readable medium; and stored instructions translatable by the processor for: enriching content by creating facets in an enriching stage of a plurality of stages; and enhancing a search and navigation experience with faceting in an enhancing stage of the plurality of stages; wherein the enriching comprises: determining semantic metadata from content ingested by an ingestion broker of the system, the ingestion broker including an ingestion pipeline for processing the content from a plurality of data sources, the determining being performed by a content analytics module or a semantic annotator and including automatically generating a piece of derived or inferred metadata from the content; adding the piece of derived or inferred metadata to the content by storing the piece of derived or inferred metadata as a facet for the content in the semantic



index, the storing performed by an engine of the system; and wherein the enhancing comprises: receiving a search query from a user through a search engine interface of the engine; searching the semantic index for documents or portions of documents that are relevant to a meaning of the search query, the searching including analyzing the search query, comparing query terms in the search query with information stored in the semantic index, the information stored in the semantic index including the facet for the content, and instantiating metadata associated with the documents or the portions of documents as facets; and responsive to the search query, presenting semantic facets to the user through the search engine interface, the semantic facets including a first semantic facet instantiated by the engine in response to the search query and a second semantic facet retrieved from the semantic index, each of the semantic facets capable of triggering its own faceted search results responsive to navigation or refinement of a search criterion by the user through the search engine interface, the presenting performed by the search engine interface on a client computer communicatively connected to the system.



[0026] The details and particulars of these aspects of the invention will now be described below, by way of example, with reference to the attached drawings.

DETAILED DESCRIPTION

[0027] Embodiments of the present invention, which are described below, enable various new techniques pertaining to the generation of semantically enriched content and to semantic search and navigation. As will be elaborated below, this technology provides innovative tools for managing and delivering content.

[0028] By way of overview, and to place the novel technology in its proper context, the novel technology should be understood as being part of one or more processes of creating semantically-enriched content and searching semantically (semantic navigation). In general, there are five main stages or processes: (1) normalizing data content, (2) enriching the data content by semantic annotations, (3) storing the data content, (4) enhancing the search and navigation experience with faceting and (5) retrieving or delivering the enriched data content in response to search queries and/or facet based filters. In the first stage, data is normalized from a variety of different sources such as, for example, crawled Web pages, CMS, blogs, forums, XML formats, text, PDF documents or MS Office documents. In the second stage, the data content is enriched by creating facets, such as, for example, people, places, organizations, concepts, categories, summaries or sentiments. In the third stage, the data content is stored in one or more of a number of different forms such as, for example, as articles, products, bios, reviews, etc. In the fourth and fifth stages, the enriched data---content is retrieved or delivered in the form of, for example, a site search, topic pages, widgets, syndication, etc.

[0029] With faceted search results, the user can rapidly navigate to other content that may be of interest. Existing structured metadata in the content and the automatically extracted metadata may be stored as facets for every document. Facets may include people, places, organizations, similar

documents, hot topics, sentiment analysis, or the like. Facets may be used to improve result relevancy, suggest related documents, create topic pages and automatically syndicate relevant and timely content.

[0030] In one implementation, searching may be enhanced by a similar-document search (i.e. a search for documents that are "similar to document X"). Alternatively, in another implementation, the interface may enable a search for an "alternative results set" in which the handler returns documents that are potentially related to the search query even if they do not contain the search terms.

[0031] Semantic navigation thus increases the efficacy of a search engine by increasing the likelihood that meaningful and relevant content will be located by a search. Semantic navigation has the effect of breaking content silos and unwrapping content with an intuitive search and discovery experience.

[0032] FIG. 1 schematically depicts an exemplary computer network environment 100 for implementing embodiments of the present invention. As illustrated, a computer-based information search system includes client computers 110 which communicate through a network 140 with a search engine server 120 to obtain information associated with a search query (which may be a structured query or a natural language query). While three client computers 110 are illustrated, any number of client computers may be in use. The client computers 110 may be geographically distributed across the network 140, collocated, or any combination thereof. While a single search engine server 120 is illustrated, it should be appreciated that the functionality of the server 120 may be distributed over multiple servers 120 which may be clustered,

geographically distributed across the network 140, or any combination thereof.

[0033] A search query may be transmitted from a client computer 110 through the network 140 to the search engine server 120. The search engine server processes the natural language or structured query and searches through its index based upon syntax and/or semantics extracted from the query. Results of the search are returned from the server 120 through the network 140 to the client computer 110.

[0034] One or more search indexes (or indices) may be stored at, or in association with, the server 120. Information in a search index may be population from a set of source information, or a document collection. For example, in a web search implementation, content may be collected and indexed from various web sites on various web servers 130 across the network 140. Such collection and indexing may be performed by software executing on the server 120 or on another computer. The collection may be performed by web crawlers or spider applications. The collected data is also ingested into a content analytics module and a semantic annotator which process the ingested information such that semantic content collected from the corpus may be indexed (e.g., at indexing 245, as shown in FIG. 3) based on both syntax and the semantics extracted by the content analytics module.

[0035] The client computers 110 may act as terminals, hypertext browser clients, graphical display clients, or other networked clients to the server 120. For example, a web browser application at the client computers 110 may support interfacing with a web server application at the server 120. Such a browser may use controls, plug-ins, or applets to support interfacing to the server 120. The client computers 110 can also use other customized programs, applications, or modules to interface with the server 120. The client computers

110 can be desktop computers, laptops, handhelds, mobile devices, mobile telephones, television set-top boxes, kiosks, servers, terminals, thin-clients, or any other computerized devices.

[0036] The network 140 may be any communications network capable of supporting communications between the client computers 110 and the server 120. The network 140 may be wired, wireless, optical, radio, packet switched, circuit switched, or any combination thereof. The network 140 may use any topology, and links of the network 140 may support any networking technology, protocol, or bandwidth such as Ethernet, DSL, cable modem, ATM, SONET, MPLS, PSTN, POTS modem, PONS, HFC, satellite, ISDN, WiFi, WiMax, mobile cellular, any combination thereof, or any other data interconnection or networking mechanism. The network 140 may be an intranet, the Internet (or the World Wide Web), a LAN, WAN, MAN, or any other network for interconnecting computers. To support high volume and load, a distributed computing environment may be implemented by using networking technologies that may include, but are not limited to, TCP/IP, RPC, RMI, HTTP, Web Services (XML-RPC, JAX-RPC, SOAP, etc.).

[0037] It should be appreciated that, in addition to the illustrated network environment, the server 120 and client computer 110 may be combined into a single computing device. Such a combined computing device can support search indexes stored locally or remotely.

[0038] Turning now to FIG. 2, an exemplary computer (also referred to herein as a computing device) can execute software components and modules for creating semantically enriched content and/or for semantic search and navigation. The computer architecture shown in FIG. 2 illustrates a conventional desktop, laptop, or server computer and may be

utilized to execute aspects of the software components presented herein. It should be appreciated, however, that the described software components can also be executed on other example computing devices such as mobile devices, handheld wireless communications devices, smart phones, tablets, set-top boxes, kiosks, etc. Any one or more of the client computers 110 or server computer 120 may be implemented as the computer system illustrated in FIG. 2.

[0039] The computer illustrated in FIG. 2 can include a central processing unit 10 (CPU), also referred to herein as a processor or microprocessor, a system memory 13, including a random access memory 14 (RAM) and a read-only memory 16 (ROM), and a system bus 11 that can couple the system memory 13 to the CPU 10. A basic input/output system (BIOS) containing the basic routines that help to transfer information between elements within the computer, such as during start-up, can be stored in the ROM 16. The computer may further include a mass storage device 15 for storing an operating system 18, software, data, and various program modules. A semantic index 250 associated with the search engine may be stored within the mass storage device 15.

[0040] The mass storage device 15 can be connected to the CPU 10 through a mass storage controller (not illustrated) which is connected to the bus 11. The mass storage device 15 and its associated computer-readable media can provide non-volatile storage for the computer. Although the description of computer-readable media contained herein refers to a mass storage device, such as a hard disk, solid-state drive or CD-ROM drive, it should be appreciated by those skilled in the art that computer-readable media can be any available computer storage media that can be accessed by the computer.

[0041] By way of example, and not limitation, computer-readable media may include volatile and non-volatile, removable and non-removable media implemented in any method or technology for storage of information such as computer-readable instructions, data structures, program modules or other data. For example, computer-readable media includes, but is not limited to, RAM, ROM, EPROM, EEPROM, flash memory or other solid state memory technology, CD-ROM, DVD, or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store the desired information and which can be accessed by the computer.

[0042] According to various embodiments, the computer may operate in a networked environment using logical connections to remote computers through the network 140 such as the Internet. The computer may connect to the network 140 through a network interface 19 which may be a communications port such as a DSL or cable modem, Ethernet connection, or any suitable wired or wireless switch or router, etc. The network interface 19 connected to the bus 11. It should be appreciated that the network interface unit 19 may also be utilized to connect to other types of networks and remote computer systems. The computer may also include an input/output controller 12 for receiving and processing input from a number of other devices, including a keyboard, mouse, or electronic stylus (not illustrated). Similarly, an input/output controller 12 may provide output to a video display, a printer, or other type of output device (also not illustrated).

[0043] A number of program modules and data files may be stored in the mass storage device 15 and RAM 14 of the computer, including an operating system 18 suitable for controlling the operation of a networked desktop, laptop,

server computer, or other computing environment. The mass storage device 15, ROM 16, and RAM 14 may also store one or more program modules. In particular, the mass storage device 15, the ROM 16, and the RAM 14 may store the search engine 125 for execution by the CPU 10. The search engine 125 can include software components for implementing portions of the processes disclosed herein. The mass storage device 15, the ROM 16, and the RAM 14 may also store other types of program modules. The mass storage device 15, the ROM 16, and the RAM 14 can also store a semantic index 250 associated with the search engine.

[0044] Referring now to FIG. 3, a functional block diagram illustrates various components for semantic content enrichment and semantic navigation according to various embodiments of the invention. As discussed above, the search engine 125 enables information searches. In order to support such searches, a content ingestion (acquisition) process is performed by an ingestion broker 200. This ingestion broker is responsible to read the source document and normalize the input format for a uniform storage into the indexes. The ingestion broker may optionally include a language detector to identify the language used in a piece of content and apply the appropriate linguistic processing. The ingestion broker may include an ingestion queue (document queue) with a logging module (for generating logs) and a document/batch prioritization module (where documents in the queue are pending, processed, failed, rejected, or skipped). Operations related to content acquisition/ingestion extract information from documents provided as text content. This may include extracting keywords from full-text parsing and/or from meta-data. The ingestion broker may include configurable and extensible ingestion pipeline with parsers for packages (zip, tar, etc.), transformers for documents (transforming to any arbitrary format to XML), processors for fields (date,

encoding, etc.). As depicted in FIG. 3, a normalization process 210 is performed to normalize content from a variety of different sources and formats, e.g. crawled web pages, feeds, blogs, MS Office documents, PDF's, etc. By generating clean XML, the enriched content may be delivered across multiple digital channels for a variety of end-user applications (e.g. mobile, port to print, syndication, etc.).

[0045] As depicted in FIG. 3, a content analytics module 225 analyzes the content using knowledge engineering and machine learning techniques and may perform one or more of the following functions: extraction of key concepts (e.g. identifying part-of-speech patterns using natural language processing algorithms), extraction of entities (e.g. places, people, organizations, products), categorization (identification of main topics) and classification, sentiment analysis (including subjectivity and tonality), summarization, similarity analysis, automatic language detection, disambiguation, and derivation of additional meaning/knowledge from the words of the text.

[0046] The content is then semantically enriched by enrichment process 240 (e.g. with a semantic annotator) by encapsulating new annotations, metadata and the like to the content. In the search and discovery experience provided by the presentation layer of the invention, these new metadata are then instantiated as facets to let the end-user navigate and refine the search criteria. Content enriched with semantic metadata makes the content vastly more findable and search-engine optimized (SEO). This semantic data can be stored in a semantic index 250 that can be used for searching. The semantic index may be part of a search index, e.g. a forward index or an inverted index. The semantic index 250 may be used in conjunction with a meta search engine (which relies on the indices of other search engines and thus does not store

its own local index) or a cache-based search engine (which permanently stores the index along with the corpus). In one embodiment, the content analytics module 225 and annotator 240 may be consolidated in a text mining engine that semantically analyzes, annotates and relates data content.

[0047] Operations related to a user search are also depicted in FIG. 3. A user-input search query (e.g. a web search) entered into a search engine can take the form of a natural language question, one or more keyword(s) with implicit or explicit Boolean logic or a structural query. The search engine analyzes the user input/query and compares the query term(s) with information stored within the search index (or semantic index 250). The content and structuring of information in the semantic index 250 can support rapid matching and retrieval of documents, or portions of documents, that are relevant to the meaning of the query.

[0048] The ingestion broker may ingest a wide variety of different types of documents, e.g. web pages, textual documents, scanned documents, databases, information listings, other Internet content, or any other information source to provide a corpus of information to be searched. Normalization and content analytics may involve syntactic parsing and semantic mapping. Preliminary language processing steps may occur before, or at the beginning of parsing. For example, the text content may be separated at paragraph and/or sentence boundaries. Proper nouns may be identified as candidates for the names of particular people, places, objects or events. Also, the grammatical properties of meaningful word endings may be determined. Parsing may be performed by a shallow syntactic analysis system. The parser can convert sentences to representations that make explicit the syntactic relations among words. The parser can apply a grammar associated with the specific language in use. For example, the parser can

apply a grammar for English. The grammar is formalized as a lexical functional grammar (LFG) and a Probabilistic Context-free Grammar (PCFG). The grammar can specify possible ways for constructing meaningful key phrases in a given language. The parser may apply the rules of the grammar to the strings of the text content.

[0049] A grammar may be provided for various languages. For example, LFG grammars have been created for different languages such as English, French, Spanish, German and Portuguese. Other grammars may be provided as well. A grammar is developed by manual acquisition when grammatical rules are defined by a linguist or a dictionary writer. Alternatively, machine learning acquisition can involve the automated observation and analysis of many examples of text from a large corpus to automatically derive grammatical rules. A combination of manual definition and machine learning can also be used in acquiring the rules of a grammar.

[0050] During a semantic mapping process, information can be extracted from the shallow syntactic structures and combined with information about the meanings of the words in the sentence. A semantic map or semantic representation of a sentence can be provided as content semantics. Semantic mapping can augment the syntactic relationships provided by the parser with conceptual properties of individual words. The results can be transformed into representations of the meaning of sentences from the text content. Semantic mapping can determine roles played by words in a sentence. For example, the subject performing an action, something used to carry out the action, or something being affected by the action. For the purposes of indexing (e.g., indexing 245), words can be stored in a semantic index 250 along with their roles. Thus, retrieval from the semantic index 250 can depend not merely on a word in isolation, but also on the meaning of the word in the sentences in which it

appears within the text content. Semantic mapping can support disambiguation of terms, determination of antecedent relationships, and expansion of terms by synonym, hypernym, or hyponym.

[0051] An index can represent a large corpus of information so that the locations of words and phrases can be rapidly identified within the index. A traditional search engine may use keywords as search terms such that the index maps from keywords specified by a user to articles or documents where those keywords appear. The semantic index 250 can represent the semantic meanings of words in addition to the words themselves. Semantic relationships can be assigned to words during both content acquisition by ingestion broker 200 and user search. Queries against the semantic index 250 can be based on not only words, but words in specific roles, with specific properties or relationships. The roles are those played by the word in the sentence or phrase as stored in the semantic index 250. The semantic index 250 can be considered an inverted index that is a rapidly searchable database whose entries are semantic words (i.e. word in a given role) with pointers to the documents, or web pages, on which those words occur. The semantic index 250 can support hybrid indexing (e.g., indexing 245 according to some embodiments). Such hybrid indexing can combine features and functions of both keyword indexing and semantic indexing.

[0052] User entry of queries (input of search query 260) can be supported in the form of keywords or structural queries. Optionally, the query can be processed for semantics. The semantic mapping of the query can also provide semantics to be used in a retrieval process 280 against the semantic index 250. The retrieval process 280 can support hybrid index queries where both keyword index retrieval and semantic index retrieval may be provided alone or in combination.

[0053] In response to a search query, results of retrieval 280 from the semantic index 250 can inform a ranking process 285. Ranking can leverage both keyword and semantic information. During ranking 285, the results obtained by retrieval 280 can be ordered by various metrics in an attempt to place the most desirable results closer to the top of the retrieved information to be provided to the user as a search result presentation 290. The results may be presented as faceted search results, as illustrated by way of example in FIG. 4-6. Faceted search results made by presented by the search engine client application on the client computer 110.

[0054] FIG. 4 depicts an improved search engine interface with faceted search results in accordance with one embodiment of the present invention. In addition to the standard search query field and search result list, the search engine interface depicted by way of example in FIG. 4 displays a plurality of facets. In this example, the facets are presented along the left side of the screen. For example, the facets depicted are Themes, People, Organizations, Geographic Locations and Types of Documents. The layout, number, and order of the facets and the specific names of the facets are presented solely to illustrate the concept. Other layouts, numbers, orders or names of facets may of course be dynamically displayed.

[0055] Under each facet are a plurality of subjects, topics or names (i.e. specific names of people, specific place names, and specific names of organizations, etc.). These are referred to herein as faceted search results. In other words, each facet triggers its own faceted search results. The number of faceted search results that are displayed for each facet is of course variable. An optional link, virtual button or other such user interface element may be provided to enable the user to view additional faceted search results. In the

example depicted in FIG. 4, a link entitled “More” is provided to enable the user to cause the device to display more results.

[0056] As depicted in the example presented in FIG. 4, faceted search results are presented in the search engine interface in response to a query for the professional hockey team “Montreal Canadiens”. The facets (themes, organizations, locations, and document types) each can trigger faceted search results if the user adds them as criteria to navigate, refine or filter the current result set. For themes (or topics), there are, for example, different types of sports. For organizations, there are, for example, different players or coaches associated with the team. For locations, there are, for example, a city (e.g. Montreal) and a province (e.g. Quebec). For document types, there are, for example, articles and blogs.

[0057] Using knowledge engineering techniques, each single facet can be dynamically filtered at indexing time (e.g., enrichment with semantics and indexing 240-245 shown in FIG. 3 according to some embodiments) or even at query time (e.g., semantic search and navigation 260-290 shown in FIG. 3 according to some embodiments). Both filtering approaches improve overall precision of the facets and search results. However, dynamically filtering at query time provides an added benefit that content does not need to be re-classified.

[0058] The user may click, touch or otherwise select any one of the faceted search results (the listed subjects, topics, names, etc. under each facet) to obtain content relevant to that particular subject, topic or name.

[0059] Beside each faceted search result, as shown by way of example in FIG. 4, is a user interface element for including that faceted search result as a search term in a subsequent search (for example, it may use a filter query or optionally perform a Boolean logical AND with the original search query

although the filter query is generally considered to be better since it does not impact the ordering when sorting by relevancy). Also displayed in association with each faceted search result is a user interface element for excluding that faceted search result as a search term (e.g. by using a filter query that does not affect ordering based on relevancy or optionally performing a Boolean logical AND NOT with the original search query). In the example depicted in FIG. 4, the user interface for including a faceted search result as a search term in a further search is a virtual button marked with a plus sign whereas the user interface element for excluding a faceted search result as a search term in a further search is a virtual button marked with a minus sign.

[0060] In other words, in the example depicted in FIG. 4, each faceted search result under each facet has its own pair of user interface elements to include or exclude that particular faceted search result in a further search. Beside each category in each facet is a user interface element that enables the user to select to include or exclude that facet from the search results. A first user interface element may be a virtual button with a first symbol such as a plus sign (+) that enables the user to include that element (faceted search result) in a subsequent search. A second user interface element may be a virtual button with a second symbol such as a minus (-) sign to enable the user to exclude that element (faceted search result) from the subsequent search.

[0061] FIG. 4 is thus an example of a semantic faceted site search in which facets are presented in response to the search query (in addition to the standard list of relevancy-ranked search results). As will be elaborated below, the user interface elements associated with each of the faceted search results facilitate semantic navigation by enabling the user to refine the search query and thus the search results based on

one or more of the facets and to navigate to more relevant content. As will be also elaborated below in greater detail, the technology also provides a method and system for automatically and dynamically identifying semantic facets within unstructured portions of the content while ingesting the content into the search indices.

[0062] In other words, this novel presentation layer depicted by way of example in FIG. 4 provides a richer, more ergonomic user experience, facilitates interaction between the user and other available content, and improves the likelihood that the user will find relevant information. Semantic faceted navigation provides a novel method to explore and discover semantically related content in a large document collection. In other words, semantic faceted navigation is useful not only for search per se but also for exploration of the document collection.

[0063] Upon selection of the plus or minus signs, a new selection summary box may be displayed as shown in FIG. 5. This selection summary box provides the status of the refined search parameters, i.e. which elements are included and which are excluded. This refined search box may include a user interface element next to each search term to enable the user to deselect the search term. This may be a virtual button with an X, for example, as shown in the figure. The selection summary is updated each time a further selection is made to include or exclude a faceted search result as a search term.

[0064] FIG. 6 depicts the search engine interface after two faceted search results have been selected for inclusion and after one faceted search result has been selected for exclusion. The selection summary box shows the current status of the search query.

[0065] As depicted in FIGS. 4-6, one aspect of the technology is a method of displaying search results on a computing device. In general terms, this method entails receiving search results in response to a search query, displaying a list of search results in a search engine interface, and displaying a plurality of editorial and semantic facets in the search engine interface, each of the facets comprising faceted search results, each of the faceted search results being displayed in association with user interface elements for including or excluding the faceted search result as additional search terms to subsequently refine the search query.

[0066] Displaying search results may optionally include displaying, for one or more of the displayed search results, a list of metadata. Displayed metadata may be editorial or manually annotated metadata or it may be automatically extracted metadata. This displayed metadata may relate to concepts, entities, summaries, etc. The displayed metadata provides a quick overview as well as another way to navigate into the collection.

[0067] This technology confers a number of advantages, such as generally improving the search experience and enabling website owners and web-based content providers to improve the retention rate. In addition, this technology can be used to achieve better organic search engine optimization (SEO) and potentially greater advertising revenue.

[0068] With reference to FIG. 7, another aspect of the present invention is an inference technique that permits the indexation of inferred metadata to improve findability. For example, consider a Web page about an event in Stockholm where the term "Sweden" is not mentioned in the content. A visitor searching for "event Sweden" will thus, usually conventional content search technology, receive zero hits. With the present

technology, however, an inference is made at indexing time that the place name "Stockholm" refers to the capital of Sweden; this inference "Sweden" is thus added as derived or inferred metadata. This logic can be generalized to any kind of entity type where a parent/child relationship is involved (e.g. holonym/meronym, hyperonym/hyponym). Inferred metadata is automatically generated and automatically added as annotations to the content. This approach is comparable to automated query expansion but, in this novel implementation, the method involves semantically enriching content at indexing time instead of doing it at query time.

[0069] In one implementation, semantic pivots (which are based on co-occurrence) may be utilized as a tool to combine different types of semantic facets to further refine search criteria using semantic facet associations. Pivots may be useful for sentiment analysis and other applications.

[0070] This novel association technique leverages semantic metadata extracted from the unstructured content and makes it possible to display advanced facets where two facets are linked together. For example, by using this approach, the semantic navigation experience can provide facets to drill down the content using advanced criteria like "display only documents in which Acme Corp is mentioned in a negative context" (reciprocal relationship between a company name and sentiment analysis) or "only documents where a politician is involved in a trial" (reciprocal relationship between a "person type" and a topic). Pivoted facets offer the user a novel way to refine a search using criteria similar to what could be obtained using a natural query language. Broadly stated, this method for semantically searching content comprises displaying a first facet and a second facet as part of faceted search results in response to a search query, associating a first facet with a second facet in response to

user input and searching a semantic index based on the first and second facets by cross-referencing metadata for the content associated with the first facet with metadata for the content associated with the second facet. For example, an ingested document describing a politician "John Doe" is inferred to be a document about a "politician" even if the phrase "politician" does not appear in the document. This semantic metadata "policitian" is added to the index to point to the ingested document about John Doe. Another ingested document may describe, for example, a court proceeding from which the semantic metadata "trial" is inferred. Associating the facets "politician" and "trial" creates a pivoted faceted search for both concepts. The search cross-references the semantic metadata from each of the different facets to make further semantically-based connections or associations. Thus, the search results for the facet "politicians" may be refined by searching the metadata of this content for anything semantically related to "trials". The faceted search may be pivoted by inverting one facet with the other facet. For example, the pivoted search would search the metadata of the content under the facet "trials" for semantic references to "politicians".

[0071] With reference to FIG. 8, another aspect of the present invention is a method for dynamically generating topic pages based on semantic connections built from semantic annotations. In other words, topic pages may be created by inferring or deriving semantic metadata and then performing a similarity analysis on the semantic metadata to thereby associate and aggregate similar content for the topic page. This method thus involves providing semantic annotations for a plurality of content presented on a web page, determining semantic connections from the semantic annotations, and then dynamically generating topic pages about a topic, a celebrity

(person attribute), a company or a product for example. Using this technique, a topic page does not need to be populated manually with content. Based on the existing semantic annotations and connections stored in the index, the system is able to generate such pages on demand. Editorial placement rules can also be applied to topic pages in order to promote specific assets based on the user query and/or the criteria used to generate the topic page.

[0072] An exemplary implementation of a semantic navigation system is depicted schematically in FIG. 9. As depicted by way of example in FIG. 9, the semantic navigation system includes a number of modules or components. As illustrated in this schematic figure, the system includes a built-in web crawler (which may include be preconfigured with ready-to-use crawling sample configurations). The system depicted in FIG. 9 also includes an ingestion broker, an index engine, a content analytics module, and a presentation module having presentation logic, themes and a widget server. The system interacts with client systems and consumers of the user experience as depicted schematically in FIG. 9.

[0073] The ingestion broker provides functionalities to ingest, parse, process, normalize and store documents into the index. In one embodiment, the ingestion broker supports extraction of textual content from static or binary files (like PDF and Microsoft Office documents) in the ingestion queue, and also supports chaining of transformers and processors. During the ingestion process, the engine can also call a content analytics module or any third-party text annotator in order to enrich documents with semantic annotations before they are indexed. The content analytics module may be configured to supply several types of semantic annotations for documents, provide keywords and normalized facets which are then stored into the indices operated by the

(index) engine. In another implementation, another (e.g. third-party) semantic annotator may be used instead of the content analytics module.

[0074] As depicted by way of example in FIG. 9, the system includes the core engine which operates the search indexes. The core engine may include one or more query-time handlers. These handlers may perform query dispatching. These handlers may optionally enable editorial and advertisement (forced) placements, facet cleanup, and other features such as semantic similarity analysis (adapted from the "more-like-this" algorithm). The engine may have preconfigured (ready-to-use) schemas and configurations for indexes in various languages such as, for example, English, French, Spanish, German and Portuguese. In one embodiment, facet attributes may be stored in the index as facets or sub-facets (for pivoted faceting). In one implementation, the core engine is tightly integrated with the content analytics. Configurable content analytic profiles (round-robin and calls) may be provided. Optionally, file-system persistence of semantic metadata enables fast re-indexing.

[0075] In a faceted search implementation, each document is associated with a list of facets (i.e. categories that it belongs to). Those facets (categories) are stored as document attributes within the search index. During the search for a specific query, the facets of all matched documents are retrieved using the index. The retrieved facets may then be used by the searcher to narrow his search to a specific facet.

[0076] A presentation module has presentation logic for implementing a presentation framework/layer that is responsible for the user interface (UI) rendering. The presentation logic uses themes, templates and widgets which

may be configured to control how the content is rendered for viewing by end users.

[0077] As illustrated in FIG. 9, the presentation module may include preconfigured (ready-to-use) themes (which may be built with a template-based system) and a library for creating different presentations and layouts, e.g. for faceted search results, topic pages, mobile search, etc. A themes folder may be provided. The themes folder contains most of the UI, CSS and images contained in the rendering of the output, such as generic themes. In one embodiment, all themes share common files which are located in a common folder. Those files can be reconfigured or modified to change the themes.

[0078] As illustrated in FIG. 9, the presentation module also includes a widget server to serve consumable content fragments as non-JavaScript widgets (e.g. search results, facet compilations, trends, etc.). The fragments can be in any format (XML, RDF, JSON, HTML) and are extensible via templates. A JavaScript library is also available to enable quick and seamless inclusion of content widgets into external web pages. Widgets are provided for semantic facets, similar documents, recommended links, top searches, top-documents, did-you-mean, etc. and may be aggregated into a theme to render a specific layout and fulfill specific business needs. In one embodiment, a pivoted facets widget supports faceted sentiments, trends, and hierarchical faceting.

[0079] In one embodiment, the presentation framework includes a configurable logger which compiles a log of events in the presentation framework.

[0080] The semantic navigation system of FIG. 9 may include an optional management console (not illustrated) which provides various administrative control features through a Web

application interface. Content expert features may be included such as slider-based algorithmic tweaking, schema customization and crawler configuration. Editorial features (forced placements configuration) may also be provided in this management console.

[0081] In one embodiment, the management console enables monitoring of the status of the ingestion broker(s), start/stop/restart their execution as well as copy configurations between different ingestion broker instances to ease configuration replication in a distributed environment.

[0082] In one embodiment, the management console enables management of the behaviour of multiple engines and their indexes in one convenient place. Management includes monitoring status, execution, create/remove indexes, schema management, configure editorial and advertisement placements, tune sort algorithm, index cleanup and optimization, configure semantic filtering and copy configurations between different engine instances.

[0083] In one embodiment, the management console enables an administrator to monitor the status and copy configurations between different presentation framework instances.

[0084] In one embodiment, the management console enables you to monitor the status and control the execution of content analytics instances.

[0085] Another aspect of the invention is a method for providing a customizable developer experience (DX) with RDF-XML widget bundles. This may involve encapsulating content fragments from different non-JavaScript widgets into one single RDF-XML response. A bundle can be customized to encapsulate different sets of content fragments.

[0086] Another aspect of the invention is a method of creating customized user interfaces using themes built from an aggregation of interconnected widgets. A widget is a fragment of (generally) structured content in any format (HTML, XML, JSON, etc.).

[0087] In one implementation, a ready-made non-JavaScript widget library (collection of widgets) is provided to quickly promote content on a site. JavaScript widgets can also be used to fetch content from external sites and thus promote content from external web sites. In the themes folder, a JavaScript snippet can be configured to fetch content and to output it on any site, in any way, in HTML format.

[0088] Any of the methods disclosed herein may be implemented in hardware, software, firmware or any combination thereof. Where implemented as software, the method steps, acts or operations may be programmed or coded as computer-readable instructions and recorded electronically, magnetically or optically on a non-transitory computer-readable medium, computer-readable memory, machine-readable memory or computer program product. In other words, the computer-readable memory or computer-readable medium comprises instructions in code which when loaded into a memory and executed on a processor of a computing device cause the computing device to perform one or more of the foregoing method(s). In a software implementation, software components and modules may be implemented using standard programming languages including, but not limited to, object-oriented languages (e.g. Java, C++, C#, Smalltalk, etc.), functional languages (e.g. ML, Lisp, Scheme, etc.), procedural languages (e.g. C, Pascal, Ada, Modula, etc.), scripting languages (e.g. Perl, Ruby, Python, JavaScript, VBScript, etc.), declarative languages (e.g. SQL, Prolog, etc.), or any other suitable programming language, version, extension or combination thereof.

[0089] A computer-readable medium can be any means that contain, store, communicate, propagate or transport the program for use by or in connection with the instruction execution system, apparatus or device. The computer-readable medium may be electronic, magnetic, optical, electromagnetic, infrared or any semiconductor system or device. For example, computer executable code to perform the methods disclosed herein may be tangibly recorded on a computer-readable medium including, but not limited to, a floppy-disk, a CD-ROM, a DVD, RAM, ROM, EPROM, Flash Memory or any suitable memory card, etc. The method may also be implemented in hardware. A hardware implementation might employ discrete logic circuits having logic gates for implementing logic functions on data signals, an application-specific integrated circuit (ASIC) having appropriate combinational logic gates, a programmable gate array (PGA), a field programmable gate array (FPGA), etc.

[0090] This invention has been described in terms of specific embodiments, implementations and configurations which are intended to be exemplary only. Persons of ordinary skill in the art will appreciate, having read this disclosure, that many obvious variations, modifications and refinements may be made without departing from the inventive concept(s) presented herein. The scope of the exclusive right sought by the Applicant(s) is therefore intended to be limited solely by the appended claims.

CLAIMS

1. A method, performed by a computing device, for annotating content, the method comprising:

automatically inferring new semantic metadata from existing semantic metadata and from content ingested by an ingestion broker of a semantic navigation system operating the computing device; and

automatically annotating the content with the new semantic metadata.

2. The method according to claim 1, wherein the new semantic metadata is automatically inferred by creating facets.

3. The method according to claim 1, wherein the content is automatically annotated by a content analytics module or a semantic annotator operatively connected to an engine operating on the computing device.

4. The method according to claim 1, further comprising:
storing the new semantic metadata in a semantic index.

5. A computer-readable medium comprising programmed instructions in code which, when loaded into a memory and executed by a processor of a computing device, causes the computing device to:

automatically infer new semantic metadata from existing semantic metadata and from content ingested by an ingestion broker of a semantic navigation system operating the computing device; and

automatically annotate the content with the new metadata.

6. The computer-readable medium of claim 5, wherein the new semantic metadata is automatically inferred by creating facets.

7. The computer-readable medium of claim 5, wherein the content is automatically annotated by a content analytics module or a semantic annotator operatively connected to an engine operating on the computing device.

8. The computer-readable medium of claim 5, wherein the programmed instructions in code when loaded into the memory and executed by the processor of the computing device further causes the computing device to:

store the new semantic metadata in a semantic index.

9. A method of semantically enriching content during content ingestion for semantic navigation during search and discovery, the method comprising:

enriching content by creating facets in an enriching stage of a plurality of stages;
and

enhancing a search and navigation experience with faceting in an enhancing stage of the plurality of stages;

wherein the enriching comprises:

determining semantic metadata from content ingested by an ingestion broker of a semantic navigation system operating on a server machine, the ingestion broker including an ingestion pipeline for processing the content from a plurality of data sources, the determining performed by a content analytics module or a semantic annotator operatively connected to an engine operating on the server machine, the determining including automatically generating a piece of derived or inferred metadata from the content;

adding the piece of derived or inferred metadata to the content by storing the piece of derived or inferred metadata as a facet for the content in the semantic index, the storing performed by the engine; and

wherein the enhancing comprises:

receiving, by the engine through a search engine interface of the engine, a search query from a user;

searching, by the engine, the semantic index for documents or portions of documents that are relevant to a meaning of the search query, the searching including analyzing the search query, comparing query terms in the search query with information stored in the semantic index, the information stored in the semantic index including the facet for the content, and instantiating metadata associated with the documents of the portions of documents as facets; and

responsive to the search query, presenting semantic facets to the user through the search engine interface, the semantic facets including a first semantic facet instantiated by the engine in response to the search query and a second semantic facet retrieved from the semantic index, each of the semantic facets capable of triggering its own faceted search results responsive to navigation or refinement of a search criterion by the user through the search engine interface,

the presenting performed by the search engine interface on a client computer communicatively connected to the server machine.

10. The method according to claim 9, wherein the content ingested by the ingestion broker from the plurality of data sources comprises text content, wherein the text content comprises documents, wherein the ingesting comprises extracting information from the documents, wherein the information comprises keywords from text associated with the document, metadata associated with the documents, or a combination thereof.

11. The method according to claim 9, wherein the ingesting comprises transforming any arbitrary format of the content from the plurality of data sources into Extensible Markup Language.

12. The method according to claim 9, wherein the content analytics module, in determining the semantic metadata from the content ingested by the ingestion broker from the plurality of data sources, performs at least one of:

- extraction of key concepts from the content;
- identification of part-of-speech patterns from the content;
- extraction of entities from the content, the entities including at least one of places, people, organizations, or products;
- categorization of the content;
- identification of main topics in the content;
- classification of the content;
- sentiment analysis of the content, the sentiment analysis including subjectivity and tonality;
- summarization of the content;
- similarity analysis of the content;
- automatic language detection of the content;
- disambiguation of the content; or
- derivation of meaning, knowledge, or both, from words of text of the content.

13. The method according to claim 9, wherein the content analytics module and the semantic annotator are embodied in a text mining engine.

14. The method according to claim 13, wherein the engine comprises the text mining engine.

15. The method according to claim 9, further comprising:

delivering the content enriched with the semantic metadata across multiple digital channels to different end-user applications.

16. The method according to claim 9, wherein the semantic index is part of a search index, wherein the retrieving comprises analysing the search query and comparing the search query with information stored within the search index or the semantic index, and wherein the results responsive to the search query comprise documents, or portions of documents, that are relevant to a meaning of the search query.

17. The method according to claim 9, further comprising:

ranking the results retrieved from the semantic index responsive to the search query, the ranking performed by a ranking process of the semantic navigation system.

18. The method according to claim 17, wherein the results responsive to the search query comprise documents, or portions of documents, that are relevant to a meaning of the search query and wherein the ranking comprises applying boosts to different types of metadata associated with the documents, or the portions of the documents.

19. A system for semantically enriching content during content ingestion for semantic navigation during search and discovery, the system comprising:

a hardware processor;

a non-transitory computer readable medium; and

stored instructions translatable by the processor for:

enriching content by creating facets in an enriching stage of a plurality of stages;

and

enhancing a search and navigation experience with faceting in an enhancing stage of the plurality of stages;

wherein the enriching comprises:

determining semantic metadata from content ingested by an ingestion broker of the system, the ingestion broker including an ingestion pipeline for processing the content from a plurality of data sources, the determining being performed by a content analytics module or a semantic annotator and including automatically generating a piece of derived or inferred metadata from the content;

adding the piece of derived or inferred metadata to the content by storing the piece of derived or inferred metadata as a facet for the content in the semantic index, the storing performed by an engine of the system; and wherein the enhancing comprises:

receiving a search query from a user through a search engine interface of the engine;

searching the semantic index for documents or portions of documents that are relevant to a meaning of the search query, the searching including analyzing the search query, comparing query terms in the search query with information stored in the semantic index, the information stored in the semantic index including the facet for the content, and instantiating metadata associated with the documents or the portions of documents as facets; and

responsive to the search query, presenting semantic facets to the user through the search engine interface, the semantic facets including a first semantic facet instantiated by the engine in response to the search query and a second semantic facet retrieved from the semantic index, each of the semantic facets capable of triggering its own faceted search results responsive to navigation or refinement of a search criterion by the user through the search engine interface, the presenting performed by the search engine interface on a client computer communicatively connected to the system.

20. The system of claim 19, wherein the content ingested by the ingestion broker from the plurality of data sources comprises text content, wherein the text content comprises documents, wherein the ingesting comprises extracting information from the documents, wherein the information comprises keywords from text associated with the document, metadata associated with the documents, or a combination thereof.

21. The system of claim 19, wherein the ingestion broker is further configured for transforming any arbitrary format of the content from the plurality of data sources into Extensible Markup Language.

22. The system of claim 19, wherein the content analytics module is further configured for, in determining the semantic metadata from the content ingested by the ingestion broker from the plurality of data sources, performing at least one of:

extraction of key concepts from the content;

identification of part-of-speech patterns from the content;

extraction of entities from the content, the entities including at least one of places, people, organizations, or products;
categorization of the content;
identification of main topics in the content;
classification of the content;
sentiment analysis of the content, the sentiment analysis including subjectivity and tonality;
summarization of the content;
similarity analysis of the content;
automatic language detection of the content;
disambiguation of the content; or
derivation of meaning, knowledge, or both, from words of text of the content.

23. The system of claim 19, wherein the content analytics module and the semantic annotator are embodied in a text mining engine.

24. The system of claim 23, wherein the engine comprises the text mining engine.

25. The system of claim 19, further comprising:
multiple digital channels configured for delivering the content enriched with the semantic metadata to different end-user applications.

26. The system of claim 19, wherein the semantic index is part of a search index, wherein the retrieval process is further configured for analyzing the search query and comparing the search query with information stored within the search index or the semantic index, and wherein the results responsive to the search query comprise documents, or portions of documents, that are relevant to a meaning of the search query.

27. The system of claim 19, further comprising:
a ranking process configured for ranking the results retrieved from the semantic index responsive to the search query.

28. The system of claim 27, wherein the results responsive to the search query comprise documents, or portions of documents, that are relevant to a meaning of the search query and wherein the ranking process is further configured for applying boosts to different types of metadata associated with the documents, or the portions of the documents.

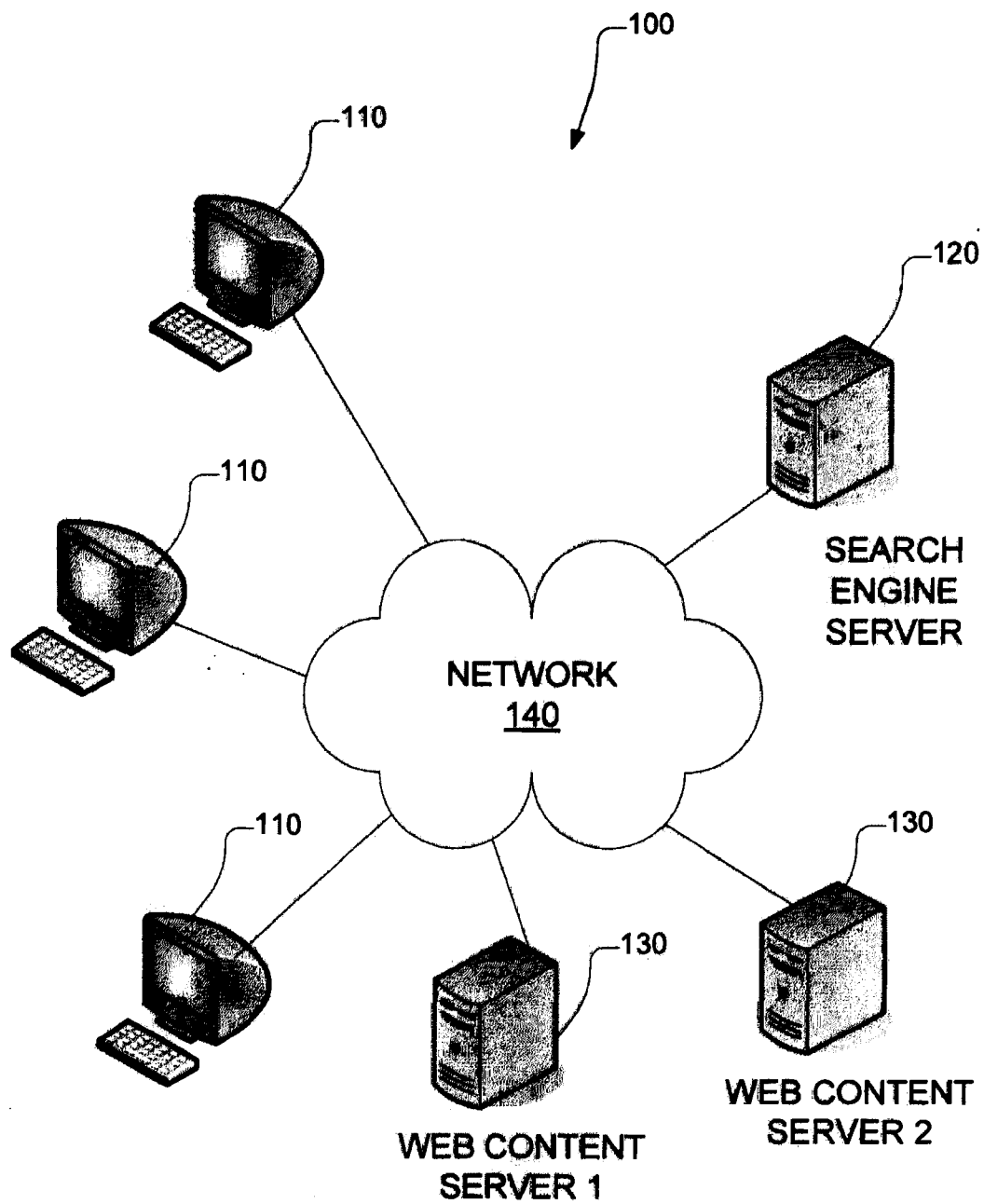


FIG. 1

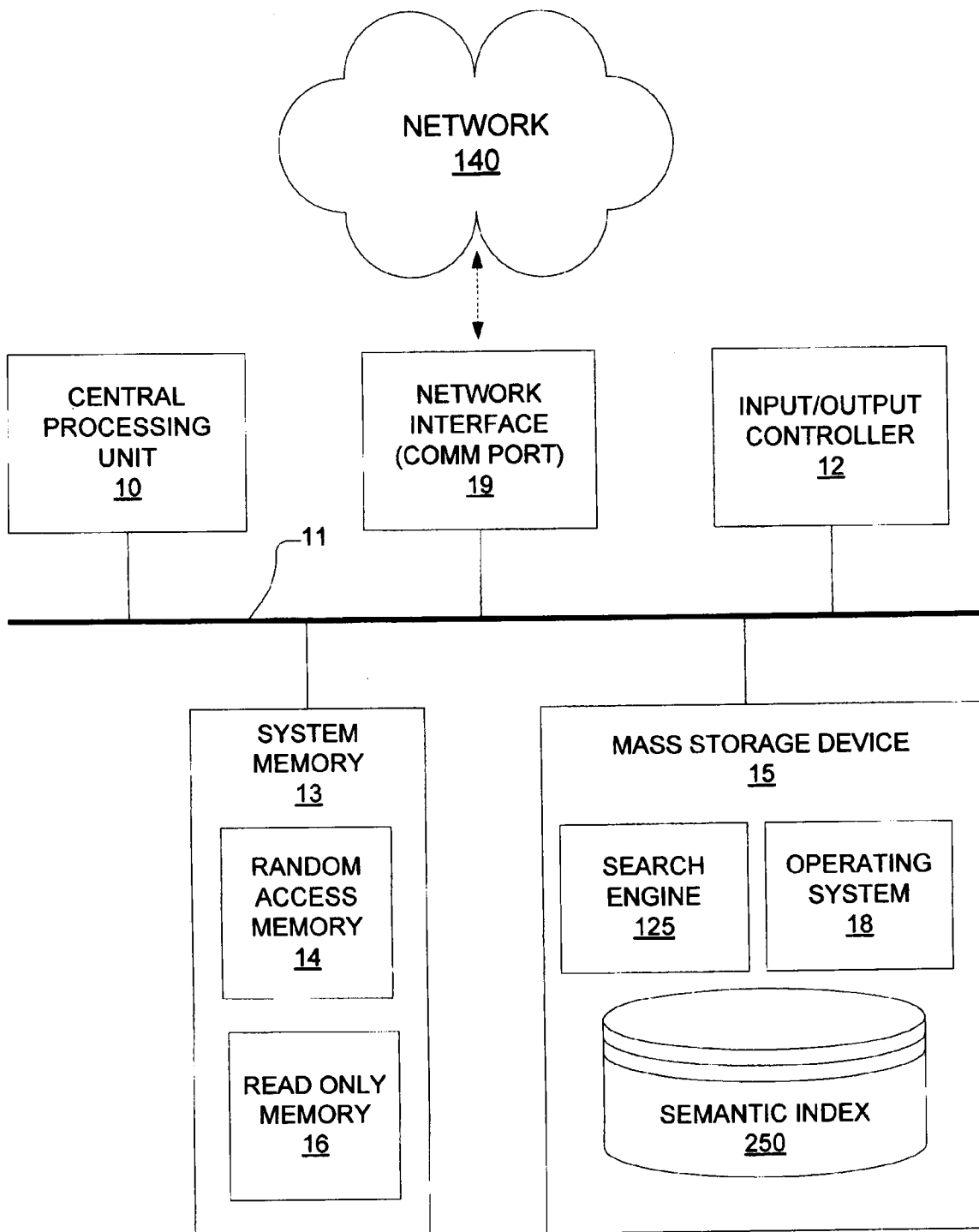


FIG. 2

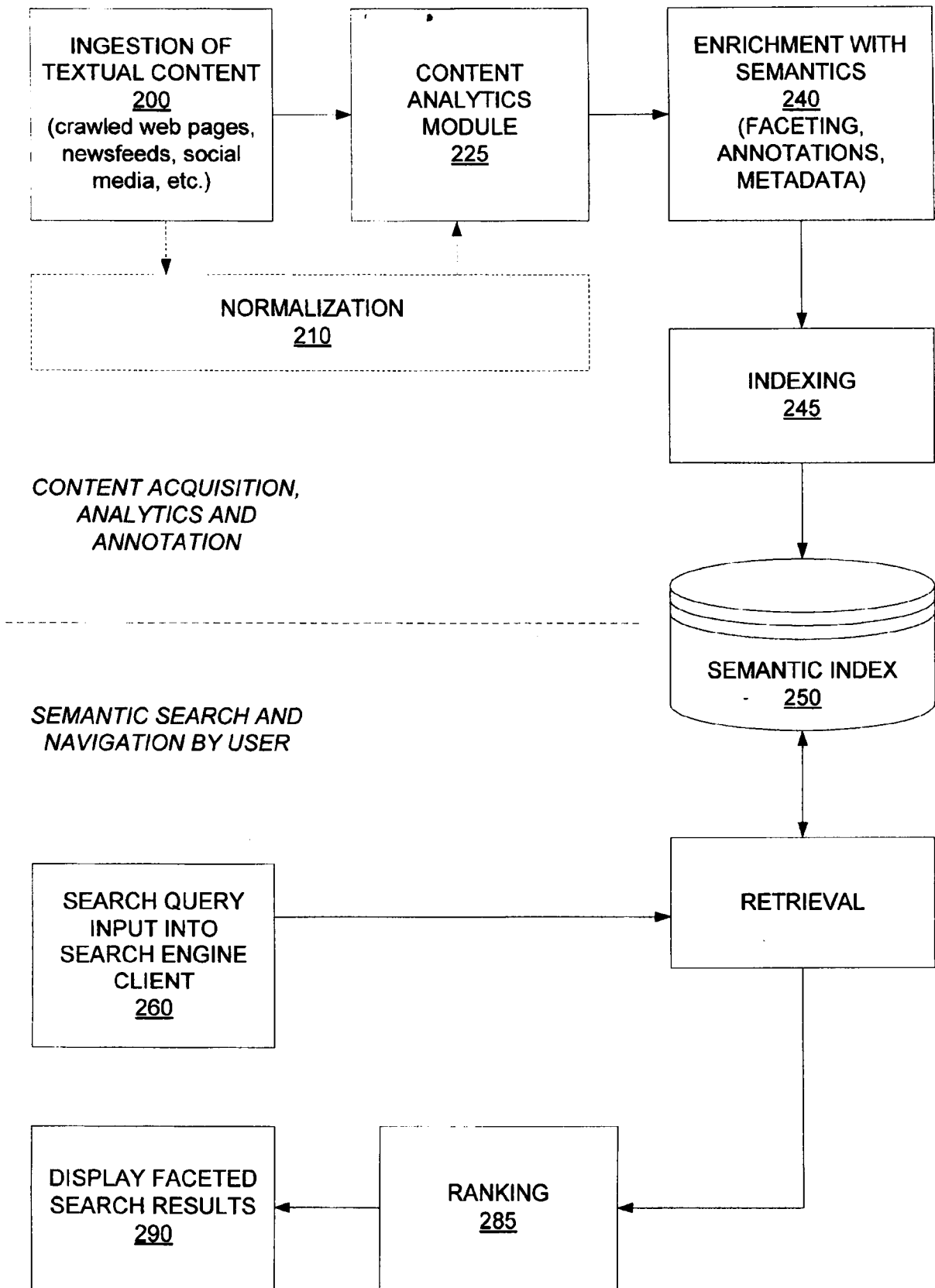


FIG. 3

Enter Key Words to Search:

"Montreal Canadiens"



Advanced Search

There are 3034 results



In the past 30 days

REFINE YOUR SEARCH ?

Themes

Hockey (2436)	☐	☐
Soccer (167)	☐	☐
Baseball (109)	☐	☐
More		

People

Carey Price (989)	☐	☐
P.K. Subban (654)	☐	☐
Jacques Martin (301)	☐	☐
More		

Organizations

Canadiens (711)	☐	☐
NHL (554)	☐	☐
Molson (80)	☐	☐
More		

Geographic Locations

Montreal (1006)	☐	☐
Quebec (987)	☐	☐
More		

Types of Documents

Articles (937)	☐	☐
Blogs (198)	☐	☐
More		

Canadiens Beat Bruins 3-2 in OT to Take Series in Five Games

Subban scores in OT as Habs clinch series 4-1.
www.hockeynews.com

Habs to Face Red Wings in Conference Finals
Game 1 of the Eastern Conference Finals will take place on Saturday night at the Bell Centre.
www.nhl.com

Habs Win to Advance
Habs win in OT to advance to next round.
www.gazette.ca

Habs Beat Bruins in Five
Montreal wins three in a row to knock out their longstanding rivals the Boston Bruins.
www.cbcsports.ca

Subban's OT goals sinks Bruins
P.K. Subban's goal at 4:31 of the first overtime period gave the Habs the win over the Bruins.
www.tsn.com

Bruins coach complains about officiating in Game 5
Boston Bruins Head Coach Claude Julien complained about the officiating after Game 5, saying that his team spent the whole night in the penalty box.
www.torontostar.ca

Habs Will Beat Red Wings in Six
If Price and Subban keep playing that well, I think the Habs will win in six games.
www.hockeyblog.com

FIG. 4

Enter Key Words to Search:

"Montreal Canadiens"

There are 121 results

Advanced Search

In the past 30 days

REFINE YOUR SEARCH ?

Your selections

Included +

Carey Price

☒

Canadiens

☒

Excluded -

P.K. Subban

☒

Themes

Hockey (99)

☒

☐

Soccer (12)

☒

☐

Baseball (5)

☒

☐

More

People

Jacques Martin (18)

☒

☐

Mike Cammalleri (15)

☒

☐

More

Organizations

NHL (80)

☒

☐

Molson (1)

☒

☐

More

Geographic Locations

Montreal (76)

☒

☐

Quebec (32)

☒

☐

More

Types of Documents

Articles (110)

☒

☐

Blogs (33)

☒

☐

More

Carey Price Lifts Canadiens in Game 5 OT Win

Carey Price stopped 42 shots as the Habs clinch their series 4-1.

www.hockeynews.com

Price Eager to Face Red Wings in Conference Finals

Habs Goalie Carey Price said he's eager for Game 1 of the Eastern Conference Finals.

www.nhl.com

Price Shuts Out Bruins as Habs Advance

Price was solid nets as the Habs win in OT to advance to next round.

www.gazette.ca

Habs Beat Bruins in Five

Montreal wins three in a row to knock out their longstanding rivals the Boston Bruins. Carey Price...

www.cbcsports.ca

Price Key to Victory Over Bruins

After P.K. Subban's goal at 4:31 of the first overtime period gave the Habs the win over the Bruins, Carey Price was asked what he thought was the turning

www.tsn.com

Bruins Admit Price was Difference

Boston Bruins Head Coach Claude Julien said goaltending made the difference. Habs goalie Carey..

www.ottawacitizen.ca

Habs Will Beat Red Wings in Six

If Price and Subban keep playing that well, I think the Habs will win in six games.

www.hockeyblog.com

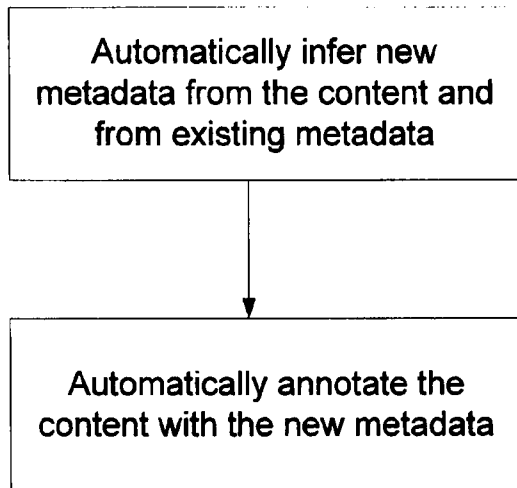


FIG. 7

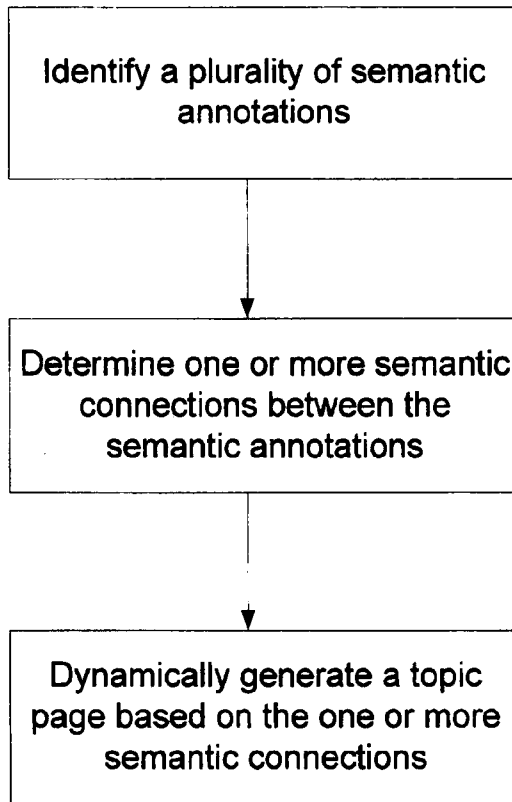


FIG. 8

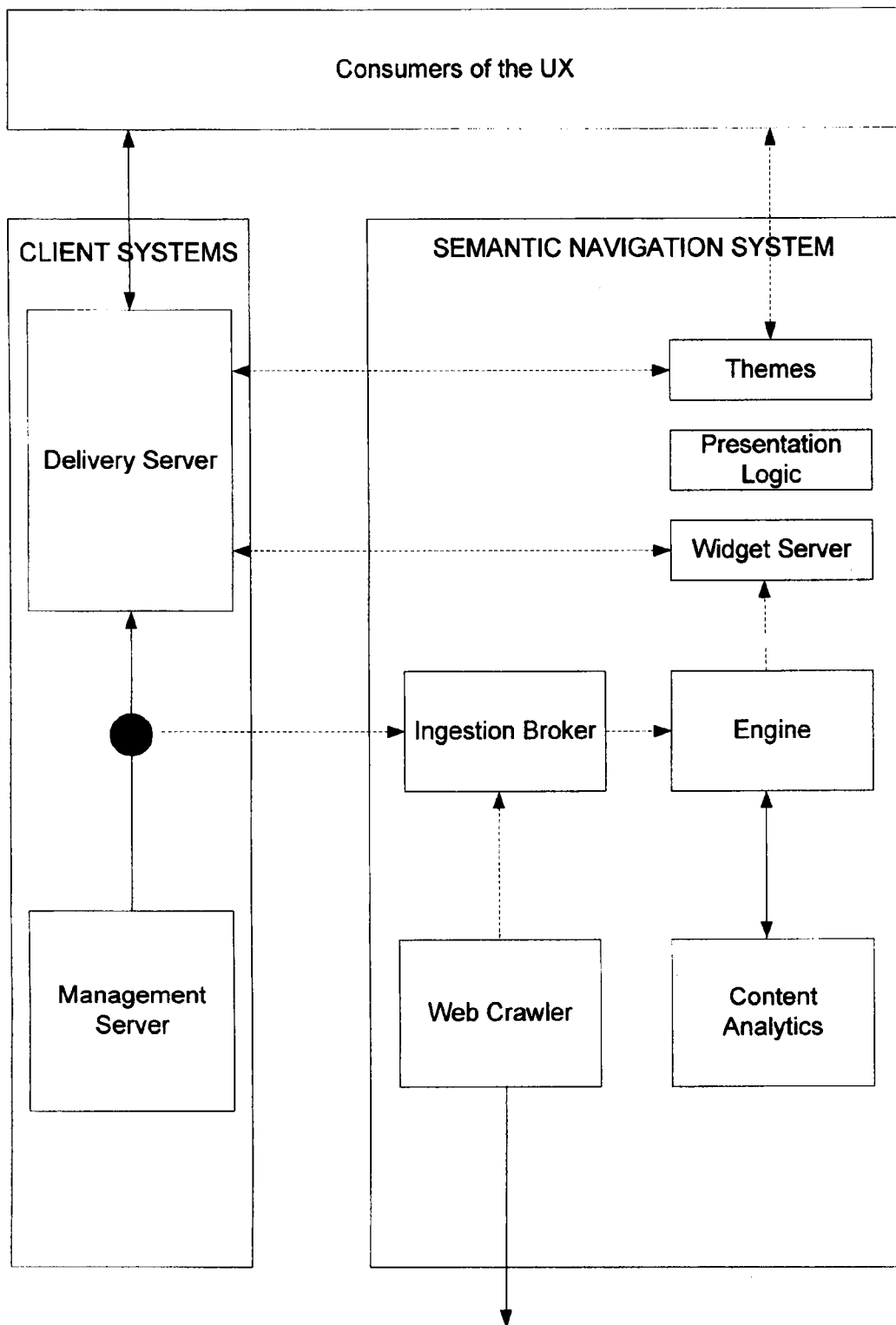


FIG. 9

