



(51) International Patent Classification:

G06N 5/00 (2006.01) G06F 17/27 (2006.01)  
G06F 17/30 (2006.01)

(21) International Application Number:

PCT/AU20 16/000096

(22) International Filing Date:

21 March 2016 (21.03.2016)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/135,560 19 March 2015 (19.03.2015) US

(71) Applicant: SEMANTIC TECHNOLOGIES PTY LTD  
[AU/AU]; Level 30, 100 Miller Street, North Sydney, New  
South Wales 2060 (AU).

(72) Inventors: TONKIN, Albert Donald; 47 Clontarf Street,  
Seaforth, New South Wales 2029 (AU). LE, Dung Xuan  
Thi; Unit 28, 14-16 Station Street, Homebush, New South  
Wales 2140 (AU).

(74) Agent: WRAYS PTY LTD; Ground Floor, 56 Ord Street,  
West Perth, Western Australia 6005 (AU).

(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, DK, DM,  
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,  
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,  
KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG,  
MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM,  
PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC,  
SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,  
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every  
kind of regional protection available): ARIPO (BW, GH,  
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,  
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,  
TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,  
DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,  
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,  
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,  
GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

(54) Title: SEMANTIC KNOWLEDGE BASE

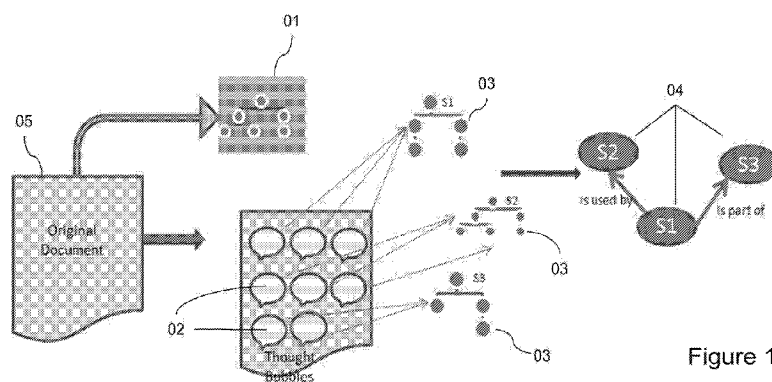


Figure 1

(57) Abstract: A system for categorising and referencing a document using an electronic processing device, wherein: the electronic processing device reviews the content of the document to identify structures within the document; wherein the identified structures are referenced against a library of structures stored in a database; wherein the document is categorised according to the conformance of the identified structures with those of the stored library of structures; and wherein the categorised structure is added to the stored library.

## SEMANTIC KNOWLEDGE BASE

### Technical Field

[0001] The present invention generally relates to a method and apparatus for extracting information from a document or set of documents into a semantically accurate knowledge base.

### Background Art

[0002] Each document, reference, patent application or patent cited in this text is expressly incorporated herein in their entirety by reference, which means that it should be read and considered by the reader as part of this text. That the document, reference, patent application, or patent cited in this text is not repeated in this text is merely for reasons of conciseness.

[0003] The rapid growth of knowledge in all areas means that it is not possible for a human being to absorb and analyse data except in increasingly narrow areas of expertise. By converting this knowledge to cognitive data, much of the analysis can be automated.

[0004] For example, IBM's Watson analysed more than 70,000 scientific articles related to *p53*, a protein that has been linked to many cancers. Automated analysis carried out by Watson helped biologists and data scientists identify six proteins that modify *p53* and that should be targeted for new research. Most important, with the use of Watson, the discovery was made in a matter of weeks, according to IBM.

[0005] More than 50 million research papers have been published, and that is doubling every three years. This rate of growth in searchable published papers is growing faster than the rate by which the data within the papers can be interpreted.

[0006] In addition to the growth in research data we have exponential growth in every other field of data, sometimes called 'Big Data'. Social communication mechanisms such as Twitter, Facebook, LinkedIn etc., generate masses of information which can be analysed. The justice system, health system, education system, manufacturing, travel, military systems and hundreds of others, all generate data at an ever increasing rate.

[0007] An objective of the present invention is to convert this information into cognitive knowledge which can be queried and analysed by computers.

[0008] This knowledge can be used by business, government, security forces and analysts of all types.

### **The Problem Space**

[0009] Traditionally data was collected in a formalised manner, by filling in values on a screen or form, for example. The data was captured in a database which was highly structured to hold exactly the data captured in this manner. The data could be analysed by a person familiar with the structure of that data, to give information in the form of reports. The data was held in the database and the knowledge of the data was accessible by the person familiar with the structure of that data. That data familiarity could be captured in a metadata repository (by the expert) thus allowing limited automated tool access to the data.

[0010] The next major advance was data which carried its own metadata in the form of xml. This created the potential for automated tools to analyse the data based upon the metadata tags carried with the data.

[0011] Assume a collection of text information which may contain information that is needed to build and use a knowledge store.

1. Is this information relevant to my potential queries?
2. Find the information which resolves my queries.
3. Which information increases my understanding of my problem?
4. What is the information in this document about?
5. Does my understanding of the problem match the concepts in the document? How does it differ?

[0012] IBM's Watson, Stanford University's DeepDive and DeepQA partially address 1 and 2, partially address 3 but do not address 4 and 5.

### **The Solution Space**

[0013] Existing methods use two techniques for determining if information is relevant to a question.

1. Understanding the information using Natural Language Processing (NLP).
2. Determining relevance statistically.

[0014] These techniques can be supplemented with Artificial Intelligence (AI) which is used to 'learn' how concepts relate. This is effectively an enhancement of the statistical technique.

- Both techniques have problems of indeterminacy of meaning.
- Neither technique is used in isolation of the other.

[0015] In both solutions a major problem is determining and transferring contextual and identity references (anaphora).

### Other Technology Solutions

#### IBM Watson

1. Is this information relevant to my queries?
  - If Q&A keywords match then hypothesise relevance.
  - Apply algorithms to test hypotheses
2. Find information which answers my queries!
  - Use Supporting Evidence Retrieval (SER)
3. Which information best answers my queries?
  - Many algorithms cleverly combined
4. What is this information about?
  - **Not a problem addressed by DeepQA or DeepDive**
5. Does my understanding of the problem match the concepts in the document?  
How does it differ?
  - **Requires** 4 and then each item in 4 is asked of the universe of 'those claims'. **Not a problem addressed by DeepQA**

[0016] IBM Watson describes a statistical solution as follows, where QA is Question Answering:

*"Early QA systems translated the natural-language question into a formal structured query against a precompiled knowledge database to obtain the answer.*

*There are two main downsides with this approach:*

- *language must be precisely and completely translated into a formal representation;*
- *underlying structured data and schema must be suitably encoded and populated.*

*This doesn't work where questions are expressed using a wide variety of linguistic expressions and span a broad range of topics.*

*Therefore Watson employs many retrieval and scoring methods based on unstructured and structured knowledge to resolve queries and answers.*

*Many of these methods do not rely on any internal representation of the natural language at all but rather employ algorithms that use unstructured sources and compare natural-language statements to each other on the basis of properties of the statements themselves".*

[0017] The IBM process consists of the following:

1. Analyse 'the question'
2. Generate candidate answers
3. Use **evidence scoring** to quantify candidates
4. Focus on 'passage scoring' i.e. the text component which contains the candidate answer keywords.
5. User SER techniques to find passages related to the candidate and provide an SER score
6. Algorithmically combine Evidence Score and SER Score to determine correctness of each hypothesis.

[0018] IBM employs the following four techniques. Each technique assigns a score:

1. Passage Term Match. By matching question terms to passage terms, regardless of grammatical relationship or word order

2. Skip-Bigram. By matching pairs of terms that are connected or nearly connected (connected skipping one other node) in the structure of the question to corresponding pairs of terms in the passage.
3. Textual Alignment. By comparing the words and word order of the passage to those of the question with the **focus** replaced by the candidate answer
4. Logical Form. On the basis of how well the structure of the question matches with that of the passage, aligning the **focus** to the candidate answer.

[0019] However, the IBM solution:

- Only addresses part of the problem space.
- Is statistically based rather than cognitive, hence may make cognitive errors.
- Non application of anaphora resolution means that **many candidates will be unrecognised.**
- Is not suitable for building a knowledge base.

## DeepDive

[0020] Stanford Universities DeepDive product uses machine learning techniques to leverage domain-specific knowledge and incorporates user feedback to improve the quality of its analysis.

[0021] DeepDive:

- computes calibrated probabilities for every assertion it makes. A fact with probability 0.9 it means the fact is 90% likely to be true;
- can extract data from millions of documents, web pages, PDFs, tables, and figures;
- allows developers to write simple rules that inform the inference (learning) process;

- is able to use the data to learn "distantly" without training for each prediction;
- is a scalable, high-performance inference and learning engine;
- uses open source tools including MADlib, Impala, a product from Oracle, and low-level techniques, such as Hogwild!.

### Google Knowledge Vault

[0022] The Google Knowledge Vault autonomously gathers and merges information from across the web into a single base of facts about the world, and the people and objects in it. As of 2014, it contained 1.6 billion facts which had been collated automatically from the internet. The difference between Google's existing Knowledge Graph and the Knowledge Vault is the way that facts are accumulated. The Knowledge Graph pulls in information from trusted sources like Freebase and Wikipedia, both of which are crowd sourced initiatives. The Knowledge Vault is an accumulation of facts from across the entire web. It is a mix of both high-confidence results and low-confidence or 'dirty' ones and machine learning is used to rank them.

[0023] The concept behind the Knowledge Vault was presented in a paper, *Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion* authored by Xin Luna Dong, Evgeniy Gabrilovich, Jeremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, Wei Zhang—all of them from Google Research.

[0024] The approach has been through various tests run by Google in other search and web products. The Official Blog Post announcing the Knowledge Graph and the transition from "Strings to Things" says that the Knowledge Graph isn't just rooted in public sources such as Freebase, Wikipedia and the CIA World Factbook. It's also augmented at a much larger scale—because we're focused on comprehensive breadth and depth".

[0025] Google has created:

- Knowledge Graph; and
- Knowledge Vault; and
- schema.org—a collection of schema for defining objects and their attributes.

[0026] Googles Knowledge Vault contains three major components:

- *Extractors*—these systems extract triples from a huge number of Web sources. Each extractor assigns a confidence score to an extracted triple, representing uncertainty about the identity of the relation and its corresponding arguments.
- *Graph-based priors*—these systems learn the prior probability of each possible triple, based on triples stored in an existing Knowledge Base (KB).
- *Knowledge fusion*—this system computes the probability of a triple being true, based on agreement between different extractors and priors.

[0027] Overall this is a system for collating and evaluating existing collated material from multiple sources. It does not have a significant semantic or cognitive component and depends heavily on confidence scoring mechanisms.

#### **Other General Products in this Space**

- CIA World Fact Book;
- FreeBase;
- Wolfram Alpha.

[0028] Reference to cited material or information contained in the text should not be understood as a concession that the material or information was part of the common general knowledge or was known in Australia or any other country.

#### **Summary of Invention**

[0029] The objective of this patent is to enable the extraction of information from a set of documents into a knowledge store, in the form of highly resolved Resource Description Framework (RDF) triples using multiple ontologies to define the knowledge store.

[0030] In this context a 'document' could be anything recording information. For example:

- a scientific research paper;
- a newspaper article;



- a medical discharge report;
- a twitter conversation;
- a computer blog;
- a video;
- an email;
- a patent.

[0031] Computers have evolved from managing data to managing information and are now trying to manage knowledge. Knowledge management requires cognitive computing and natural language processing and the existence of a comprehensive knowledge model. Ontologies provide a framework for this knowledge store.

[0032] Most techniques for extracting knowledge are based on keyword search and match techniques. As such they are really information extraction techniques, containing little or no knowledge and hence unable to accurately infer any knowledge based facts.

[0033] The technique of the present invention is based on building a knowledge model based utilising ontologies. It depends heavily on Natural Language Processing (NLP) to determine the semantics and hence the meaning. Without the ability to convert the document to semantics, cognitive processing is not possible.

[0034] The success of this tool would be dependent upon the use of an efficient semantic matcher. This tool understands context specific vocabularies, for example:

- a medical vocabulary;
- a business financial services vocabulary;
- a twitter vocabulary;
- a court reporting vocabulary;
- a medical diagnosis vocabulary;
- emails;
- etcetera.

[0035] The semantic matcher achieves understanding by transforming words, phrases and multi-words (e.g. myocardial infarction) into RDF triples. Thus the meaning of 'Prime Minister' would have a set of RDF triples such as (pseudo RDF):

- Person has role of minister;
- PM leads a political party;

- Political party forms a government;
- Government is of a nation;
- PM is elected;
- Elected by party representatives;
- Representatives elected by universal franchise.

[0036] These relationships would also be described in an ontology of government, with the specific subtype of government being the 'Westminster Government'.

[0037] In order to provide the answer to a question it is necessary to take the concepts imbedded in the question and search for matching concepts in the available media.

### **Example 1—Who is the Prime minister of Australia?**

[0038] Using a statistical, keyword based technique. Keywords 'Prime Minister' and 'Australia'.

- Find the documents which have the most hits on the keywords. This could include many millions of documents (in the order of 123,000,000 through a Google search conducted in 2015 including all 28 PMs).
- Determine statistical probability of correct answer based upon the number of keywords matched in each document. Sequence the documents based upon match quality.
- Manually browse each document to see if it provides the knowledge required.

[0039] Using our technique = Knowledge based:

- Parse the question to determine the semantics. This will resolve the entity types in question:
  - The verb 'is' implies 'who is the current Prime Minister'
  - The subject and predicate of the query 'who' and 'is' can be translated semantically to 'what is the name of
  - Prime Minister: Entity type = personal Role, which is a sub-class of Person (Person has data property name)

- Australia: Entity type = Country subclass of Nation which has a data property of 'leader of government' and 'title of leader' = 'Prime Minister'
- Check if answer already in knowledge base. If answer not available continue.
- Find the documents which have the most hits on the keywords 'Prime Minister' and 'Australia'. This search could be semantically enhanced by using the word 'current'.
- Parse the documents into RDF triples and determine entity types.
- Use knowledge of the Entity types in each document to incrementally build knowledge base:
  - Prime Minister: Entity type = personal Role, Person (Person has name)
  - Australia: Entity type = Country, Nation (Has a Prime Minister)
- Info Requested: Name of PM of Australia.
- Get exact answer or near exact answer.

[0040] The present invention seeks to obtain an understanding of the objective and contents of the documents being examined, in a manner which allows inferences to be drawn, understood, and related between documents.

[0041] In a first aspect, the present invention seeks to provide a system for categorising and referencing a document using an electronic processing device, wherein the electronic processing device reviews the content of the document to identify structures within the document;

wherein the identified structures are referenced against a library of structures stored in a database;

wherein the document is categorised according to the conformance of the identified structures with those of the stored library of structures; and

wherein the categorised structure is added to the stored library.

[0042] Preferably, elements of the identified structure are referenced through Natural Language Processing.

[0043] Preferably, when there is no conformance between the identified structure and the stored library of structures the system searches through documents in an external network for structures that match the identified structure.

[0044] Preferably, a searchable reference is associated with the document and linked to the document.

[0045] In a second aspect, the present invention seeks to provide a system for arranging text and extracting information from the text using an electronic processing device;

wherein the system scans the text and assigns it a classification; and

wherein on assigning the classification the system associates a template to the text based on the classification.

[0046] Preferably, the template separates the text into distinct sections.

[0047] Preferably, each distinct section is assigned a subject, an object and a predicate.

[0048] Preferably, the subject, object and predicate are grouped in a resource description framework.

[0049] In a third aspect, the present invention seeks to provide an electronic system for classifying the content of text wherein the system reviews the text and assigns different regions of the text a primary subject, a primary object and a primary predicate and stores the assignment as a resource description framework.

[0050] Preferably, the different regions of text are reviewed and further allocated at least one additional subject, at least one additional object and at least one additional predicate.

[0051] In a fourth aspect, the present invention seeks to provide a system for categorising and referencing the contents of a document using an electronic processing device, wherein the electronic processing device reviews the contents of the documents and identifies the expression of at least one concept within the document; and

wherein the identified expression of the at least one concept is assigned a searchable resource description framework.

[0052] Preferably, the searchable resource description framework consists of a subject, predicate and object in a resource description framework triplet.

[0053] Preferably, a plurality of resource description framework triplets are associated with the contents of the document.

[0054] Preferably, if a non-specifically identifying subject, object or predicate is used within a first resource description framework an anaphora is applied to the surrounding resource description frameworks to identify the non-specifically identifying subject, object or predicate.

[0055] Preferably, if a subordinate conjunction appears first in the resource description framework followed by a personal pronoun and then a dependent clause, the system will associate the non-specifically identifying subject, object or predicate with the corresponding subject, object or predicate in the following resource description frameworks.

[0056] In a fifth aspect, the present invention seeks to provide a method for configuring data for storage in a knowledge base, the method comprising:

- providing an electronic document having document type attributes;

- capturing the document type attributes in a template that is based on an ontology language;

- based upon the capturing by the template, defining a structure of the document determined by document type attributes to generate a defined structure as output; and

- storing the defined structure for further processing.

[0057] Preferably, the document further comprises base concepts, the further comprising based upon the capturing by the template:

- defining base concepts and generating base concepts as output; and

- storing the base concepts for further processing.

[0058] Preferably, the method further comprises:

- arranging the defined structure and base concepts into concept maps based upon the ontology language; and

- storing the concept maps in a knowledge base.

[0059] Preferably, a base concept is either explicit or implicit and when the base concept is an implicit concept, the method further comprises accessing external resources via the Internet to transform an implicit base concept into an explicit base concept.

[0060] Preferably, the method further comprises defining a sub-structure of the document based upon the sub-structure being captured by a sub-structure template.

[0061] Preferably, the method further comprises building a template to capture the document type attributes.

[0062] Preferably, the method further comprises generating a conformance report to generate validation output as to the use of a particular template with a particular document.

[0063] In a sixth aspect, the present invention seeks to provide a method of configuring data for storage in a knowledge base, the method comprising:

- providing an electronic document having contextual properties;

- expressing the contextual properties as resource description framework units representative of a subject, an object and a predicate;

- arranging the resource description framework units so that the contextual properties can be referenced between two or more resource description framework unit to generate arranged resource description framework units;

- performing anaphoric resolution between the arranged resource description framework units to generate one or more resource description framework units concepts from the resource description framework units; and

- storing the resource description framework concepts into concept maps based upon an ontology language for further processing.

[0064] Preferably, the method further comprises storing the concept maps in a knowledge base.

[0065] Preferably, a resource description framework concept is either explicit or implicit and when the resource description framework concept is an implicit resource description framework concept, the method further comprises accessing external resources via the Internet to transform a resource description framework implicit concept into a resource description framework explicit concept.

[0066] Preferably, the method further comprises merging base concepts and resource description framework concepts to form merged concepts wherein a merged concept is either explicit or implicit and when the merged concept is an implicit merged concept, the method further comprises:

accessing external resources via the Internet to transform an implicit merged concept into an explicit merged concept; and

storing explicit merged concepts in knowledge base.

[0067] Preferably, the method further comprises merging the defined structure and resource description framework concepts to form merged concepts wherein a merged concept is either explicit or implicit and when the merged concept is an implicit merged concept, the method further comprises:

accessing external resources via the Internet to transform an implicit merged concept into an explicit merged concept; and

storing explicit merged concepts in knowledge base.

[0068] Preferably, the method further comprises querying the knowledge base to retrieve an explicit merged concept.

### **Brief Description of the Drawings**

[0069] Further features of the present invention are more fully described in the following description of several non-limiting embodiments thereof. This description is included solely for the purposes of exemplifying the present invention. It should not be understood as a restriction on the broad summary, disclosure or description of the invention as set out above. The description will be made with reference to the accompanying drawings in which:

Figure 1 is a Diagram of Components used in Knowledge Extractor;

Figure 2 is a Diagram of Using external sources to add implicit detail;

Figure 3 is a Diagram of Some components of a News Report DTT shown as an ontology;

Figure 4 is a screen shot Sample of a Twitter document at [twitter.com/TEDTalks/status/5699](https://twitter.com/TEDTalks/status/5699). ..;

Figure 5 is a flow chart illustrating Extracting Knowledge through Document Type Template only;

Figure 6 is a flow chart illustrating Extracting Knowledge through Thought Bubbles Only;

Figure 7 is a flow chart illustrating Extracting Knowledge through Anaphor Resolution Only;

Figure 8 is a flow chart illustrating Extracting Knowledge through Concept Map Only; and

Figure 9 is a flow chart illustrating Extracting Knowledge using all components.

[0070] In the drawings like structures are referred to by like numerals throughout the several views. The drawings shown are not necessarily to scale, with emphasis instead generally being placed upon illustrating the principles of the present invention.

### **Description of Embodiments**

[0071] Computers have evolved from managing data to managing information and are now trying to manage knowledge. Knowledge management requires cognitive computing and natural language processing and the existence of a comprehensive knowledge model. Ontologies provide a framework for this knowledge model and RDF provides a model to store the data.

[0072] Most techniques for extracting knowledge are based on keyword search and match techniques. As such they are actually information extraction techniques, containing little or no knowledge, no cognitive understanding and hence unable to infer any knowledge based facts.

[0073] The exponential growth of knowledge in all areas means that it is not possible for a human being to absorb and analyse data except in increasingly narrow areas of expertise. By converting this knowledge to cognitive data, much of the analysis can be automated.

[0074] More than 50 million research papers have been published, and that is doubling every three years *"not only are our databases growing; they are growing faster than we can interpret all the data that they contain"*, says Dr. Olivier Lichtarge, a computational biologist and professor of molecular and human genetics at Baylor Med.



[0075] In addition to the growth in research data we have exponential growth in every other field of data, sometimes called 'Big Data'. Social communication mechanisms such as Twitter, Facebook, LinkedIn etcetera, generate masses of information which can be analysed. The justice system, health system, education system, manufacturing, travel, military systems and hundreds of others, all generate data at an ever increasing rate.

[0076] The objective of this toolset is to convert this information into cognitive knowledge which can be queried and analysed by computers.

[0077] The technique is based on building a knowledge model based utilising ontologies. It recognises that every document uses a vocabulary specific to the type of document. For example a scientific paper on diabetes, uses a different vocabulary to a newspaper report of a football game and to a hospital patient discharge report. For each document type the structure of the document can be broadly defined in terms of the roles of the subjects, the types of subjects, the types of artefacts, or events recorded. Document types always include specific components, such as an 'Abstract', a 'Conclusion', an 'Event description', an 'Author', a 'Reporter' and so forth.

[0078] The document type attributes can be captured in a Document Type Template (DTT) written in an ontology language such as Web Ontology Language (OWL) to describe the concepts and structure of documents of a particular type. These ontologies define information using semantic terminology which can be 'understood' by a computer.

[0079] Each of the concepts in the template and other concepts in the document itself can also be defined as OWL ontologies. For example a football player is a person who has the role of footballer. Ontologies for these two concepts exist and can be used as a framework for capturing details from the document in semantic terms.

[0080] As the document is read, the sentences are converted to RDF triples using Natural Language Processing (NLP), and the concepts are extracted by using subject typing as per the SSAP semantic matcher. Thus the phrase *"right winger Jones scored an own goal"* would have a completely different meaning in a football report and a political commentary. In the case of a football report Jones is a person (concept with attribute name = jones) playing football (activity concept with position = winger, objective = scoring goals, no own goals). In a political report Jones is a person (concept with attribute name = jones) supporting right wing politics (ideology with type =

right wing) who performed an act (activity concept, type unknown) which damaged his reputation (own goal).

## Technical Overview

[0081] The process and tool described in this patent is a 'Knowledge Extractor' (KnowExt) and is a stand-alone process or tool.

[0082] It uses Natural Language Processing (NLP) in a sophisticated manner.

- The NLP is typically performed sentence by sentence.
- The context of one sentence frequently is implied in following sentences allowing the resolution of references between sentences. Such references are called anaphors. Resolving these references is called 'anaphor resolution'.
  - For example, *Tom* climbed the Matterhorn. *He* used spurs (*to climb the mountain*).
  - '*He*' resolves to '*Tom*' and '*the mountain*' resolves to '*the Matterhorn*'.
- A paragraph is used to describe and elaborate upon an aspect of a single concept or event.
- Frequently context is determined by a previous paragraph or sentence or title etcetera.
  - For example, from there, our hero could see all Switzerland. '*There*' is '*the Matterhorn*' and '*Our hero*' is '*Tom*', both resolved from a previous paragraph.
- Contextual information may also be determined by the source document type (e.g. scientific paper, political article etc.).
- NLP is transformed to RDF graphs in which anaphors have been resolved.

[0083] The process can be summarised as follows:

- Identifying, manipulating and populating the concepts from a document in a machine readable manner (RDF).
- Identifying and resolving contextual references.
- Finding and organising implicit information.
- Understanding and validating relationships between concepts.
- Managing the scope of information explicitly and implicitly available.
- Producing a reusable framework to capture the information extracted (Such as Knowledge Vault).

[0084] Machine learning techniques can be applied to each of these processes. The application of Deep Learning techniques allows the use of multiple contexts to be evaluated simultaneously and the detection and minimisation of noise words. The difference between Deep learning as normally applied to NLP is that a specific deep learning context is available in terms of the various context specific templates being used (DTT, Concept Maps (CM) and Thought Bubbles (TB)).

### **Similar Concepts**

[0085] Googles Knowledge Vault is the most similar concept, however the construction methods are different and the scale and granularity are also different.

[0086] The present invention consists of a database of concepts, their meaning and relationships between them defined as ontologies. Data instances are created using the meanings captured in the ontologies.

[0087] The present invention assumes that all data has structure, even if that structure is only partially known. The structure may be explicit as in a Relational Database Management System (RDBMS), or implicit as in a scientific report.

[0088] With reference to Figures 1 to 9, the present invention solution is a cognitive solution based on conversion of text from an original document 05 to RDF using NLP. It has four components which address NLP problems through these new techniques:

- *Document Type Templates (DTT) 01*—Defines the high level data structure of the source documents 05

- *Thought Bubbles (TB) 02*—Capture the concepts in each paragraph of data and extract the entity types from the data
- *Concept Maps (CM) or Micro-Vaults (MV) 03*—Pre-constructed ontologies based upon entity types.
- *Knowledge Base (KB) 04*—The transcribed content of the document with all data related to the CMs 03 and the CMs 03 related using information from the source document.

[0089] The CM's 03 and KB 04 together create a knowledge vault which can be created from scratch and updated as new information becomes available.

[0090] Associated with each of the templates mentioned above is a set of template specific ontologies. Any Device or Event identified in these ontologies has a set of associated time components. In the case of an 'event', an exact time may be known. In the case of a future event whose time is unknown, the time may be captured as after the publication date of the article in which it was mentioned. In the case of a historic event whose time is unknown, the time may be captured as before the publication date of the article in which it was mentioned. Events may be automatically time-lined based upon information in the article.

[0091] Sometimes may be assigned by period or era, For example, 'the middle ages', 'the Jurassic' and so forth. For some objects there may be dates such as 'Blunderbuss 1600-1800', 'IPad 2000'. All objects and events are initially assigned a date of 'Unknown'. This date is progressively refined using information and inferences from the document and the web.

[0092] The time associated with an object or event has both type and value. Some illustrative types and values are shown:

- Actual time as in UTC or GMT.
- Approximate time we may make assumptions:
  - 10:00-this day;
  - Date-this century;
  - Circa.
- Relative time:
  - Before or after another event.

- Shortly before, long before. Relate to lifespan of principal class e.g. shortly before Elijah was born implies less than one lifetime of 70 years. Long before Elijah was born suggests more than one lifetime of 70 years.
- Geological time.
- Astronomical Time.
- Time anomaly-either an error or science fiction.

[0093] The relationship between the components is shown in Figure 1.

[0094] We now examine these components in detail.

## **Part 1—Document Type Templates**

[0095] A 'DTT 01 is an ontology based on the type of document source 05, for example:

- Scientific paper;
- News article;
- Discharge document;
- Pathology Report;
- Sensor input;
- Social media:
  - Twitter;
  - Facebook;
  - LinkedIn;
- Emails;
- Patents;
- Etcetera.

[0096] The DTT 01 specifies the major concepts or classes to be found in a document 05 of a particular type. Each concept is related to an initial set of data properties, also set by the type template. The list of classes could be initialised from schema.org for example, or by various methods described below.

[0097] The DTT 01 typically contains subtype templates. For example a scientific paper typically contains:

- Authors;

- Abstract;
- Findings;
- Chapters;
- Conclusion;
- Bibliography.

[0098] Many of these subtype templates would be shared between different DTTs 01.

[0099] There are axioms specific to a particular DTT 01. For example, it is axiomatic that concepts referenced in the 'Abstract' must also appear in the 'Findings' and in the 'Conclusion' of a scientific paper.

[00100] The DTT 01 is a document type specific ontology which is used to create data property instances from the input document(s). Data acquired from the document is termed 'Explicit data'.

[00101] The DTT 01 data properties may optionally be updated with data properties from other sources 07 such as DBPedia/Wikipedia. Related data acquired from outside the document is termed 'Implicit data' or 'Inferred Data'. For example the text "*Tom climbed the Matterhorn*" would result in the Object '*Matterhorn*' a proper noun being explicitly defined. 'Tom' and the 'Matterhorn' would obtain their URI from the URI of the document in which they were defined.

[00102] '*Tom*' is a proper noun which is the name of a person '*climbed the Matterhorn*' is an Event. The Event consists of the Activity '*climb*', '*Person*', '*Event*' and '*Activity*' are concepts for which a Concept Map would exist as an ontology.

[00103] The fact that the '*Matterhorn*' is a mountain in Switzerland is implicit data which could be determined by a Wikipedia search. '*Switzerland*' is a location with a massive amount of implicit data. A '*mountain*' also has a number of data properties which can be implicitly included and updated with data properties specific to the '*Matterhorn*' (height, max/min temperature, geolocation, type of rock, etc.). This implicit data would obtain its URI from its source 07, viz., Wikipedia or DBPedia.

[00104] The acquisition of implicit data is shown diagrammatically in Figure 2.

[00105] The DTT 01 forms part of a pre-constructed knowledgebase framework. It may be created by a user to resolve any new document type. It can use some subtype DTTs 01 which may be shared across DTTs 01.

[00106] Ideally the DTT 01 would be created using tool assistance and/or AI. For example, by performing deep indexing of newspaper articles it would be possible to determine that there are categories of newspaper articles such as:

- Political:
  - Local;
  - National;
  - International;
- Sport;
- Fashion;
- Education;
- Entertainment.

[00107] Each of these categories results in a sub-type of the 'Newspaper Article' template. The main subject of a political report would be:

- Politicians;
- Governments.

[00108] An illustrative DTT 01 for publications including News Reports 15 is shown in Appendix A. A small sample of some of the concepts used in a News Report 15 is shown diagrammatically in Figure 3. It shows that a News Item 17 consists of a Headline 19 and an Event 21. The Event 21 can consist of many Events 45, 47, 49 Activities 25, 27, 29, utilising objects 30, and with active 31 and passive 33 participants. It also shows that News Item 17 originated through a journalist 18 and News Organisation 20 as an organisation 41 and involved party 43. It also shows that the event 21 took place at a location 51 and involved a person 53.

[00109] The DTT 01 also determines many of the Uniform Resource Identifiers (URIs) associated with the document (see Figure 4). The URI for the document is plain. The 'Title' can be extracted from the first tweet in the conversation. Each tweet defines a paragraph or 'thought bubble' 05, 07, 09 in our terminology.

[00110] A cursory examination of Figure 4, providing an example of a twitter dialogue 55, would enable most of the components of a suitable DTT 01 to be designed. For example the title could be extracted from the first entry 23.

[001 11] A DTT conformance report can be produced to validate the use of a particular DTT 01 with a particular document. For example:

- Bibliography missing;
- Oldest and newest reference in Bibliography.

## **Part 2—Thought Bubbles**

[001 12] If a document is considered a 'body' of knowledge, then each paragraph in the document is like an 'organ' in that body. Or to use another analogy, a 'Thought Bubble' (TB) 02 is a 'molecule of information' composed of 'atoms' which correspond to RDF statements. In a Twitter conversation each 'Tweet' would be a TB 02.

[001 13] The Thought Bubble has contextual properties expressed as RDF triples. Each TB 02 or paragraph will contain a main:

- Subject (s);
- Object (p);
- Predicate (o); and
- Numerous subordinate s, p, o's.

[001 14] TB's 02 are generally arranged linearly as in the original document. IE TBs 02 can be numbered and arranged hierarchically. In more complex documents such as Twitter conversations the arrangement may resemble a network rather than a linear sequence. However, even as a network the TB's 02 are sequenced. This is important to allow references from one TB 02 to be passed to another TB 02 for anaphor resolution.

[001 15] TBs can inherit contextual properties from the previous TBs 02 through the use of anaphora analysis.

[001 16] Typically a TB 02 will correspond to a paragraph from the source document. Other sources are:

- Chapter/Section headings
- Captions
- Tweets
- Figures
- Tables



- ...

[001 17] Within a TB 02 contextual properties allow us to:

- Use NLP to perform named entity recognition, to create and reify RDF triples and produce an RDF graph of the text.
- Apply a template based on the type and subtype of the source document.
  - For example, Scientific paper, News article, Discharge report, etc.
  - Template may assume language constructs, vocabulary context etc.
- Infer the appropriate RDF subject in each ambiguous reference. For example, Tom, He, Our hero.
  - If all references are ambiguous then attempt to infer subject from previous TB or from information captured in the DTT 01
- Extract and classify each unambiguous subject and object from the TB 02 into individual, ontology based Concept Maps (CM) 03.

[001 18] The processing of a TB 02 will be specific to the type of TB 02. For example a table with headings and columns will relate the column entries to the column headings. The main s p o will be inherited from the caption, and the subordinate s,p,o's from the column headings and the column entries. Resolving these relationships is a form of anaphora.

### Anaphora

[001 19] Broadly, Anaphora is any linguistic referencing. In the context of the Knowledge Extractor (KE), see below, this referencing is extended to include various forms of document components such as tables, figures, captions etcetera.

[00120] Three types are identified:

1. Endophora-Forward referencing.
2. Cataphora-Backward referencing.

### 3. Exophora-Situational referencing.

[00121] Simple examples follow. In reality references can be separated by many sentences or paragraphs.

#### Endophora

[00122] The anaphor appears after the referenced item.

- Susan dropped the plate. **It** shattered loudly—the pronoun *it* is an anaphor; it points to the left toward its antecedent *the plate*. (Also note Object to Subject transfer).
- The music stopped, and **that** upset everyone—the demonstrative pronoun *that* is an anaphor; it points to the left toward its antecedent *The music stopped*. (Note first sentence reified).
- Fred was angry, and **so** was I—the adverb *so* is an anaphor; it points to the left toward its antecedent *angry*. (Also note Subject Object inversion).
- If Sam buys a new bike, I will do it as well—the verb phrase (or predicate) *do it* is anaphor; it points to the left toward its antecedent *buys a new bike*.

[00123] A related form of endophora is relating the contents of a column in a table to the heading of that column.

#### Cataphora

[00124] The anaphor appears before the referenced item.

- Because **he** was cold, David put on his coat—the pronoun *he* is a cataphor; it points to the right toward its postcedent *David*. (Complex reification of first phrase is required to manage 'because').
- **His** friends have been criticising Jim for exaggerating—the possessive adjective *his* is a cataphor; it points to the right toward its postcedent *Jim*.

- Although Sam might **do so**, I will not buy a new bike—the verb phrase *do so* is a cataphor; it points to the right toward its postcedent *buy a new bike*. (Complex reification of both phrases is required to manage 'although').
- In their free time, the kids play video games—the possessive adjective 'their' is a cataphor; it points to the right toward its postcedent the kids.

[001 25] A related form of cataphora is relating the caption of a figure or table to its contents.

### Exophora

[00126] Exophoric reference occurs when an expression refers to something that is not directly present in the linguistic context, but in the situational context. TBs are especially useful in resolving exophors. Deictic proforms are stereotypical exophors, for example:

- **This** hose is better than **that** one. The demonstrative adjectives *this* and *that* are exophors; they point to entities in the situational context.
- Jerry is standing over there. The adverb *there* is an exophor; it points to a location in the situational context.

### **Part 3—Concept Maps**

[001 27] A 'Concept Map' (CM) 03 is a bit like a subset of Googles Knowledge graph, created or supplemented by the source text being analysed. It is a small ontology detailing all the properties associated with a particular concept.

- Each CM 03 is a collection of mini ontologies such as Person, Event, Activity.
- CM 03 has ontology template based on subject/object type e.g. as determined by Schema.org:
  - Person;
  - Event;
  - Location;
  - Device.

- Possible relationships between named entity types are prescribed (see YAGO-1 00,000 types and their relationships). For example a person cannot 'sing' a mountain.
- CM's 03 initial set of data properties set by subject/object type template.
- Update the CM 03 with data properties from the TBs 02 (Explicit data).
- Optionally update the CM 03 with data properties from other sources 07 e.g. Dbpedia, Wikipedia (Implicit data).

[00128] This example shows the more common high level properties associated with a Person.

#### Person

- Identifiers
  - Name
  - Social Security Number
  - Date of birth
  - Father
  - Mother
- Physical Properties
  - Structure
    - Limbs
    - Organs
  - Senses
    - Sight
    - Smell
- Activities
  - Natural
    - Walk

- Eat
  - Run
- Assisted (denotes an object which would have its own CM 03)
  - Fly-Aeroplane, Kite
  - Drive-Car, boat, golf ball
  - Shoot-Weapon, camera
- Abstract
  - Work
  - Play
- Relationships
  - People
    - Marriage
    - Genealogical
    - Work
  - Objects
    - Own
    - Use

## Part 4— Knowledge Base

[00129] Introduce the concept of a 'Knowledge Bases' (KB) 04 graph.

- KB 04 is RDF graph based on the relationships between the CMs 03 identified by NLP analysis of the document. For example:
  - DTT 01 type = Scientific paper
  - DTT 01 components: Title: 'Computing Minimal Mappings', Authors: 'Jones', Abstract: '...', Conclusion: '...'
  - TBs 02 Text converted to RDF and anaphors resolved

- CM's 03 Ontologies, minimal mappings, redundant mappings populated with details from TBs.. .
- KB 04: An ontology has many mappings; some mappings are redundant; a minimal mapping has no redundant mappings; ...
- The relationships are determined by the TBs 02.
- The Object properties of the KB 04 can be validated against the CM 03 properties referenced.
  - For example, the Matterhorn is a mountain. Mountains can be climbed. Mountain climbers use crampons. This is not consistent with the data implicit in the CMs. If the sentence had read 'He used crampons', then the sentence would have been consistent. Spurs are used to climb trees, not mountains.

[001 30] After the various components (DTTs 01, CMs 03, TBs 02) have been assembled into a raw Knowledge Base 04 a number of processes are performed to refine the result. These include:

- Ambiguity reduction. For example is 'Tom' in chapter 6, 'Tom Jones' or 'Tom Brown', Are Liz and Lisa the same person?
- Duplicate reduction. Often an entity is identified as a new entity even though it has been previously identified with the same name.
- Time determination. Examine related events to more accurately determine times. This typically requires several iterations.
- Validate implicit information based upon refined data.
- Update URIs to reflect source as explicit, implicit or inferred from one or both explicit and implicit sources.
- Update provenance information.
- Load into the final version of Knowledge Base.

1. Is this information relevant to my query?
  - Do the CM's 03 map?
  - Does the KB 04 map?
  - Does the KB 04 contain information not in the query?
2. Find information which answers my query!
  - As above
3. Which information best answers my query?
  - Scoring technique based on CM's 03 and KB's 04
4. What is this information about?
  - CMs 03 and DCs
5. Do these claims match those claims?
  - As for 1.

### **What the Knowledge Extractor Does**

[00131] The major use of the Knowledge Extractor (KE) is in building a KB 04 or knowledge vault (KV). As each document is read the appropriate components of the knowledge vault are updated. The KV could be topic specific, such as diabetes, or it could cover a range of topics, or even all topics as per the Google KV.

[00132] The KE is a tool which:

- Extracts information from unstructured sources and transforms the information into knowledge in a controlled and consistent manner.
- Resolves internal references in the source documents through the use of its artefacts and processes.
- Transforms information into knowledge using the vocabulary and concepts of the subject area.
- Identifies concepts within a document and:

- Identifies the attributes of those concepts as defined in the document.
- Optionally adds additional attributes by using external sources on the www.
- Identifies relationships between the concepts identified in the document.
- Compares the knowledge identified with existing knowledge allowing the differences to be highlighted.
- Resolves language and dialect differences to determine a language neutral definition of the concepts in a document.
- Creates a comprehensive knowledge base from the documents analysed. This knowledge base can be queried and manipulated using the SPARQL query language. Using SPARQL against the knowledge base it becomes possible to ask detailed and complex queries.

[00133] As such, the KE can be used for the following types of knowledge management.

- *Research*—by collating and classifying knowledge in many research papers new knowledge can be gained.
- *Profiling*—by collating disparate forms of social media a person or organisation can be profiled. For example, information from Facebook, Twitter, LinkedIn and other social media sources can be combined and transformed into a single vocabulary. This is information of use to Call Centres, Telcos, Banks and any large customer facing organisation.
- *Forensic analysis*—by analysing all links to an event described in Twitter or Facebook etc., a comprehensive picture of an event can be obtained. See the Steubenville case.
- *Security*—detection of terrorism, illegal immigration, and all other aspects of homeland security can be improved.
- Technological advances can be evaluated by extracting the knowledge from industry journals, scientific paper, manufacturers white papers etc.



[00134] Figures 5 to Figure 9 illustrate some of the ways in which the various components may be used individually or combined to create a knowledge base.

[00135] Using the DTT 01 only will result in a consistent knowledge framework for all documents with the various concepts consistently defined and URIs assigned (see section "*Function\_1: Document Type Template (DTT) 01 Processing*" below for further explanation). Although all details of the input document would not have been analysed in full detail, the major subjects and context would be available and loaded into the KB 04.

[00136] Using TBs 04, only a complete set of RDF triples may be extracted but because there has been no anaphor resolution 09 the triples will contain many blank nodes and consequently much of the meaning will be irretrievable. However, the main subject, object and predicate of each bubble will have been identified, plus some supporting information (see section "*Function\_2: Thought Bubbles (TB) Processing*" below for a detailed explanation).

[00137] Anaphor resolution 09 applied alone would resolve anaphors within a TB 02 but would not be able to resolve anaphors across TVs 02 (see section "*Function\_3: Anaphor Resolution*" for a detailed explanation).

[00138] Referring to Figure 8, it is possible to construct a knowledge base using only the concept maps 03. The input document is parsed using an NLP to RDF transformation. Subjects and objects in the RDF statements can then be mapped to concept types and the knowledge base populated accordingly. This is shown in more detail in the section "*Function\_4: Concept Map Processing*".

[00139] Referring to Figure 9, a scenario which optimises the concurrent use of the various components is shown.

## **FunctionM : Document Type Definition (DTT) Processing**

### **Process Definition**

Input: Documents, potentially DTTs.

Method: If DTT does not exist

Manually create a new DTT

Assistant Tool (incorporates NLP)

DTT does exist (It either pre-existed, or was created)

Determine which DTT to use

Apply template

Use NLP to determine mappings to template

Determine individual Base URIs

Determine individual Base Concepts

Output: DTT, Abstract ontology (AO), Based URIs, Base Concepts

**End**

## **Function\_2: Thought Bubbles (TB) Processing**

### **Process Definition**

Input: Documents, optionally Base URIs, Base Concepts

Method: **Error Reduction:** for each document, errors such as compound sentences, incomplete sentences, etc., will be corrected.

**Extract Text:** using NLP technique:

- Within a document paragraphs will be split.
- Within a paragraph sentences will be split.
- Within a sentence it performs a tokenisation, grammar tagging and produces a Penn tree.

### **Extract RDF components**

- Within a Penn tree, we identify and derive subjects, predicates and objects.
- Using AI techniques and prepared training data, we perform a classification to extract the concepts/entities for subjects and objects.
- IF anaphor:

Perform Function\_3: Anaphor Resolution (RDF triples, Penn Tree)

END IF

- Based URI, Subjects, predicates, objects and concepts are used to identify the concept maps and construct a set of RDF triples.
- Using External resources to enrich RDF triples.

Output: RDF triples and concepts from Concept maps

**End**

**Function\_3: Anaphor Resolution****Process Definition**

Input: RDF triples, Penn Tree, External Resources

Method: IF SBAR (clause introduced by a (possibly empty) subordinating conjunction—for example, if I see Matt at work, I'll invite him over this evening (-> if)) leads the sentence and PRP appears following by THEN

Cataphora is identified. Apply backward referencing

IF PRP leads the sentence THEN

Endophora is identified. Apply forward referencing

IF WDT (determiner e.g. which, that, whatever, what, whichever) is found in the sentence THEN

Exophora is identified. Apply situation referencing

Output: Updated Subjects and objects

**End**

**Function\_4: Concept Map Processing****Process Definition**

Input: RDF triples, External Resource, concepts, Abstract Ontology

Method: Merge concepts in Abstract Ontology with concepts in Thought Bubbles

IF predefined relationship not present THEN

Query relationship from external resources using subjects and objects in RDF triples

Link concepts based on extracted relationship

Output: New knowledge base

**End**

**Variations and Modifications**

[00140] Modifications and variations such as would be apparent to the skilled addressee are considered to fall within the scope of the present invention. The present invention is not to be limited in scope by any of the specific embodiments described herein. These embodiments are intended for the purpose of exemplification only. Functionally equivalent products, formulations and methods are clearly within the scope of the invention as described herein.

[00141] Reference to positional descriptions, such as lower and upper, are to be taken in context of the embodiments depicted in the figures, and are not to be taken as limiting the invention to the literal interpretation of the term but rather as would be understood by the skilled addressee.

[00142] Throughout this specification, unless the context requires otherwise, the word "*comprise*" or variations such as "*comprises*" or "*comprising*", will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

[00143] Also, future patent applications maybe filed in Australia or overseas on the basis of, or claiming priority from, the present application. It is to be understood that the following provisional claims are provided by way of example only, and are not intended to limit the scope of what may be claimed in any such future application. Features may be added to or omitted from the provisional claims at a later date so as to further define or re-define the invention or inventions.

## APPENDIX A

### Sample Publication Ontology or DTT

The following ontology provides a sample DTT 01 for publications. The URI of this ontology is in the highlighted line starting '@base'. Every Class, Object Property and Data Property in this ontology can be referenced via this URI.

```
@prefix : <http://localhost/ontologies/2014/7/PaperNew#> .
@prefix ex: <http://localhost/ontologies/2014/7/PaperNew#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@base <http://localhost/ontologies/2014/7/PaperNew#> .
```

```
<http://localhost/ontologies/2014/7/PaperNew#> rdf:type owl:Ontology .
```

```
#####
#
#   Object Properties
#
#####
```

```
### http://localhost/ontologies/2014/7/PaperNew#Consists_of
```

```
Consists_of rdf:type owl:ObjectProperty ,
              owl:TransitiveProperty ;
```

```
    rdfs:comment "An Activity or an Event may consist of activities" ;
```

```
    rdfs:range :Activity ;
```

```
    rdfs:domain :Event .
```

```
### http://localhost/ontologies/2014/7/PaperNew#Contains
```

```
Contains rdf:type owl:ObjectProperty ,
            owl:TransitiveProperty ;
```

```
    rdfs:comment "A discharge report contains all details of the patient and their treatment associated with a particular admission" ;
```

```
    rdfs:domain :Discharge_Report ;
```

```
    rdfs:range :Medical_Advice ,
               dedication ,
               Treatment .
```

```
### http://localhost/ontologies/2014/7/PaperNew#Edits
```

```
:Edits rdf:type owl:ObjectProperty ;
```

```
    rdfs:domain :Editor ;
```

```
    rdfs:range Journal .
```

```
### http://localhost/ontologies/2014/7/PaperNew#Employs
```

```
:Employs rdf:type owl:ObjectProperty ;
```

rdfs:comment "A research Institution employs Authors" ;

rdfs:range :Authors ;

rdfs:domain :Research\_Institution .

### http://localhost/ontologies/20\_14/7/PaperNew#Gives

:Gives rdf:type owhObjectProperty ;

rdfs:comment "A Practitioner gives medical advice" ;

rdfs:range :Medical\_Advice ,  
:Medication ;

rdfs:domain Practitioner .

### http://localhost/ontologies/20\_14/7/PaperNew#Has\_Author\_Role

:Has\_Author\_Role rdf:type owhAsymmetricProperty ,  
owhObjectProperty ;

rdfs:comment "An Author will have an Author type role such as Doctor, Research Scientist  
etc."^^xsd:string ;

rdfs:domain :Author\_Type ;

rdfs:range :Authors .

### http://localhost/ontologies/20\_14/7/PaperNew#Has\_Journalist\_Role

:Has\_Journalist\_Role rdf:type owhObjectProperty ;

rdfs:comment "A journalist will have a journalist type such as 'political commentator',  
'Science writer' etc"^^xsd:string ;

rdfs:range Journalist ;

rdfs:domain :Journalist\_Type .

### http://localhost/ontologies/20\_14/7/PaperNew#OccursAt

OccursAt rdf:type owhAsymmetricProperty ,  
owhObjectProperty ;

rdfs:comment "An event occurs at a location" ;

rdfs:domain :Event ;

rdfs:range :Location ;

owhinverseOf :is\_venue\_for .

### http://localhost/ontologies/20\_14/7/PaperNew#Organises

Organises rdf:type owhAsymmetricProperty ,  
owhObjectProperty ;

```

    rdfs:comment "A person organises an event" ;

    rdfs:range :Event ;

    rdfs:domain Person .
### http://localhost/ontologies/20_14/7/PaperNew#Participates_Actively
:Participates_Actively rdf:type owhAsymmetricProperty ,
                        owhObjectProperty ;

    rdfs:comment "A participant may participate actively in an event by performing some action
which defines the event. For example, a singer, a speaker" ;

    rdfs:range :Active_Participation ;

    rdfs:domain Participant .

### http://localhost/ontologies/20_14/7/PaperNew#Participates_Passively
:Participates_Passively rdf:type owl:AsymmetricProperty ,
                        owhObjectProperty ;

    rdfs:comment "A participant may participate passively in an event by attending the event.
For example, attending a concert, observing an accident." ;

    rdfs:domain Participant ;

    rdfs:range :Passive_Participation .

### http://localhost/ontologies/20_14/7/PaperNew#Performs
Performs rdf:type owhAsymmetricProperty ,
            owhObjectProperty ;

    rdfs:comment "A medical practioner performs treatment" ;

    rdfs:domain Practitioner ;

    rdfs:range Treatment .

### http://localhost/ontologies/20_14/7/PaperNew#Produces
Produces rdf:type owhAsymmetricProperty ,
            owhObjectProperty ;

    rdfs:comment "An Practitioner Produces an article such as a Medical Discharge Report" ;

    rdfs:range :Discharge_Report ;

    rdfs:domain Practitioner .

### http://localhost/ontologies/20_14/7/PaperNew#Publishes
Publishes rdf:type owhAsymmetricProperty ,
            owhObjectProperty ;

    rdfs:comment "A news Organisation publishes a News Item" ;

    rdfs:range :News_Item ;
```

rdfs:domain :News\_Organisation .

### http://localhost/ontologies/20\_14/7/PaperNew#Sponsors

:Sponsors rdf:type owhAsymmetricProperty ,  
owhObjectProperty ;

rdfs:comment "An involved party sponsors ann event" ;

rdfs:range :Event ;

rdfs:domain :Involved\_party .

### http://localhost/ontologies/20\_14/7/PaperNew#Subject\_of

:Subject\_of rdf:type owhAsymmetricProperty ,  
owhObjectProperty ;

rdfs:comment "A Patient is the Subject of an article such as a Medical Discharge Report" ;

rdfs:range :Discharge\_Report ;

rdfs:domain Patient .

### http://localhost/ontologies/20\_14/7/PaperNew#Submits

:Submits rdf:type owhAsymmetricProperty ,  
owhObjectProperty ;

rdfs:comment "A journalist submits a News Item " ;

rdfs:domain Journalist ;

rdfs:range :News\_Item .

### http://localhost/ontologies/20\_14/7/PaperNew#SubmittedBy

:SubmittedBy rdf:type owhObjectProperty ;

rdfs:comment "A News Item is submitted by a Journalist" ;

rdfs:range Journalist ;

rdfs:domain :News\_Item ;

owhinverseOf :Submits .

### http://localhost/ontologies/20\_14/7/PaperNew#Writes

Writes rdf:type owhAsymmetricProperty ,  
owhObjectProperty ;

rdfs:comment "An Author writes an article such as a Scientific Paper or News Item" ;

rdfs:domain :Authors ;

rdfs:range :Research\_Paper .

### http://localhost/ontologies/20\_14/7/PaperNew#has



:has rdf:type owhAsymmetricProperty ,  
owhObjectProperty ;

rdfs:comment "An Article has a component such as a Title or Abstract" ;

rdfs:range :Abstract ,  
:Authors ;

rdfs:domain :Body ;

rdfs:range :Conclusion ,  
Observations ,  
:Prior\_Art ,  
References ;

rdfs:domain :Research\_Paper ;

rdfs:range :Section ,  
Title .

### http://localhost/ontologies/20\_14/7/PaperNew#hasRole

:hasRole rdf:type owhAsymmetricProperty ,  
owhObjectProperty ;

rdfs:comment "An involved party may have a role such as Owner, Leader, Journalist  
etc"^^xsd:string ;

rdfs:range :Involved\_party ;

rdfs:domain :Role .

### http://localhost/ontologies/20\_14/7/PaperNew#has\_Staff

:has\_Staff rdf:type owhAsymmetricProperty ,  
owhObjectProperty ;

rdfs:range Journalist ;

rdfs:domain :News\_Organisation .

### http://localhost/ontologies/20\_14/7/PaperNew#has\_Topic

:has\_Topic rdf:type owhAsymmetricProperty ,  
owhObjectProperty ;

rdfs:comment "A news article has a topic of headline etc" ;

rdfs:range :Event ;

rdfs:domain :News\_Item .

### http://localhost/ontologies/20\_14/7/PaperNew#has\_active\_Participants

:has\_active\_Participants rdf:type owhObjectProperty ;

rdfs:comment "An activity may have many active participants. For example a football  
team." ;

rdfs:range :Active\_Participation ;

rdfs:domain :Activity .

### http://localhost/ontologies/2014/7/PaperNew#has\_headline

:has\_headline rdf:type owhObjectProperty ;

rdfs:range :Headline ;

rdfs:domain :News\_Item .

### http://localhost/ontologies/2014/7/PaperNew#has\_passive\_participants

:has\_passive\_participants rdf:type owhObjectProperty ;

rdfs:comment "An activity may have many passive participants. For example, attendees at a football match, viewers of an activity on TV"^^xsd:string ;

rdfs:domain :Activity ;

rdfs:range :Passive\_Participation .

### http://localhost/ontologies/2014/7/PaperNew#includes

:includes rdf:type owhObjectProperty ,  
owhTransitiveProperty ;

rdfs:comment "An abstract includes description and findings" ;

rdfs:domain :Abstract ;

rdfs:range :Description ,  
:Findings .

### http://localhost/ontologies/2014/7/PaperNew#includesFigure

:includesFigure rdf:type owhObjectProperty ;

rdfs:range :Figure ;

rdfs:domain :Section .

### http://localhost/ontologies/2014/7/PaperNew#isComposedOf

:isComposedOf rdf:type owhObjectProperty ,  
owhTransitiveProperty ;

rdfs:comment "An activity can be composed of other activities"^^xsd:string ;

rdfs:range :Activity ;

rdfs:domain :Activity .

### http://localhost/ontologies/2014/7/PaperNew#isWrittenBy

:isWrittenBy rdf:type owhObjectProperty ;

```

    rdfs:comment "A scientific paper id written by Authors" ;

    rdfs:range :Authors ;

    rdfs:domain :Research_Paper .

### http://localhost/ontologies/20_14/7/PaperNew#is_involved_in

:is_involved_in rdf:type owl:AsymmetricProperty ,
    owl:ObjectProperty ;

    rdfs:comment "A non-living physical or ephemeral thing. For example, a painting, a statue, a
computer program, a patent." ;

    rdfs:range :Activity ;

    rdfs:domain :Involved_party ,
        :Location ,
        :Object .

### http://localhost/ontologies/20_14/7/PaperNew#is_organised_by

:is_organised_by rdf:type owl:ObjectProperty ;

    owl:inverseOf Organises .

### http://localhost/ontologies/20_14/7/PaperNew#is_venue_for

:is_venue_for rdf:type owl:AsymmetricProperty ,
    owl:ObjectProperty ;

    rdfs:comment "A location is the venue for an event" ;

    rdfs:range :Event ;

    rdfs:domain :Location .

#####
#
# Data properties
#
#####

### http://localhost/ontologies/20_14/7/PaperNew#ActivityID

:ActivityID rdf:type owl:DatatypeProperty ,
    owl:FunctionalProperty ;

    rdfs:comment "Unique Identifier for an activity" ;

    rdfs:domain :Activity ;

    rdfs:range xsd:positiveInteger .

### http://localhost/ontologies/20_14/7/PaperNew#Activity_Name

:Activity_Name rdf:type owl:DatatypeProperty ;

    rdfs:comment "The name given to an activity. For example, painting, skiing, sailing, listening." ;

```

rdfs:domain :Activity ;

rdfs:range xsd:string .

### http://localhost/ontologies/20\_14/7/PaperNew#ArticleID

:ArticleID rdf:type owhDatatypeProperty ,  
owhFunctionalProperty ;

rdfs:comment "Uniquely identifies an Article" ;

rdfs:domain Publication ;

rdfs:range xsd:positiveInteger .

### http://localhost/ontologies/20\_14/7/PaperNew#EventID

:EventID rdf:type owhDatatypeProperty ,  
owhFunctionalProperty ;  
rdfs:comment "Unique Identifier for an event" ;

rdfs:domain :Event ;

rdfs:range xsd:positiveInteger .

### http://localhost/ontologies/20\_14/7/PaperNew#Event\_Description\_RDF

:Event\_Description\_RDF rdf:type owhDatatypeProperty ;

rdfs:comment "Event description converted to RDF, anaphorised and reified if required." ;

rdfs:domain :Event ;

rdfs:range xsd:string .

### http://localhost/ontologies/20\_14/7/PaperNew#Event\_Description\_Text

:Event\_Description\_Text rdf:type owhDatatypeProperty ;

rdfs:comment "A textual description of the event" ;

rdfs:domain :Event ;

rdfs:range xsd:string .

### http://localhost/ontologies/20\_14/7/PaperNew#Event\_Name\_Subject

:Event\_Name\_Subject rdf:type owhDatatypeProperty ;

rdfs:comment "The subject extracted from the event title. Eg 'Day Out', 'Show', 'War' . " ;

rdfs:domain :Event ;

rdfs:range xsd:string .

### http://localhost/ontologies/20\_14/7/PaperNew#Event\_Name\_Text

:Event\_Name\_Text rdf:type owhDatatypeProperty ;

rdfs:comment "The name of an event. For example, The big day out', The Easter Show', The Iraqi War' ." ;

    rdfs:domain :Event ;

    rdfs:range xsd:string .

### http://localhost/ontologies/2014/7/PaperNew#Event\_Start\_Time

:Event\_Start\_Time rdfs:type owhDatatypeProperty ;

    rdfs:comment "The time at which an event is scheduled to start. Time may include GMT offset" ;

    rdfs:domain :Event ;

    rdfs:range xsd:dateTime .

### http://localhost/ontologies/2014/7/PaperNew#Event\_end\_time

:Event\_end\_time rdfs:type owhDatatypeProperty ;

    rdfs:comment "The time at which an event is scheduled to finish. Time may include GMT offset" ;

    rdfs:domain :Event ;

    rdfs:range xsd:dateTime .

### http://localhost/ontologies/2014/7/PaperNew#Given\_Name

:Given\_Name rdfs:type owhDatatypeProperty ;

    rdfs:comment "Name given, not inherited."^^xsd:string ;

    rdfs:domain Person ;

    rdfs:range xsd:string .

### http://localhost/ontologies/2014/7/PaperNew#HeadlineID

:HeadlineID rdfs:type owhDatatypeProperty ,  
    owhFunctionalProperty ;

    rdfs:domain :Headline ;

    rdfs:range xsd:positiveInteger .

### http://localhost/ontologies/2014/7/PaperNew#Headline\_Object

:Headline\_Object rdfs:type owhDatatypeProperty ;

    rdfs:comment "The object component of the headline RDF" ;

    rdfs:domain :Headline ;

    rdfs:range xsd:string .

### http://localhost/ontologies/2014/7/PaperNew#Headline\_Predicate

:Headline\_Predicate rdf:type owhDatatypeProperty ;

    rdfs:comment "The predicate component of the headline RDF" ;

    rdfs:domain :Headline ;

    rdfs:range xsd:string .

### http://localhost/ontologies/20\_14/7/PaperNew#Headline\_RDF

:Headline\_RDF rdf:type owhDatatypeProperty ;

    rdfs:comment "The headline text converted to RDF, reified as required." ;

    rdfs:domain :Headline ;

    rdfs:range xsd:string .

### http://localhost/ontologies/20\_14/7/PaperNew#Headline\_Subject

:Headline\_Subject rdf:type owhDatatypeProperty ;

    rdfs:comment "The subject component of the headline RDF" ;

    rdfs:domain :Headline ;

    rdfs:range xsd:string .

### http://localhost/ontologies/20\_14/7/PaperNew#Headline\_Tag

:Headline\_Tag rdf:type owhDatatypeProperty ;

    rdfs:comment "A tag associated with the headline. For example 'Iraqi War:'" ;

    rdfs:domain :Headline ;

    rdfs:range xsd:string .

### http://localhost/ontologies/2014/7/PaperNew#Headline\_Text

:Headline\_Text rdf:type owhDatatypeProperty ;

    rdfs:comment "The text of a headline not including any tags that may have been included in the te3xt" ;

    rdfs:domain :Headline ;

    rdfs:range xsd:string .

### http://localhost/ontologies/20\_14/7/PaperNew#IPName

:IPName rdf:type owhDatatypeProperty ;

    rdfs:comment "A name associated with the involved party. In the case of an organisation then it would be the company name. In the case of a person it would be the person name" ;

    rdfs:domain :Involved\_party ;

rdfs:range xsd:string .

### http://localhost/ontologies/20\_14/7/PaperNew#InvolvedPartyID

:InvolvedPartyID rdfs:type owhDatatypeProperty ,  
owhFunctionalProperty ;

rdfs:comment "Unique Identifier for an involved party" ;

rdfs:domain :Involved\_party ;

rdfs:range xsd:positiveInteger .

### http://localhost/ontologies/20\_14/7/PaperNew#Journal\_Number

:Journal\_Number rdfs:type owhDatatypeProperty ;

rdfs:comment "Publisher assigned journal number"^^xsd:string ;

rdfs:domain :Journal ;

rdfs:range xsd:integer .

### http://localhost/ontologies/20\_14/7/PaperNew#Journal\_Title

:Journal\_Title rdfs:type owhDatatypeProperty ;

rdfs:comment "Title of Journal"^^xsd:string ;

rdfs:domain Journal ;

rdfs:range xsd:string .

### http://localhost/ontologies/20\_14/7/PaperNew#Journalist\_Submission\_Date

:Journalist\_Submission\_Date rdfs:type owhDatatypeProperty ;

rdfs:comment "The date that the journalist submitted the article for  
publication"^^xsd:dateTime ;

rdfs:domain :News\_Item ;

rdfs:range xsd:dateTime .

### http://localhost/ontologies/20\_14/7/PaperNew#Keywords

:Keywords rdfs:type owhDatatypeProperty ;

rdfs:comment "Keywords associatesd with a paper"^^xsd:string ;

rdfs:domain :Research\_Paper ;

rdfs:range xsd:string .

### http://localhost/ontologies/20\_14/7/PaperNew#Latitude

:Latitude rdfs:type owhDatatypeProperty ;

rdfs:domain :Location ;

rdfs:range xsd:long .

### http://localhost/ontologies/2014/7/PaperNew#LocationID

:LocationID rdfs:type owhDatatypeProperty ,  
owhFunctionalProperty ;

rdfs:domain :Location ;

rdfs:range xsd:positiveInteger .

### http://localhost/ontologies/2014/7/PaperNew#Location\_Address

:Location\_Address rdfs:type owhDatatypeProperty ;

rdfs:comment "A location has an Address" ;

rdfs:domain :Location ;

rdfs:range xsd:string .

### http://localhost/ontologies/2014/7/PaperNew#Longitude

:Longitude rdfs:type owhDatatypeProperty ;  
rdfs:comment "The longitude of a location" ;  
rdfs:domain :Location ;  
rdfs:range xsd:long .

### http://localhost/ontologies/2014/7/PaperNew#Organisation\_Name

:Organisation\_Name rdfs:type owhDatatypeProperty ;  
rdfs:comment "The registered name of an Organisation" ;  
rdfs:domain Organisation ;  
rdfs:range xsd:string .

### http://localhost/ontologies/2014/7/PaperNew#Pages

Pages rdfs:type owhDatatypeProperty ;  
rdfs:comment "Number of pages in article"^^xsd:string ;  
rdfs:domain :Research\_Paper ;  
rdfs:range xsd:integer .

### http://localhost/ontologies/2014/7/PaperNew#Publication\_Date

:Publication\_Date rdfs:type owhDatatypeProperty ;  
rdfs:comment "The date the article was published"^^xsd:string ;  
rdfs:domain :News\_Item ;  
rdfs:range xsd:dateTime .

### http://localhost/ontologies/2014/7/PaperNew#Qualifications

:Qualifications rdfs:type owhDatatypeProperty ;



```
    rdfs:comment "University degrees and name of conferring institution."^^xsd:string ;  
    rdfs:domain :Authors ;  
    rdfs:range xsd:string .
```

```
### http://localhost/ontologies/20_14/7/PaperNew#RoleID
```

```
:RoleID rdf:type owl:DatatypeProperty ,  
         owhFunctionalProperty ;  
  
    rdfs:comment "A Unique Identifier for Role"^^xsd:positiveInteger ;  
    rdfs:domain :Role ;  
    rdfs:range xsd:positiveInteger .
```

```
### http://localhost/ontologies/20_14/7/PaperNew#Role_Name
```

```
:Role_Name rdf:type owl:DatatypeProperty ;  
  
    rdfs:comment "The name of a role which an IP may assume."^^xsd:string ;  
    rdfs:domain :Role ;  
    rdfs:range xsd:string .
```

```
### http://localhost/ontologies/20_14/7/PaperNew#ScPap_Identifier
```

```
:ScPap_Identifier rdf:type owhDatatypeProperty ,  
                    owhFunctionalProperty ;  
    rdfs:comment "Unique identifier for a scientific paper" ;  
    rdfs:domain :Research_Paper ;  
    rdfs:range xsd:positiveInteger .
```

```
### http://localhost/ontologies/20_14/7/PaperNew#Surname
```

```
:Surname rdf:type owhDatatypeProperty ;  
    rdfs:comment "Inherited name."^^xsd:string ;  
    rdfs:domain Person ;  
    rdfs:range xsd:string .
```

```
### http://localhost/ontologies/20_14/7/PaperNew#TitleID
```

```
:TitleID rdf:type owhDatatypeProperty ,  
           owhFunctionalProperty ;  
    rdfs:comment "Unique identifier for the title of an Article"^^xsd:positiveInteger ;  
    rdfs:domain Title ;  
    rdfs:range xsd:positiveInteger .
```

```
### http://localhost/ontologies/20_14/7/PaperNew#Title_Object
```

```
:Title_Object rdf:type owhDatatypeProperty ;  
    rdfs:comment "The object component of the RDF triple derived from the title."^^xsd:string ;  
    rdfs:domain Title ;  
    rdfs:range xsd:string .
```

```
### http://localhost/ontologies/20_14/7/PaperNew#Title_Predicate
```

```
:Title_Predicate rdf:type owhDatatypeProperty ;
    rdfs:comment "The predicate component of the RDF triple derived from the title."^^xsd:string ;
    rdfs:domain Title ;
    rdfs:range xsd:string .
```

```
### http://localhost/ontologies/201_4/7/PaperNew#Title_RDF
```

```
:Title_RDF rdf:type owhDatatypeProperty ;
    rdfs:comment "The RDF triple derived from the title."^^xsd:string ;
    rdfs:domain :Title ;
    rdfs:range xsd:string .
```

```
### http://localhost/ontologies/201_4/7/PaperNew#Title_Subject
```

```
:Title_Subject rdf:type owhDatatypeProperty ;
    rdfs:comment "The subject component of the RDF triple derived from the title."^^xsd:string ;
    rdfs:domain :Title ;
    rdfs:range xsd:string .
```

```
### http://localhost/ontologies/201_4/7/PaperNew#Title_Text
```

```
:Title_Text rdf:type owhDatatypeProperty ;
    rdfs:comment "The text of the title."^^xsd:string ;
    rdfs:domain Title ;
    rdfs:range xsd:string .
```

```
### http://localhost/ontologies/201_4/7/PaperNew#Volume
```

```
:Volume rdf:type owhDatatypeProperty ;
    rdfs:comment "Publisher assigned volume number"^^xsd:string ;
    rdfs:domain Journal ;
    rdfs:range xsd:integer .
```

```
#####
#
# Classes
#
#####
```

```
### http://localhost/ontologies/201_4/7/PaperNew#Abstract
```

```
:Abstract rdf:type owhClass ;
    rdfs:label "Abstract of Research Paper"^^xsd:string ;
    rdfs:comment "A comprehensive and succinct description of the article." .
```

```
### http://localhost/ontologies/201_4/7/PaperNew#Active_Participation
```

```
:Active_Participation rdf:type owhClass ;
    rdfs:comment "A participant has an active participation type in an activity if they perform  
aprt of the action. "^^xsd:string .
```

```
### http://localhost/ontologies/201_4/7/PaperNew#Activity
```

```
:Activity rdf:type owhClass ;
    rdfs:comment "Actions performed by participants in an event" .
```

```
### http://localhost/ontologies/201_4/7/PaperNew#Author_Type
```

:Author\_Type rdf:type owhClass ;  
    rdfs:subClassOf :Role ;  
    rdfs:comment "Types of authors of scientific papers. For example; Professor, Doctor, Scientist, Engineer" .

### http://localhost/ontologies/2014/7/PaperNew#Authors

:Authors rdf:type owhClass ;  
    rdfs:subClassOf Person ;  
    rdfs:comment "An Author writes an article" .

### http://localhost/ontologies/2014/7/PaperNew#Body

:Body rdf:type owhClass ;  
    rdfs:comment "The main part of the scientific paper" .

### http://localhost/ontologies/2014/7/PaperNew#Conclusion

:Conclusion rdf:type owhClass ;  
    rdfs:label "Conclusion"^^xsd:string ;  
    rdfs:comment "Described the outcome of applying the theory explored in this Article" .

### http://localhost/ontologies/2014/7/PaperNew#Content

:Content rdf:type owhClass ;  
    rdfs:comment "The bulk of an article. Not explicitly covered by any other component" .

### http://localhost/ontologies/2014/7/PaperNew#Definitions

:Definitions rdf:type owhClass ;  
    rdfs:comment "The meaning of terms defined in an article" .

### http://localhost/ontologies/2014/7/PaperNew#Description

:Description rdf:type owhClass ;  
    rdfs:comment "A textual explanation of the major subject or object of an article" .

### http://localhost/ontologies/2014/7/PaperNew#Discharge\_Report

:Discharge\_Report rdf:type owhClass ;  
    rdfs:subClassOf Publication ;  
    rdfs:comment "A structured document with some natural language components utilising medical jargon whose structure follows patient discharge report conventions" .

### http://localhost/ontologies/2014/7/PaperNew#Editor

:Editor rdf:type owhClass ;  
    rdfs:subClassOf :Role ;  
    rdfs:comment "A person who edits a journal, newspaper etc"^^xsd:string .

### http://localhost/ontologies/2014/7/PaperNew#Event

:Event rdf:type owhClass ;  
    rdfs:comment "A set of activities which have time, duration and location" .

### http://localhost/ontologies/2014/7/PaperNew#Figure

:Figure rdf:type owhClass ;  
    rdfs:comment "A graphic representation of facts in a section of a scientific paper"^^xsd:string .

### http://localhost/ontologies/2014/7/PaperNew#Findings

:Findings rdf:type owhClass ;  
rdfs:comment "The findings of a scientific paper as described in the abstract"^^xsd:string .

### http://localhost/ontologies/2014/7/PaperNew#Headline

:Headline rdf:type owl:Class ;  
rdfs:comment "A terse summary of a news item"^^xsd:string .

### http://localhost/ontologies/2014/7/PaperNew#Involved\_party

:Involved\_party rdf:type owl:Class ;  
rdfs:comment "Any person or organisation involved in an Article or Event in any way" .

### http://localhost/ontologies/2014/7/PaperNew#Journal

Journal rdf:type owhClass .

### http://localhost/ontologies/2014/7/PaperNew#Journal\_Paper

:Journal\_Paper rdf:type owhClass ;  
rdfs:subClassOf :Research\_Paper .

### http://localhost/ontologies/2014/7/PaperNew#Journalist

Journalist rdf:type owhClass ;  
rdfs:subClassOf Person ;  
rdfs:comment "An Individual who submits a News Item" .

### http://localhost/ontologies/2014/7/PaperNew#Journalist\_Type

:Journalist\_Type rdf:type owhClass ;  
rdfs:subClassOf :Role ;  
rdfs:comment "Role type fro a journalist. For example; Political, Fashion, Motoring, Financial, Editor" .

### http://localhost/ontologies/2014/7/PaperNew#Location

:Location rdf:type owhClass ;  
rdfs:comment "A geo-spatial coordinate" .

### http://localhost/ontologies/2014/7/PaperNew#Medical\_Advice

:Medical\_Advice rdf:type owhClass ;  
rdfs:comment "A written opinion pprovided by a medical practitioner"^^xsd:string .

### http://localhost/ontologies/2014/7/PaperNew#Medication

:Medication rdf:type owhClass ;  
rdfs:comment "medicines prescribed by a medical practitioner"^^xsd:string .

### http://localhost/ontologies/2014/7/PaperNew#News\_Item

:News\_Item rdf:type owhClass ;  
rdfs:subClassOf Publication ;  
rdfs:comment "A partially structured natural language document whose structure follows news item conventions" .

### http://localhost/ontologies/20\_14/7/PaperNew#News\_Organisation

:News\_Organisation rdf:type owhClass ;  
rdfs:subClassOf Organisation ;  
rdfs:comment "An organisation which collects and disseminates news via news articles" .

### http://localhost/ontologies/20\_14/7/PaperNew#Object

:Object rdf:type owhClass ;  
rdfs:label Object "^^xsd:string ;  
rdfs:comment "A physical device used in performing an activity"^^xsd:string .

### http://localhost/ontologies/20\_14/7/PaperNew#Observations

Observations rdf:type owhClass ;  
rdfs:comment "The facts used to reach the conclusion as described in the abstract of a scientific paper"^^xsd:string .

### http://localhost/ontologies/20\_14/7/PaperNew#Organisation

Organisation rdf:type owhClass ;  
rdfs:subClassOf :Involved\_party ;  
rdfs:comment "A named collection of people collected for some purpose" .

### http://localhost/ontologies/20\_14/7/PaperNew#Participant

Participant rdf:type owhClass ;  
rdfs:subClassOf Person ;  
rdfs:comment "A person who participates, actively or passively in an event" .

### http://localhost/ontologies/20\_14/7/PaperNew#Passive\_Participation

:Passive\_Participation rdf:type owhClass ;  
rdfs:comment "A participant is a paassive participant in an activity f they observe or attend or comment on an activity."^^xsd:string .

### http://localhost/ontologies/20\_14/7/PaperNew#Patient

Patient rdf:type owhClass ;  
rdfs:subClassOf Person ;  
rdfs:comment "The subject of a discharge report" .

### http://localhost/ontologies/20\_14/7/PaperNew#Person

Person rdf:type owhClass ;  
rdfs:subClassOf :Involved\_party ;  
rdfs:comment "A unique human being" .

### http://localhost/ontologies/20\_14/7/PaperNew#Practitioner

Practitioner rdf:type owhClass ;  
rdfs:subClassOf Person ;  
rdfs:comment "The author and producer of a discharge report" .

### http://localhost/ontologies/20\_14/7/PaperNew#Prior\_Art

Prior\_Art rdf:type owhClass ;  
rdfs:comment "Historic solutions or attempted solutions to the problem being analysed" .

### http://localhost/ontologies/20 14/7/PaperNew#Publication

Publication rdf:type owhClass ;  
rdfs:label "Publication"^^xsd:string ;  
rdfs:comment "Any text document is an article" .

### http://localhost/ontologies/20 14/7/PaperNew#References

References rdf:type owhClass ;  
rdfs:comment "Bibliography of papers quoted in this document" .

### http://localhost/ontologies/20 14/7/PaperNew#Research\_Institution

:Research\_Institution rdf:type owhClass ;  
rdfs:subClassOf Organisation ;  
rdfs:comment "An organisation which performs research and disseminates the findings of that research via Scientific Papers" .

### http://localhost/ontologies/20 14/7/PaperNew#Research\_Paper

:Research\_Paper rdf:type owhClass ;  
rdfs:subClassOf Publication ,  
:Research\_Paper ;  
rdfs:comment "A scientific paper researching new knowledge"^^xsd:string ,  
"A structured natural language document whose structure follows scientific paper conventions" .

### http://localhost/ontologies/20 14/7/PaperNew#Role

:Role rdf:type owhClass ;  
rdfs:label "Role of Person"^^xsd:string ;  
rdfs:comment "The role which a person may assume. For example a journalist may be a Political Correspondent of a Fashion Correspondent" .

### http://localhost/ontologies/20 14/7/PaperNew#Section

:Section rdf:type owhClass ;  
rdfs:label "Logical Component of Research Paper"^^xsd:string ;  
rdfs:comment "A chapter of a scientific paper"^^xsd:string .

### http://localhost/ontologies/20 14/7/PaperNew#Title

Title rdf:type owhClass ;  
rdfs:comment "The name given to a research paper"^^xsd:string .

### http://localhost/ontologies/20 14/7/PaperNew#Treatment

Treatment rdf:type owhClass ;  
rdfs:comment "A course of action to be taken by a patient"^^xsd:string .

### http://localhost/ontologies/20 14/7/PaperNew#University

:University rdf:type owhClass ;  
rdfs:subClassOf :Research\_Institution ;  
rdfs:comment "A not for profit research and training institution" .

#####

```

#
#  Individuals
#
#####

### http://localhost/ontologies/20_14/7/PaperNew#A1

:A1 rdf:type :Activity ,
      owhNamedIndividual ;
:ActivityID "6000"^^xsd:positiveInteger ;
:Activity_Name "demonstrates Leadership"^^xsd:string ;
rdfs:comment "An activity that Tony Abbot performs" .

### http://localhost/ontologies/20_14/7/PaperNew#A2

:A2 rdf:type :Activity ,
      owhNamedIndividual ;
:ActivityID "6001"^^xsd:positiveInteger ;
:Activity_Name "Remains in Custody"^^xsd:string ;
rdfs:comment "The activity of remaining in custody performed by Harriet Wran"^^xsd:string .

### http://localhost/ontologies/20_14/7/PaperNew#A3

:A3 rdf:type owhNamedIndividual ;
:ActivityID "6002"^^xsd:positiveInteger ;
:Activity_Name "Charged with Murder"^^xsd:string ;
rdfs:comment "The activity of being charged for murder performed by Harriet Wran"^^xsd:string .

### http://localhost/ontologies/20_14/7/PaperNew#Abbott_a_Leader

:Abbott_a_Leader rdf:type :News_Item ,
      owhNamedIndividual ;
:ArticleID "1000"^^xsd:positiveInteger ;
:SubmittedBy :Andrew_Bolt .

### http://localhost/ontologies/20_14/7/PaperNew#Andrew_Bolt

:Andrew_Bolt rdf:type Journalist ,
      owhNamedIndividual ;
:InvolvedPartyID "10"^^xsd:positiveInteger ;
:Given_Name "Andrew"^^xsd:string ;
:Surname "Bolt"^^xsd:string ;
rdfs:comment "Andrew Bolt is a Journalist" ;
:Submits :Abbott_a_Leader ;
:Has_Journalist_Role :PoliticalCommentator .

### http://localhost/ontologies/20_14/7/PaperNew#Crime_Editor

:Crime_Editor rdf:type :Journalist_Type ,
      owhNamedIndividual ;
:RoleID "2001"^^xsd:positiveInteger ;
:Role_Name "Crime Editor"^^xsd:string .

### http://localhost/ontologies/20_14/7/PaperNew#E_1

:E_1 rdf:type :Event ,
      owhNamedIndividual ;
:Event_Start_Time "2012-04-12T12:00:00"^^xsd:dateTime ;
:EventID "5000"^^xsd:positiveInteger ;

```

```

:Event_Description_Text "THE best thing about Tony Abbott's first year as Prime Minister has been —
thank God — he isn't Labor. But that's no longer the main reason he's looking hard to beat, after months
of trailing in the polls"^^xsd:string ;
:Event_Name_Subject "Tony Abbott"^^xsd:string ;
:Event_Description_RDF ""Tony_Abbott isNot Labor;
Tony_Abbott isLooking Hard_to_beat;
Tony_Abbott wasTrailingIn Polls
""^^xsd:string ;
:Consists_of :A1 .

```

```

### http://localhost/ontologies/20_14/7/PaperNew#E2

```

```

:E2 rdf:type :Event ,
      owhNamedIndividual ;
:Event_Start_Time "2014-10-08T12:53:00"^^xsd:dateTime ;
:EventID "5002"^^xsd:positiveInteger ;
:Event_Description_RDF ""Harriet Wran to remain in custody .
Harriet Wran charged with murder .""^^xsd:string ;
:Event_Description_Text "Harriet Wran to remain in custody charged with murder"^^xsd:string ;
:Consists_of :A2 ,
              :A3 .

```

```

### http://localhost/ontologies/20_14/7/PaperNew#E3

```

```

:E3 rdf:type :Event ,
      owhNamedIndividual ;
:EventID "5003"^^xsd:positiveInteger ;
:Event_Description_Text "The daughter of the late NSW premier Neville Wran has been remanded in
custody following a brief mention of the murder case against her in a Sydney court."^^xsd:string .

```

```

### http://localhost/ontologies/20_14/7/PaperNew#HWran

```

```

:HWran rdf:type :News_Item ,
              owhNamedIndividual ;
:ArticleID "1001"^^xsd:positiveInteger ;
:SubmittedBy :Nic_Ralston .

```

```

### http://localhost/ontologies/20_14/7/PaperNew#Harriet_Wran

```

```

:Harriet_Wran rdf:type Person ,
                 owhNamedIndividual ;
:InvolvedPartyID "13"^^xsd:positiveInteger ;
:Given_Name "Harriet"^^xsd:string ;
:Surname "Wran"^^xsd:string ;
:Participates_Actively :A2 ,
                       :A3 .

```

```

###

```

```

http://localhost/ontologies/2014/7/PaperNew#Harriet_Wran_to_remain_in_custody_charged_with_murder

```

```

:Harriet_Wran_to_remain_in_custody_charged_with_murder rdf:type :Headline ,
                                                             owhNamedIndividual ;
:HeadlineID "7001"^^xsd:positiveInteger ;
:Headline_Subject "Harriet Wran"^^xsd:string ;
:Headline_Text "Harriet Wran to remain in custody charged with
murder"^^xsd:string .

```

```

### http://localhost/ontologies/20_14/7/PaperNew#Labor

```



:Labor rdf:type Organisation ,  
owkNamedIndividual ;

:InvolvedPartyID "110"^^xsd:positiveInteger .

### http://localhost/ontologies/2014/7/PaperNew#Leader

:Leader rdf:type :Role ,  
owl:NamedIndividual ;

:RoleID "2002"^^xsd:positiveInteger ;

:Role\_Name "Leader"^^xsd:string .

### http://localhost/ontologies/2014/7/PaperNew#Nic\_Ralston

:Nic\_Ralston rdf:type Journalist ,  
owkNamedIndividual ;  
:InvolvedPartyID "12"^^xsd:positiveInteger ;  
:Surname "Nic"^^xsd:string ;  
:Given\_Name "Ralston"^^xsd:string ;  
:Has\_Journalist\_Role :Crime\_Editor ;  
:Submits :HWran .

### http://localhost/ontologies/2014/7/PaperNew#Parliament\_House

:Parliament\_House rdf:type :Location ,  
owkNamedIndividual ;  
:LocationID "4001"^^xsd:positiveInteger ;  
:Location\_Address "Canberra"^^xsd:string ;  
:is\_venue\_for :E1 .

### http://localhost/ontologies/2014/7/PaperNew#PoliticalCommentator

:PoliticalCommentator rdf:type :Journalist\_Type ,  
owkNamedIndividual ;  
:RoleID "2005"^^xsd:positiveInteger ;  
:Role\_Name ""Political Commentator  
""^^xsd:string .

### http://localhost/ontologies/2014/7/PaperNew#SMH

:SMH rdf:type :News\_Organisation ,  
owkNamedIndividual ;  
:InvolvedPartyID "120"^^xsd:positiveInteger ;  
:Organisation\_Name "Fairfax"^^xsd:string ;  
Publishes :HWran ;  
:has\_Staff :Nic\_Ralston .

### http://localhost/ontologies/2014/7/PaperNew#The\_Australian

:The\_Australian rdf:type :News\_Organisation ,  
owkNamedIndividual ;  
:InvolvedPartyID "100"^^xsd:positiveInteger ;  
:Organisation\_Name "News Corporation"^^xsd:string ;  
rdfs:comment ""The Australian is a news Organisation  
"" ;  
Publishes :Abbott\_a\_Leader ;  
:has\_Staff :Andrew\_Bolt .

### http://localhost/ontologies/20\_14/7/PaperNew#Tony\_Abbott

```
:Tony_Abbott rdf:type Person ,
              owhNamedIndividual ;
  :InvolvedPartyID "1"^^xsd:positiveInteger ;
  :Surname "Abbott"^^xsd:string ;
  :Role_Name "Leader"^^xsd:string ;
  :Given_Name "Tony"^^xsd:string ;
  :Participates_Actively :A1 ;
  :hasRole :Leader .
```

### http://localhost/ontologies/20\_14/7/PaperNew#Tony\_Abbott\_has\_arrived\_as\_a\_true\_leader

```
:Tony_Abbott_has_arrived_as_a_true_leader rdf:type :Headline ,
              owhNamedIndividual ;
  :HeadlineID "7000"^^xsd:positiveInteger ;
  :Headline_Subject "Tony Abbot"^^xsd:string ;
  :Headline_Text "Tony Abbott has arrived as a true leader"^^xsd:string ;
  :Headline_Object "True Leader"^^xsd:string ;
  :Headline_Predicate "hasArrivedAs"^^xsd:string .
```

### Generated by the OWL API (version 3.4.2) <http://owlapi.sourceforge.net>

**CLAIMS:**

1. A system for categorising and referencing a document using an electronic processing device, wherein:
  - the electronic processing device reviews the content of the document to identify structures within the document;
  - wherein the identified structures are referenced against a library of structures stored in a database;
  - wherein the document is categorised according to the conformance of the identified structures with those of the stored library of structures; and
  - wherein the categorised structure is added to the stored library.
2. A system for categorising and referencing a document according to Claim 1, wherein elements of the identified structure are referenced through Natural Language Processing.
3. A system for categorising and referencing a document according to Claim 1, wherein when there is no conformance between the identified structure and the stored library of structures the system searches through documents in an external network for structures that match the identified structure.
4. A system for categorising and referencing a document according to Claim 1, wherein a searchable reference is associated with the document and linked to the document.
5. A system for arranging text and extracting information from the text using an electronic processing device wherein:
  - the system scans the text and assigns it a classification; and
  - wherein on assigning the classification the system associates a template to the text based on the classification.
6. The system as claimed in Claim 5, wherein the template separates the text into distinct sections.
7. The system as claimed in Claim 6, wherein each distinct section is assigned a subject, an object and a predicate.
8. The system as claimed in Claim 7, wherein the subject, object and predicate are grouped in a resource description framework.

9. An electronic system for classifying the content of text wherein the system reviews the text and assigns different regions of the text a primary subject, a primary object and a primary predicate and stores the assignment as a resource description framework.
10. The electronic system as claimed in Claim 9, wherein the different regions of text are reviewed and further allocated at least one additional subject, at least one additional object and at least one additional predicate.
11. A system for categorising and referencing the contents of a document using an electronic processing device, wherein:
  - the electronic processing device reviews the contents of the documents and identifies the expression of at least one concept within the document; and
  - wherein the identified expression of the at least one concept is assigned a searchable resource description framework.
12. A system for categorising and referencing the contents of a document as claimed in Claim 11, wherein the searchable resource description framework consists of a subject, predicate and object in a resource description framework triplet.
13. A system for categorising and referencing the contents of a document as claimed in Claim 11, wherein a plurality of resource description framework triplets are associated with the contents of the document.
14. A system for categorising and referencing the contents of a document as claimed in any one of Claims 11 to 13, wherein if a non-specifically identifying subject, object or predicate is used within a first resource description framework an anaphora is applied to the surrounding resource description frameworks to identify the non-specifically identifying subject, object or predicate.
15. A system for categorising and referencing the contents of a document as claimed in Claim 14, wherein if a subordinate conjunction appears first in the resource description framework followed by a personal pronoun and then a dependent clause, the system will associate the non-specifically identifying subject, object or predicate with the corresponding subject, object or predicate in the following resource description frameworks.

16. A method for configuring data for storage in a knowledge base, the method comprising:

providing an electronic document having document type attributes;

capturing the document type attributes in a template that is based on an ontology language based upon the capturing by the template, defining a structure of the document determined by document type attributes to generate a defined structure as output; and

storing the defined structure for further processing.

17. The method of Claim 16, wherein the document further comprises base concepts, the further comprising:

based upon the capturing by the template, defining base concepts and generating base concepts as output; and

storing the base concepts for further processing.

18. The method of Claim 17 further comprising:

arranging the defined structure and base concepts into concept maps based upon the ontology language; and

storing the concept maps in a knowledge base.

19. The method of Claim 17, wherein a base concept is either explicit or implicit and when the base concept is an implicit concept, the method further comprises accessing external resources via the Internet to transform an implicit base concept into an explicit base concept.

20. The method of Claim 16 further comprising defining a sub-structure of the document based upon the sub-structure being captured by a sub-structure template.

21. The method of Claim 16, further comprising building a template to capture the document type attributes.

22. The method of Claim 17, further comprising generating a conformance report to generate validation output as to the use of a particular template with a particular document.

23. A method of configuring data for storage in a knowledge base, the method comprising:

providing an electronic document having contextual properties;

expressing the contextual properties as resource description framework units representative of a subject, an object and a predicate;

arranging the resource description framework units so that the contextual properties can be referenced between two or more resource description framework unit to generate arranged resource description framework units;

performing anaphoric resolution between the arranged resource description framework units to generate one or more resource description framework units concepts from the resource description framework units; and

storing the resource description framework concepts into concept maps based upon an ontology language for further processing.

24. The method of Claim 23, further comprising storing the concept maps in a knowledge base.

25. The method of Claim 8, wherein a resource description framework concept is either explicit or implicit and when the resource description framework concept is an implicit resource description framework concept, the method further comprises accessing external resources via the Internet to transform a resource description framework implicit concept into a resource description framework explicit concept.

26. The method for configuring data for storage in a knowledge base of Claim 2 and Claim 8, further comprising merging base concepts and resource description framework concepts to form merged concepts wherein a merged concept is either explicit or implicit and when the merged concept is an implicit merged concept, the method further comprises:

accessing external resources via the Internet to transform an implicit merged concept into an explicit merged concept; and

storing explicit merged concepts in knowledge base.

27. The method for configuring data for storage in a knowledge base of Claim 1 and Claim 8, further comprising merging the defined structure and resource description framework concepts to form merged concepts wherein a merged concept is either explicit or implicit and when the merged concept is an implicit merged concept, the method further comprises:

accessing external resources via the Internet to transform an implicit merged concept into an explicit merged concept; and

storing explicit merged concepts in knowledge base.

28. The method of Claim 11 or Claim 12 further comprising querying the knowledge base to retrieve an explicit merged concept.
29. The method for configuring data for storage in a knowledge base of Claim 23, wherein the contextual properties have a time element associated with them.
30. The method for configuring data for storage in a knowledge base of Claim 30, wherein the time element is specific or extends over a period.
31. The system for categorising and referencing a document as claimed in Claim 1, wherein a measure of time is associated with the identified structure.

1/9

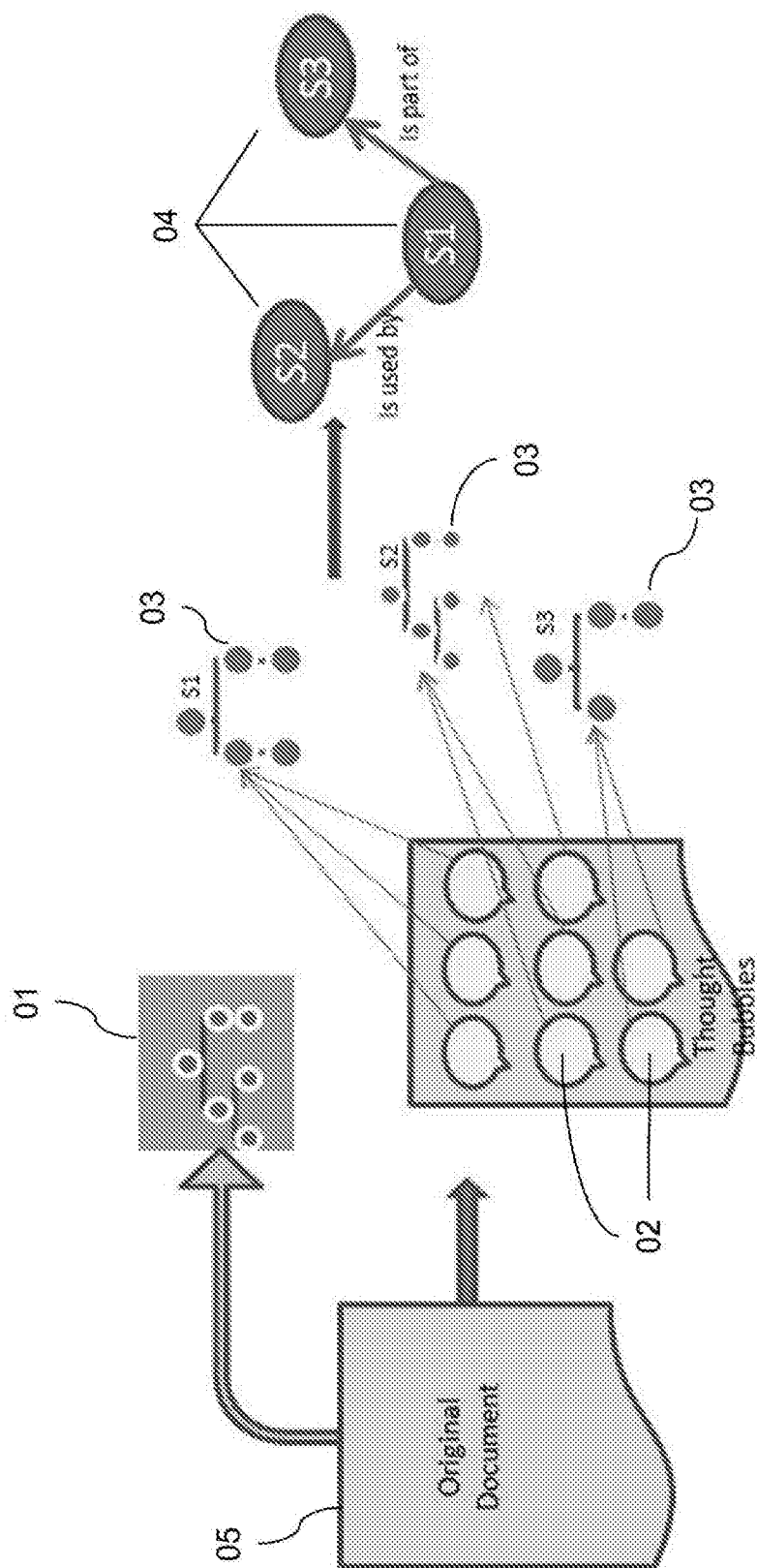


Figure 1



2/9

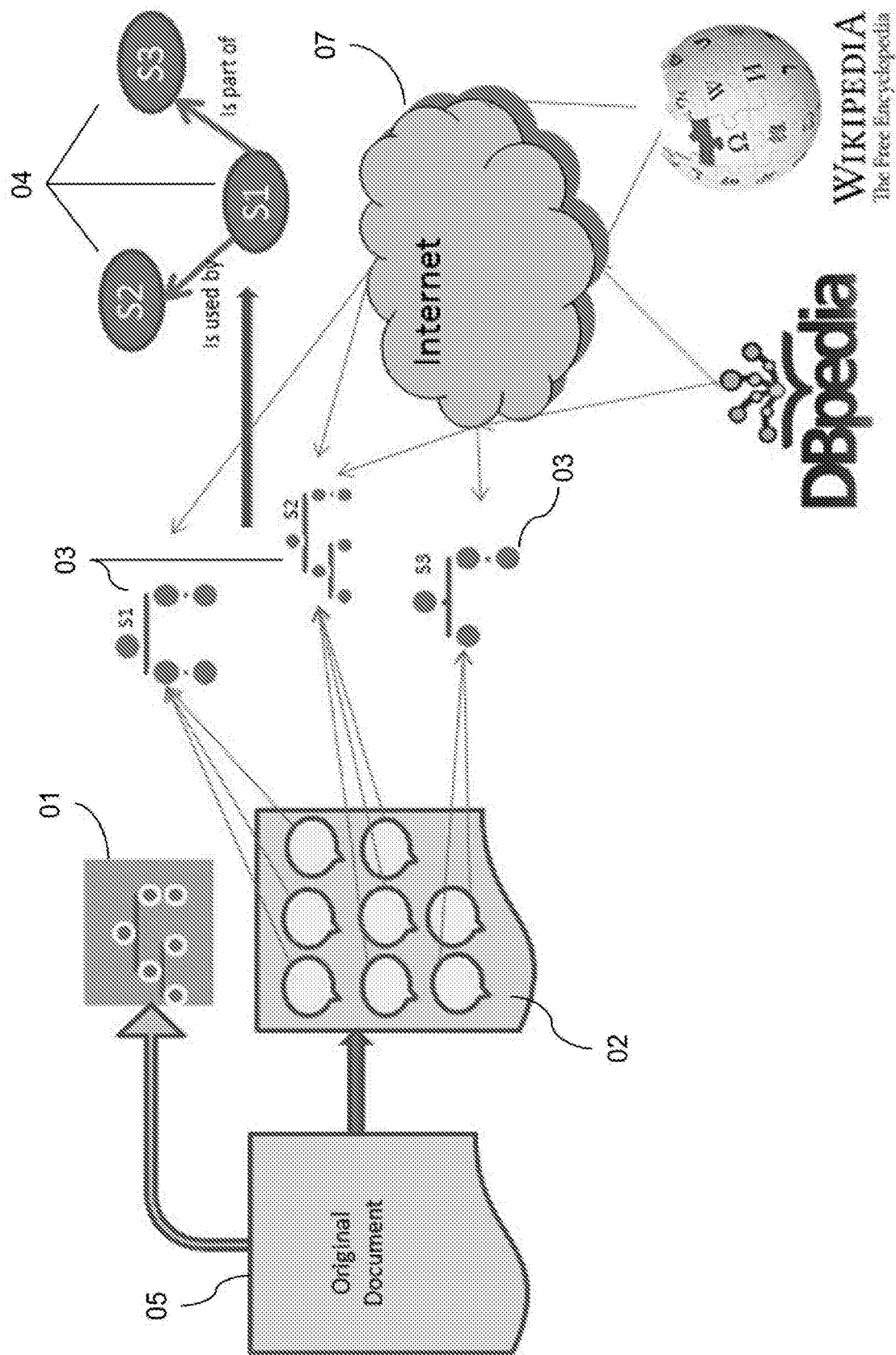


Figure 2

3/9

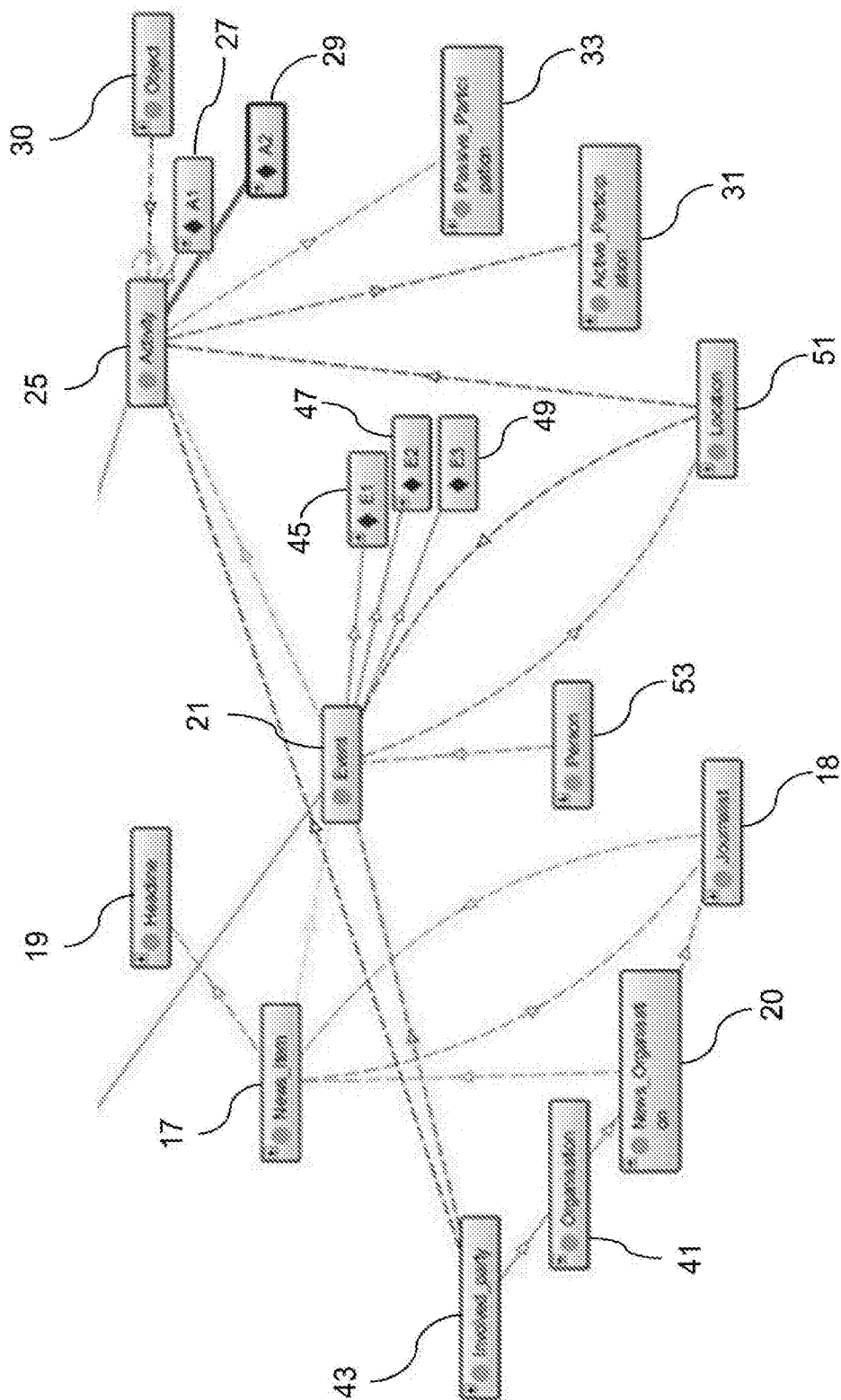


Figure 3



Figure 4

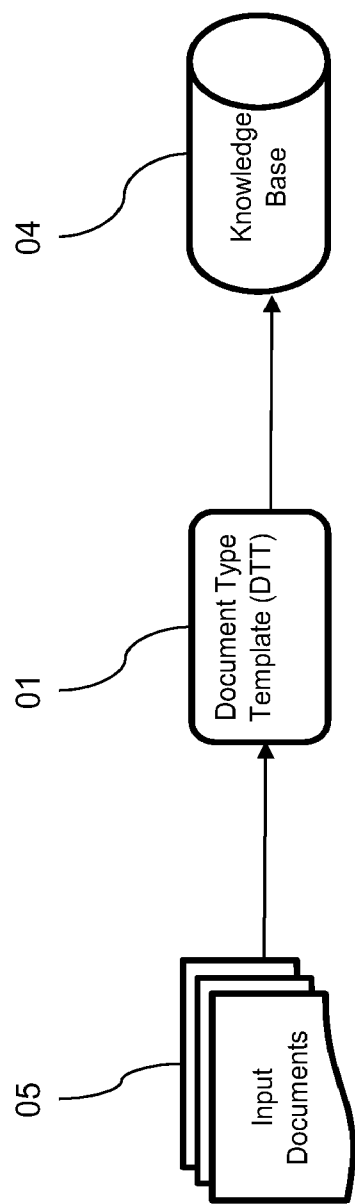


Figure 5

6/9

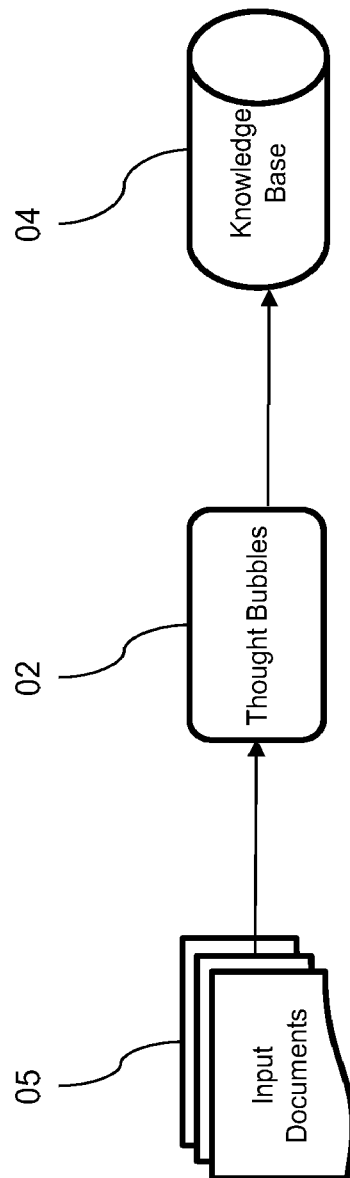


Figure 6

7/9

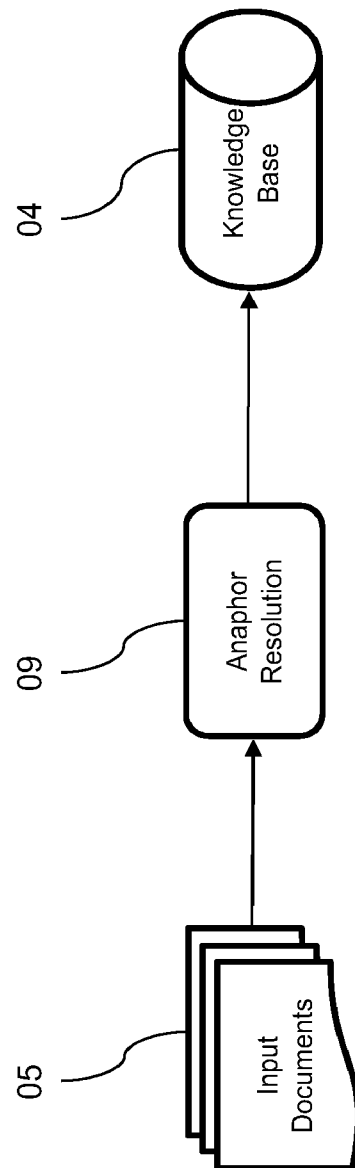


Figure 7

8/9

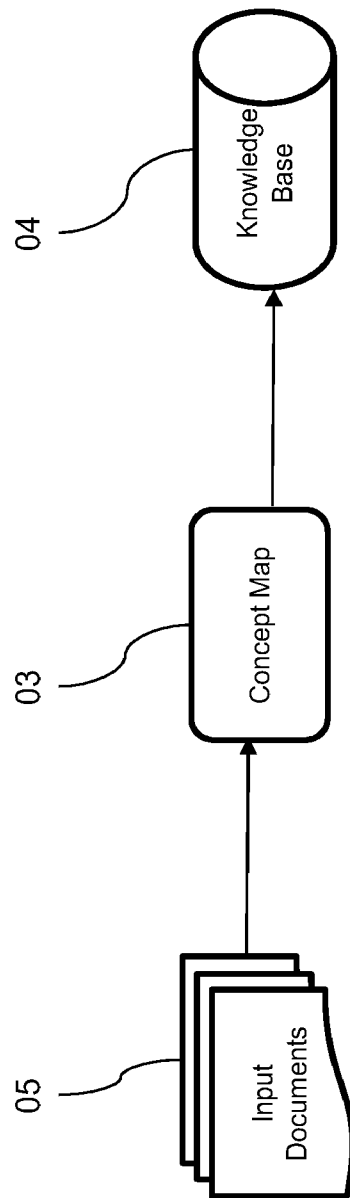


Figure 8

9/9

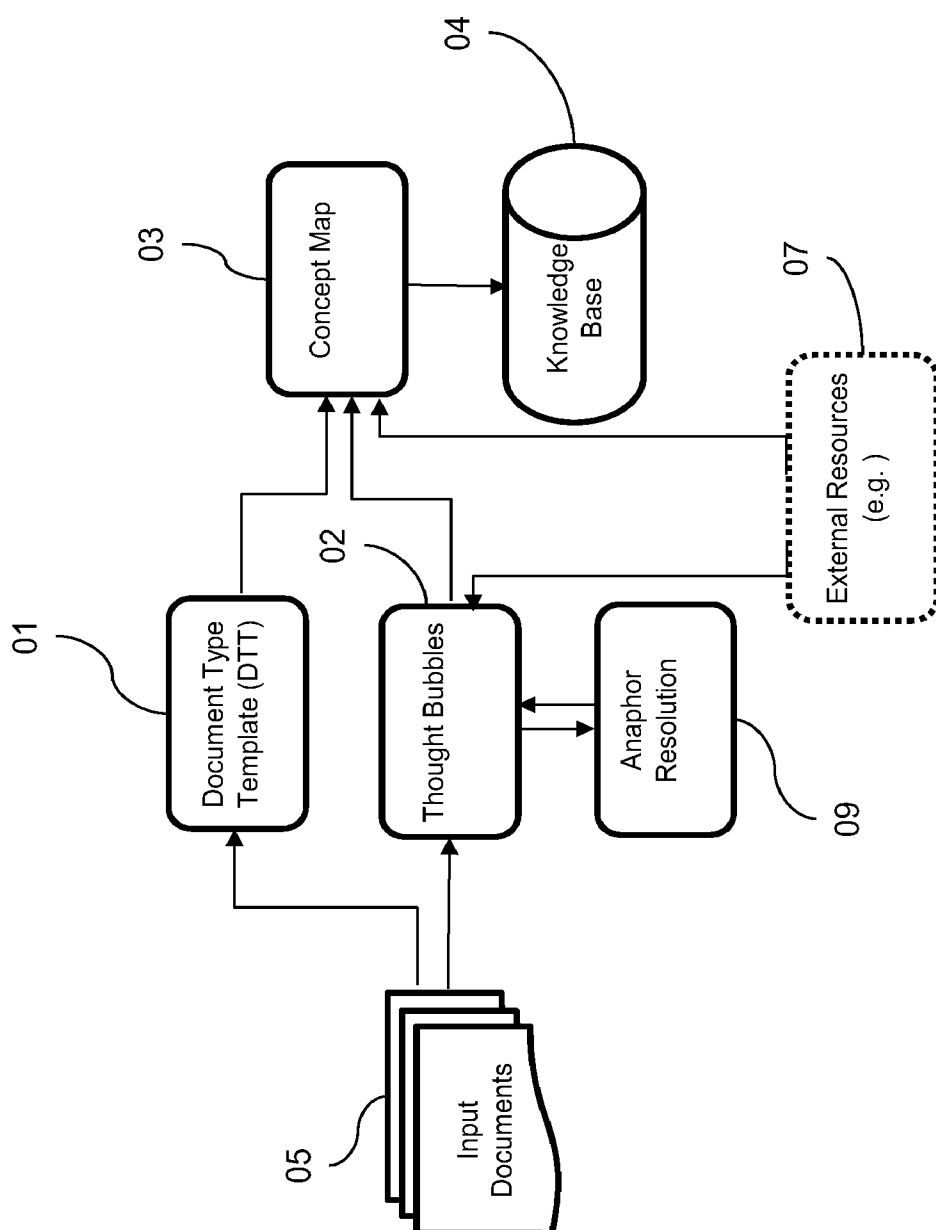


Figure 9



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/AU2016/000096

## A. CLASSIFICATION OF SUBJECT MATTER

G06N 5/00 (2006.01) G06F 17/30 (2006.01) G06F 17/27 (2006.01)

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)

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

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

WPIAP, EPODOC, TXTE: IPC G06N5, O06F17, G06N5/025, G06F17/30734, G06F17/2785 & Keywords (knowledge, ontology, construct, build, generate, extract, mine, develop, create, derive, infer, concept, fact, information, nip, semantic, context, concept, meaning, class, category, document, anaphoric, coreference, noun, resolution, type, template, vocabulary, lexis, language, structure, rdf, owl, triple, predicate, compare, match, attribute, metadata) & like terms; GOOGLE PATENTS: Keywords (document, text, language, lexis, topic, taxonomy, vocabulary, template, type, kind, construct, create, build, generate, extract, mine, develop, ontology, knowledge, natural language processing, nip, identify, extract, detect, semantic, meaning) & like terms; GOOGLE SCHOLAR: Keywords (anaphoric, coreference, bridging, rdf, framework, semantic, ontology, context, concept, triple, predicate, subject, object, owl, xml, document, type) & like terms, Inventor Names; ESPACENET: Applicant & Inventors Name; IP AUSTRALIA'S INTERNAL DATABASES: Inventor & Applicant Names

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Documents are listed in the continuation of Box C	



Further documents are listed in the continuation of Box C



See patent family annex

* "A"	Special categories of cited documents: document defining the general state of the art which is not considered to be of particular relevance	"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
"E"	earlier application or patent but published on or after the international filing date	"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
"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)	"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
"O"	document referring to an oral disclosure, use, exhibition or other means	"&"	document member of the same patent family
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search  
27 May 2016

Date of mailing of the international search report  
27 May 2016

## Name and mailing address of the ISA/AU

AUSTRALIAN PATENT OFFICE  
PO BOX 200, WODEN ACT 2606, AUSTRALIA  
Email address: pct@ipaustalia.gov.au

## Authorised officer

Olutope Omogbenigun  
AUSTRALIAN PATENT OFFICE  
(ISO 9001 Quality Certified Service)  
Telephone No. 0262832876

INTERNATIONAL SEARCH REPORT		International application No.
C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		PCT/AU2016/000096
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2014/105867 A2 (XSB, INC.) 03 M y 2014 See the whole document, especially abstract, paragraphs [0020],[0048],[0049],[0057],[0062],[0063],[00168],[00232]	1-22,3 1
X	US 2004/0083 199 A1 (GOVINDUGARI et al.) 29 April 2004 See the whole document, especially paragraphs [0040],[005 1],[0075],[008 1],[0082],[0085],[0093],[0094],[0103],[01 18],[0129],[0130],[013 1],[01 84]	Claims 1-22,3 1
X	US 2008/0126399 A1 (MACGREGOR) 29 May 2008 See the whole document, especially abstract, paragraphs [0005]-[0007],[0032],[0043],[0062]-[0079],[0086]-[0089],[01 11],[01 12],[01 15],[0170],[0198]	5-30
X	US 201 1/0106807 A1 (SRIHARI et al.) 05 May 201 1 See the whole document, especially abstract, paragraphs [0020],[002 1],[0052],[0061],[0064],[0068],[0069],[0077],[0097],[0103], figs. 2,10	5-30

Form PCT/ISA/210 (fifth sheet) (July 2009)

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/AU2016/000096

## Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
the subject matter listed in Rule 39 on which, under Article 17(2)(a)(i), an international search is not required to be carried out, including
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

## Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

**See Supplemental Box for Details**

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☒ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☒ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☒ No protest accompanied the payment of additional search fees.

**Supplemental Box****Continuation of: Box III**

This International Application does not comply with the requirements of unity of invention because it does not relate to one invention or to a group of inventions so linked as to form a single general inventive concept.

This Authority has found that there are different inventions based on the following features that separate the claims into distinct groups:

- Claims 1-4,31 are directed to a system for categorising and referencing a document. The feature of wherein the document is categorised according to the conformance of the identified structures with those of the stored library of structures is specific to this group of claims.
- Claims 5-8,16-22 are directed to a system for arranging text and extracting information from the text and a method for configuring data for storage in a knowledge base. The feature of associating a template to the text is specific to this group of claims.
- Claims 9-15,23-30 are directed to a system for categorising and referencing the contents of a document and a method of configuring data for storage in a knowledge base. The feature of storing the assignment/concept/context as a resource description framework is specific to this group of claims.

PCT Rule 13.2, first sentence, states that unity of invention is only fulfilled when there is a technical relationship among the claimed inventions involving one or more of the same or corresponding special technical features. PCT Rule 13.2, second sentence, defines a special technical feature as a feature which makes a contribution over the prior art.

When there is no special technical feature common to all the claimed inventions there is no unity of invention.

In the above groups of claims, the identified features may have the potential to make a contribution over the prior art but are not common to all the claimed inventions and therefore cannot provide the required technical relationship. Therefore there is no special technical feature common to all the claimed inventions and the requirements for unity of invention are consequently not satisfied *a priori*.

<b>INTERNATIONAL SEARCH REPORT</b> Information on patent family members		International application No. <b>PCT/AU2016/000096</b>	
This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.			
<b>Patent Document/s Cited in Search Report</b> <b>Publication Number                      Publication Date</b>		<b>Patent Family Member/s</b> <b>Publication Number                      Publication Date</b>	
WO 2014/105867 A2	03 July 2014	WO 2014105867 A2	03 Jul 2014
		AU 20 13370424 A1	23 Jul 2015
		EP 2939104 A2	04 Nov 2015
		US 20141889 17 A1	03 Jul 2014
US 2004/0083 199 A1	29 April 2004	US 2004083 199 A 1	29 Apr 2004
US 2008/0126399 A1	29 May 2008	US 2008126399 A1	29 May 2008
US 201 1/0106807 A1	05 May 201 1	US 201 1106807 A1	05 May 201 1
<b>End of Annex</b>			
Due to data integration issues this family listing may not include 10 digit Australian applications filed since May 2001. Form PCT/ISA/210 (Family Annex)(July 2009)			