

(51) International Patent Classification:
G06F 17/30 (2006.01)(21) International Application Number:
PCT/CN2011/071298(22) International Filing Date:
25 February 2011 (25.02.2011)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): **EMPIRE TECHNOLOGY DEVELOPMENT LLC** [US/US];
2711 Centerville Road, Suite 400, Wilmington, Delaware 19808 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **FANG, Jun** [CN/CN];
No. 127, Youyi XiLu, Xi'an, Shaanxi 710072 (CN).(74) Agent: **KING & WOOD MALLESONS LAWYERS**;
20th Floor, East Tower, World Financial Centre, No.1 Dongsanhuan Zhonglu, Chaoyang District, Beijing 100020 (CN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, 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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, 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, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

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

(54) Title: ONTOLOGY EXPANSION

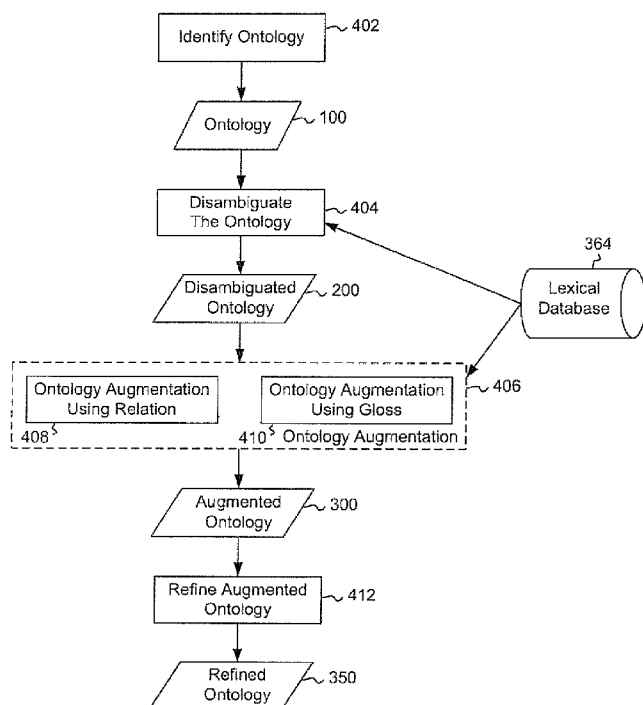


Figure 4

(57) Abstract: Techniques for ontology expansion are provided. In one disclosed method, an ontology is first identified. Terms in the ontology are disambiguated to identify term senses. Additional senses are identified using a lexical database. The additional senses are then added to the ontology to generate an augmented or expanded ontology.

ONTOLOGY EXPANSION

BACKGROUND

[001] The World Wide Web ("Web") provides a wealth of information and services to people around the world. The ability to discover information from around the globe often requires no more than a click of a mouse. At the same time, the Web is best suited for use by people. For example, tasks such as finding a specific translation of a word, searching for the lowest price on an item for sale, or making reservations at a restaurant or with an airline are often difficult for a machine to accomplish without human assistance.

[002] As a result, work is being done to make the Web more understandable. The Semantic Web, for example, tries to provide a framework to make the Web more understandable to both humans and machines by defining the meaning of information and services available on the Web. The goal is to enable the understanding and satisfaction of requests from various sources. The Semantic Web aims to enable machines, for example, to perform some of the tasks that are performed by humans today.

[003] Making the Web more understandable has many applications that include data integration, data classification, searching, content rating, data description, or the like. In order for these applications to come to fruition, however, it is necessary to identify the meaning or semantics of data and/or services on the Web.

[004] One of the tools used to determine the semantics of data and services on the Web is ontologies. Ontologies are used to express relationships among resources. For instance, there are many different terms that can be used to describe the same things in various data sets. Ontologies can identify these relationships and make it easier to determine the semantics of these data sets. Unfortunately, constructing ontologies is a labor intensive and costly process. In addition, ontologies are often incomplete and unfocused.

SUMMARY

[005] Embodiments relate to ontology construction and more particularly to expanding ontologies. In one embodiment, a method for expanding an ontology may include identifying or selecting the ontology. The ontology may be a part of

an existing ontology. After the ontology is selected, the selected ontology is disambiguated, which may include disambiguating terms in the ontology to identify term senses. The disambiguation of the terms may include the use of context to identify additional senses to be included in the expansion of the ontology. The ontology is then expanded with the additional senses related with the identified senses.

[006] In another embodiment, a method for expanding an ontology includes identifying a nearest word set and/or a context for each term in the ontology. A term sense for each term is determined and the terms senses are disambiguated using multiple sense sets and the nearest word set and/or context to identify additional senses. An augmented ontology is then generated by expanding the ontology with the additional senses. In some instances, the ontology can be refined, for example, by merging and/or removing unnecessary links.

[007] In another embodiment, a system for generating an expanded or augmented ontology from an ontology includes a lexical database and a disambiguation component. The lexical database stores multiple elements in sense sets. The disambiguation component expands the ontology using the sense sets in the lexical database to generate the augmented ontology. The disambiguation component identifies sense sets and determines a relatedness score between each term sense in the ontology and each of the elements in the sense sets. In one example, candidates are identified or generated from the sense sets. Each candidate includes may include an element from one of the sense sets. The elements in the candidate that has the best or maximum relatedness score is added to the ontology. More generally, elements whose relatedness score exceeds a threshold relatedness score are added to the ontology to generate the augmented ontology.

[008] In another example, a method for augmenting an ontology includes identifying an ontology to augment, disambiguating the ontology to generate a disambiguated ontology, augmenting the disambiguated ontology to generate an augmented ontology, and refining the augmented ontology. The disambiguated ontology can be augmented, for example, using relations between senses and/or gloss between senses.

[009] The foregoing summary is illustrative only and is not intended to be in any way limiting. In addition to the illustrative aspects, embodiments, and features

described above, further aspects, embodiments, and features will become apparent by reference to the drawings and the following detailed description.

BRIEF DESCRIPTION OF THE FIGURES

- 5 [010] Figure 1 illustrates an example of an ontology.
- [011] Figure 2 illustrates an example of the ontology of Figure 1 after the terms have been disambiguated.
- [012] Figure 3A shows an illustrative embodiment of a system for expanding an ontology.
- 10 [013] Figure 3B shows an illustrative embodiment of an expanded ontology.
- [014] Figure 3C shows an illustrative embodiment of the expanded ontology illustrated in Figure 3B after refinement.
- [015] Figure 4 shows an illustrative example of a method for expanding an ontology.
- 15 [016] Figure 5 shows a flowchart of an illustrative embodiment of a method for expanding an ontology.
- [017] Figure 6 is a block diagram of an illustrative embodiment of a computing device that is arranged for expanding an ontology.

20 DETAILED DESCRIPTION

[018] In the following detailed description, reference is made to the accompanying drawings, which form a part hereof. In the drawings, similar symbols typically identify similar components, unless context dictates otherwise. The illustrative embodiments described in the detailed description, drawings, and

25 claims are not meant to be limiting. Other embodiments may be utilized, and other changes may be made, without departing from the spirit or scope of the subject matter presented herein. It will be readily understood that the aspects of the present disclosure, as generally described herein, and illustrated in the Figures, can be arranged, substituted, combined, separated, and designed in a wide variety

30 of different configurations, all of which are explicitly contemplated herein.

[019] Embodiments disclosed herein relate to ontology construction including ontology expansion. An ontology can define concepts (e.g., words, terms) and roles, for example, in an area of knowledge. The ontology can provide the information (e.g., axioms) needed to relate words or terms that are used in the area

of knowledge. By providing relationships among words or terms, the ontology can be used to make the Web more understandable to both humans and machines. Ontology expansion improves, for example, the Semantic Web by refining relationships and/or adding to relationships in existing ontologies.

5 [020] Generally, ontology expansion or ontology augmentation may include identifying or selecting an ontology or a fragment of the ontology. Next, each term in the identified ontology is disambiguated and clarified as a term sense. The disambiguation process uses information from the context of each term to determine which sense (or senses) of a given term is being used in the ontology.
10 Then, the ontology can be expanded by appending terms which are related to term senses in the clarified ontology.

[021] Figure 1 illustrates an example of an ontology 100. The ontology 100 includes concepts and roles in this example. The concepts and roles are examples of the terminological aspects of the ontology 100 (e.g., a TBox and/or ABox).
15 The concepts and roles can be expressed in various ways. Figure 1 illustrates the concepts and roles in a graphical form. However, the ontology 100 can also be represented using statements, as a set of classes and properties, or the like.

[022] The ontology 100 may be a portion or fragment of a larger ontology. In this example, the ontology 100 may correspond to a fragment or portion of an arts ontology. The ontology 100 may contain two kinds of terms: a concept term and a role term. Each node in the ontology 100 identifies a term and, as depicted in Figure 1, the ontology 100 includes concept terms 102, 104, 106, 108, 110, 112, and role term 114. In Figure 1, the term 102 is arts. As further described herein, disambiguating the term 102 is substantially the same as disambiguating the word or term “arts” in this example. Thus, the words or terms in the ontology 100 are examples of concepts.
20
25

[023] The term 102 includes an arts concept and the terms 104, 106, and 108 include, respectively, movie, music, and literature concepts. In this example, the terms 104, 106, and 108 are each a subclass of the term 102. The term 102 is a superclass of the terms 104, 106, and 108. The terms 110 and 112 include, respectively, song and story. The terms 110 and 112 are subclasses of, respectively, the terms 104 and 108.
30

[024] The ontology 100 also includes a term 116, which is people. In this example, the terms 102 and 116 are associated by a role term 114, which is create

in this example. The role term 114 may indicate that some instances of the term 102 and some instances of the term 116 have the relation or role term 114 of create. As previously stated, the ontology 100 may include an ABox, a TBox, axioms, or the like that identify the concepts and/or roles illustrated graphically in Figure 1.

[025] The ontology 100 can be selected from existing ontologies or created without reference to existing ontologies. After the ontology 100 has been identified or selected for expansion, the terms 102, 104, 106, 108, 110, 112, 114, and/or 116 in the ontology 100 are clarified. Clarification of the ontology 100 often includes identifying term senses that correspond to the terms or words in the ontology 100. Identifying the term senses includes disambiguating term senses by taking other terms in the ontology as context. In other words, other terms in the ontology 100 may provide a context for disambiguating terms senses.

[026] When disambiguating a word (or more generally a term), the meaning of the word may be selected from a selection of sense possibilities. The sense possibilities may be predetermined and may be stored in a lexical database. In other words, the lexical database may store term senses. The lexical database may group words or term senses into groups or sense sets. For example, the words in the lexical database may be grouped into sense sets. The sense sets may also provide definitions and provide relationships between the words in the sets and between different sets. The lexical database can be used to disambiguate both concept terms and role terms.

[027] The sense sets in the lexical database may reflect relationships between a term sense and other term senses. Examples of sense sets for nouns in the lexical database may include, by way of example only, hypernyms, hyponyms, holonyms, meronyms, or the like. Relationships or sense sets for verbs may include hypernyms, troponyms, or the like. Relationships or sense sets for adjectives may include related nouns, participles, or the like. Relationships or sense sets for adverbs may include root adjectives. Words may also be connected through other lexical relationships, including antonyms.

[028] More generally, the lexical database typically stores lexical information about words and information about the meanings of words. The lexical database can distinguish between different types of speech (nouns, verbs, adjectives, adverbs), identify collocations, provide definitions, or the like. The lexical

database provides data for assigning senses to words in context. The lexical database may include morphemes, part of speech designation, sample usage, cultural annotations, semantic relationships with other words and morphemes.

[029] When provided with context, the lexical database can be used to identify term senses that are similar to a test word or a test term. Words that occur together in an ontology (e.g., in the ontology 100) are often related to some degree and the disambiguation of words or terms takes other words or terms in the ontology 100 into account. The right sense is assigned to a target word or target term by maximizing the relatedness between the target word and its context, which may be obtained from the ontology 100.

[030] Disambiguating the ontology may disambiguate all of the terms in the ontology at the same time or substantially at the same time. When obtaining the correct word or term senses for words or terms in the ontology, let word set $T = \{t_1, \dots, t_i, \dots, t_n\}$ be all words in the ontology, and let $ST = \{st_1, \dots, st_j, \dots, st_m\}$ be all possible sense combination sets for these words. The cardinality of ST is the product of cardinalities of all sense set $Sense(t_i)$. $Sense(t_i)$ denotes all possible senses of word t_i . In this example, $st_j = \{s_{j,1}, \dots, s_{j,i}, \dots, s_{j,n}\}$ is one possible combination of all senses. The sense set st which has the maximum relatedness score is chosen as the correct sense for all words in the ontology. In one example, the following formula determines the total relatedness among all sense $s_{j,i}$.

$$[031] \quad \sum Rel(s_{j,i}, s_{j,k}), \forall k, j(i > k).$$

[032] For example, an ontology may contain three terms: a, b, and c. Each of the terms a, b, and c, have possible term senses a_1, a_2, b_1, b_2 and c_1, c_2, c_3 in the lexical database, which can be used to disambiguate the terms in the ontology.

[033] The disambiguation of the terms may occur as follows, by way of example only. First, the potential disambiguation results or candidates are $\{a_1, b_1, c_1\}, \{a_1, b_1, c_2\}, \{a_1, b_1, c_3\}, \{a_1, b_2, c_1\}, \dots, \{a_2, b_2, c_3\}$. There may be 12 candidates in this example.

[034] Next, a relatedness score is determined or calculated for each candidate. The final or ultimate relatedness score may include a sum of relatedness scores among all pairs in a candidate. For instance, the relatedness score for the candidate $\{a_1, b_1, c_1\}$ is $Rel(a_1, b_1) + Rel(a_1, c_1) + Rel(b_1, c_1)$;

[035] The candidate with the maximum relatedness score or the best relatedness score is the result of the disambiguation. The candidate with the maximum or best relatedness score can be expressed as $\{a_i, b_j, c_k\}$. The desired sense of the terms a, b, c is a_i , b_j and c_k .

5 [036] Given two senses $s_{j,i}$ and $s_{j,k}$, the relatedness score $Rel(s_{j,i}, s_{j,k})$ can be obtained as described in "S. Banerjee, T. Pedersen, Extended gloss overlaps as a measure of semantic relatedness, in: Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, Acapulco, 2003, pp. 805–810, which is incorporated herein by reference in its entirety.

10 [037] The relatedness scores can be used to identify semantic relatedness to obtain a sense for every word in the ontology. A relatedness score can be measured by using an extended gloss overlap method in one embodiment. The extended gloss overlap method computes the relatedness between two input senses A and B by comparing the glosses of senses that are related to A and B through
15 relations provided in the lexical database.

[038] Let $RELS(A)$ be a set of relations that include one or more of the relations defined in the lexical database for the sense A, i.e., $RELS(A) = \{r | r \text{ is a relation defined in the lexical database for A}\}$. Each relation r in $RELS$ has a function of the same name that accepts a sense as input and returns the gloss of the senses
20 related to the input senses by the designated relation. For example, assume r represents the hypernym relation, then $r(A)$ returns the gloss of the hypernym synset of A.

[039] The $score()$ is a function that accepts as input two glosses, finds the phrase that overlaps words between them and returns a score. When comparing two
25 glosses, pronouns, prepositions, articles and/or conjunctions are often dropped. The sizes of the overlaps found are added together to arrive at the score for the given pair of glosses.

[040] The relatedness score between the input senses A and B can be determined as follows, where $gloss(A)$ denotes the gloss of A.

30 [041]
$$Rel(A, B) = \frac{score(gloss(A), gloss(B)) + \sum score(r(A), r(B)), \forall r \in (RELS(A) \cap RELS(B))}{0.5 \times (|gloss(A)| + |gloss(B)|)}$$

[042] Figure 2 is an example of the ontology after the terms have been disambiguated. In other words, Figure 2 illustrates an ontology 200 represented by senses or term senses while Figure 1 illustrates the ontology 100 represented

by terms. The ontology 200 is clarified in part because the ontology 200 illustrates term senses rather than terms. For example, an arts term sense 202 corresponds to the art term 102 after disambiguation in one embodiment. Each of the term senses illustrated in Figure 2 may include and/or be associated with sense sets 220. The sense sets 220 may be different for each of the term senses and may be separated such that the sense sets of a particular term sense are only associated with that term sense. Similarly, the other nodes in the ontology 200 include or represent term senses, which may also be expressed as words in one embodiment and also include or be associated with their own sense sets. The arrows 250 identify subclasses and/or superclasses for each term sense. For example, the term sense 208 is a subclass of the term sense 202 and a superclass of the term sense 212.

[043] When the ontology 200 is clarified, a lexical database and a relatedness score measurement can be used to augment these term senses. As previously stated, the lexical database may include relationships or sense sets between word senses such as hypernymy sense sets, hyponymy sense sets, antonymy sense sets, synonymy sense sets, meronymy sense sets, holonymy sense sets, or the like.

[044] In one example of ontology expansion, a subset of these relationships is used to perform word or term augmentation: $\text{hypernym}(X)$, $\text{hyponym}(X)$, and $\text{synonym}(X)$, where X is the sense. The lexical database can be used to identify the following sense sets:

$\text{hypernym}(X) = \{a1, a2, \dots\};$

$\text{hyponym}(X) = \{b1, b2, \dots\};$ and

$\text{synonym}(X) = \{c1, c2, \dots\}.$

[045] These sense sets include elements that are candidates to generate ontology augmentation. Generally, when a relatedness score for an element in one of the sense sets exceeds a threshold or predetermined value, then the corresponding element may be added to a class of the sense. In one example, the corresponding element can be added to a superclass, a subclass, or an equivalent class. In one example, the corresponding elements may be added as term senses in the ontology being augmented.

[046] Hypernyms typically identify senses that have a broader meaning than the sense. As a result, elements from the hypernym sense set are typically added to a superclass or added as senses that have a superclass relationship to the sense being

evaluated. Hyponyms typically identify senses that have a more specific meaning and elements from the hyponym sense set are typically added to a subclass of the senses being evaluated. Synonyms typically identify words that are similar semantically to the sense being disambiguated. As a result, elements from the synonym sense set may be added as siblings or as an equivalent class of the sense or as sibling senses in the ontology.

[047] The elements selected from the sense sets, as previously mentioned, are those that have sufficient relatedness based on some measurement, such as a relatedness score. More specifically, a hypernym relatedness score can be obtained between X and each element $\{a_1, a_2, \dots\}$. When the hypernym relatedness score is greater than a predetermined hypernym value, then the corresponding element is appended to a superclass set of X. A hyponym relatedness score can also be determined between X and each element $\{b_1, b_2, \dots\}$. When the hyponym relatedness score is greater than a predetermined hyponym value, then the corresponding element is appended to a subclass of X. Each element in $\text{synonym}(X)$ is directly placed into a sibling set or equivalent class of X, as synonyms are the most related.

[048] As discussed previously, ontology expansion may begin with ontology disambiguation. After the ontology is disambiguated, each sense in the disambiguated ontology is augmented using information related to the sense in the lexical database. As described in more detail herein, augmentation may include one or more steps.

[049] The ontology may be augmented using relation. For every sense, by way of example only, hypernymy, hyponymy and synonym relations are used to augment the sense. In this case, $\text{hypernym}(X)$, $\text{hyponym}(X)$ and $\text{synonym}(X)$ represent the result sense sets relation to word sense X - $\text{hypernym}(X) = \{a_1, a_2, \dots\}$, $\text{hyponym}(X) = \{b_1, b_2, \dots\}$ and $\text{synonym}(X) = \{s_1, s_2, \dots\}$

[050] If a relatedness score between X and an element in the sense set $\{a_1, a_2, \dots\}$ is greater than a hypernym value, the element is appended to the superclass set of X. If a relatedness score between X and element in the sense set $\{b_1, b_2, \dots\}$ is greater than a hyponym value, the element is appended to the subclass set of X; and elements in the sense set $\{s_1, s_2, \dots\}$ are placed into an equivalent class set of X. In addition, the ontology may also be augmented using gloss.

[051] After the ontology is augmented, the ontology is refined. Refining the ontology, as described in more detail herein, merges and removes some unnecessary links in the augmented ontology to refine the ontology.

5 [052] In one example, the relatedness score calculation method described previously is used in two steps: $Rel(s_{i,j}, s_{i,k})$ in ontology disambiguation, and in the calculation of a relatedness score between an augmented sense and a related element such as a_i in $hypernym(X)$ when augmenting the ontology using relations.

10 [053] Figure 3A shows an illustrative embodiment of a system for expanding or augmenting an ontology. Figure 3A illustrates the ontology 100 that is being expanded to generate an expanded or augmented ontology 300. A disambiguation component 368, which may be a computer or network of computers in one example or configured to operate on a computer or a network of computers, uses a lexical database 364 to disambiguate the ontology 100 and produce the augmented ontology 300. A display 360 and user interface 362 used in conjunction with the disambiguation component 368 can be used to view the ontology 100 and/or the augmented ontology 300 as well as to receive input from a user or other source during the generation of the augmented ontology 300.

15 [054] Figure 3B shows an illustrative embodiment of the augmented ontology 300. More specifically, with continued reference to Figures 1-2, the expansion of the ontology 100 in Figure 1 results in the augmented ontology 300 shown in Figures 3A and 3B. The ontology 100 may be expanded by determining a related sense set for each term sense, as illustrated in the ontology 200, and then adding term senses (or words) to the ontology 200 to generate the expanded ontology 300. More specifically, term senses or words in the related sense set that are sufficiently related are included in the expanded ontology 300.

20 [055] In one example, the ontology can also be expanded using gloss. In this example, all stopwords are dropped from the gloss. Stopwords may include articles, adjective, adverbs, preposition, etc. After dropping stopwords, the gloss may only contain nouns, verbs and pronouns. For a concept term sense cs , all pairs (verb, word) which match a pattern (cs , verb, word) or (word, verb, cs) in the gloss of cs are found (in this case a word is a noun or a pronoun). In one example, the verb is an assumption and will be validated in the following process. The word may be abstracted to a noun when the word is a pronoun. Next, (verb, con)

25 30

is disambiguated by using sense cs as the context. If the disambiguated result of the verb is not a verb sense, which means the above assumption for (cs, verb, word) is not exact, the pair is eliminated or dropped. Finally, if relatedness(cs, verb), relatedness(cs, con) and relatedness(verb, con) are greater than the gloss value, (verb, con) is appended to the concept term sense cs.

[056] For a role term sense rs, all pairs (w1, w2) which match pattern (w1, rs, w2) in the gloss of rs are found. In this example, w1 and w2 are nouns or pronouns. Then, the words are abstracted into general concepts con1 and con2 if necessary. Next, (con1, con2) is disambiguated by using sense rs as the context. Finally, if relatedness(con1, rs), relatedness(rs, con2) and relatedness(con1, con2) are greater than the gloss value, (verb, noun) is appended to the role term sense rs.

[057] For instance, in the disambiguated example ontology, the gloss of the concept term sense 212 (story#n#2) may contain "he writes stories for the magazines." The gloss is changed to "he writes stories magazines" by dropping stopwords (for and the) and the pair (write, he) is obtained. The pronoun "he" is abstracted to concept "people". The pair (write#v#1, people#n#1) is obtained using story#n#2 as the disambiguation context. By setting gloss value to 0.6, relatedness(people#n#1, write#v#1), relatedness(people#n#1, story#n#2) and relatedness(story#n#2, write#v#1) are greater than 0.6. As a result, (write#v#1, people#n#1) is appended to the story concept term sense 212, as depicted in Figure 3B by the role term sense 316 and the people concept term sense 318.

[058] In another example, the related sense sets of arts term sense 202 may include:

hypernym(arts) = {creation};

hyponym(arts) = {arts and crafts, decal, drawing, painting, perfumery, origami, sculpture};

synonym(arts) = {art creation}.

[059] In this example, the predetermined value of each of the hypernym, hyponym, and synonym relatedness scores may be set to 0.9. The words to be added to the expanded ontology 300 from the sense sets are then identified by generating relatedness scores for each element in each sense set. The relatedness score between arts and painting (relatedness (arts, painting)) may be 0.933. The relatedness score between the arts concept term sense 202 and drawing (from the elements identified in the hyponym(arts)) may be 0.923.

[060] One of skill in the art can appreciate that these relatedness scores are disclosed by way of example only. A similar process is performed for each term sense in the ontology 200. After (or during) the process of determining relatedness scores for the elements in the sense sets, the elements that exceed or otherwise satisfy the relatedness requirements are added to the augmented ontology 300.

[061] The augmented ontology 300 illustrated in Figure 3B depicts the results of the ontology expansion. The additions illustrated in the ontology 300 reflect the words or term senses whose relatedness scores satisfied (exceed in one example) the relevant threshold relatedness score.

[062] In Figure 3B, a painting concept term sense 302 has been added as subclasses of the arts concept term sense 202. A whistling concept term sense 306 is added as a subclass of the music concept term sense 204. A show concept term sense 310 has been added as a superclass of the movie concept term sense 206. A film concept term sense 308 has been added as an equivalent of the movie concept term sense 206. Concept term senses 312, 314, and 318 and role term senses 304 and 316 have been similarly added with similar relationships (e.g., subclass, superclass, equivalent class, etc.).

[063] The concept or role term senses added to the ontology 200 to generate the expanded ontology 300 can be added as superclasses, subclasses, and/or equivalent classes as previously described. The augmented ontology 300 improves the Semantic Web by providing an ontology that can be used to obtain a better understanding of the data and services available from the Web.

[064] Figure 3C shows an illustrative embodiment of the augmented ontology illustrated in Figure 3B after refinement. Refinement of the ontology 300 results in a refined ontology 350. Refining the ontology 350 merges links and removes unnecessary links. For instance, concept term senses and role term senses that are the same, are merged. Figure 3B, for example, illustrates two fiction concept term senses 312 and 314, two write role term senses 304 and 316, and two people concept term senses 216 and 318. When these are the same, they are merged as people concept term sense 352, fiction concept term sense 356, and role term sense 354.

[065] The removal of unnecessary links may be performed using certain rules. If one link can be inferred from other links, by way of example, then the link is

omitted. For example, the story concept term sense 212 is a subclass of the literature concept term sense 208 in Figure 3B. This relationship can be inferred from the relationships where the fiction concept term sense 312 is a subclass of the literature concept term sense 208 and the fiction concept term sense 314 is a superclass of the story concept term sense 212. Because the fiction concept term senses 312 and 314 are merged, the inferred relationship is illustrated in Figure 3C where the story concept term sense 212 is a subclass of the fiction concept term sense 356, which is a subclass of the literature concept term sense 208. Figure 3C thus illustrates an example of refining an augmented ontology by merging concept term senses or role term senses that are the same and by removing unnecessary links.

[066] Figure 4 shows a flowchart of an illustrative method for expanding an ontology. In box 402 ("Identify Ontology"), the ontology 100 is identified. As previously stated, the ontology 100 can be a portion of an existing ontology or a newly generated ontology, or a combination thereof.

[067] In box 404 ("Disambiguate the Ontology"), the ontology 100 is disambiguated to generate the disambiguated ontology 200. The lexical database 364 may be used to disambiguate the ontology 100 as described herein. This may include, for example, identifying nearest word sets, identifying word sense sets, identifying context for each sense or the like.

[068] In box 406 ("Augment the Ontology"), the expanded ontology 200 is augmented to generate the augmented ontology 300. The expanded ontology 200 can be augmented using relations in box 408 ("Ontology Augmentation Using Relation") and/or using gloss in box 410 ("Ontology Augmentation Using Gloss").

[069] In box 412 ("Refine the Augmented Ontology"), the augmented ontology 300 is refined to generate the refined ontology 350. The augmented ontology may be refined, for example, by combining concept term senses, role term senses, etc., as previously described.

[070] Figure 5 shows a flowchart of an illustrative embodiment of a method 500 for expanding an ontology. In block 502 ("Identify Ontology"), an ontology is identified. The ontology can be identified, for example, by searching an existing ontology for an area of knowledge or a particular concept. The ontology may be identified by selecting some of the nodes surrounding the particular concept or within the area of knowledge. In some examples, the extent or size of the

ontology being identified can be predetermined. In some examples, the process of expanding the ontology is performed on existing ontologies and the expansion is performed when a particular ontology or portion of an ontology is identified or otherwise selected.

5 [071] When expanding the ontology, the ontology to be expanded may already be identified or predetermined and embodiments of the ontology expansion may begin with an existing or predetermined ontology. As a result, there may be no need to search for the ontology or otherwise select the ontology. Further, the ontology being expanded may be a fragment or portion of a larger ontology.

10 [072] In block 504 (“Disambiguate Each Term in the Identified Ontology”), each term in the identified or predetermined ontology is disambiguated. Disambiguating each term in the ontology can include identifying the right or correct term senses for each term in the ontology which have the maximum relatedness score. The disambiguation of the terms in the ontology can clarify the
15 terms.

[073] The related sense sets can be identified using a lexical database, for example WordNet. The related sense sets may include a set of hypernyms, a set of hypernyms, a set of synonyms, or the like. The related sense sets are not limited to these examples, but may include others as previously stated.

20 [074] After the related sense sets are identified, a relatedness score may be determined, by way of example only, for each element in each related sense set using, for example, candidates as previously described. The elements used to expand the ontology are those elements whose relatedness score exceeds a predetermined relatedness score. The elements can be added as a superclass, a
25 subclass, or an equivalent class to the term sense being expanded.

[075] In block 506 (“Expand the Ontology”), the ontology is expanded with the elements identified during the disambiguation process. This may include adding the qualifying elements as superclasses, subclasses, equivalent classes or some related classes to the term senses in the ontology.

30 [076] Ontology expansion can be performed automatically or semi-automatically. After the ontology is identified or constructed, the ontology can be revised or expanded in accordance with the ontology expansion described herein. For instance, the ontology can be travelled (since it often has a tree-like organization) automatically to identify nearest word sets or to identify context and the like. The

context can be determined from other terms or term senses in the ontology. Further, the related sense sets and relatedness scores can also be obtained automatically without user input. Finally, the ontology can be expanded by adding nodes to the ontology. The expanded ontology can be saved automatically for use, reviewed by a person, or the like.

[077] One skilled in the art will appreciate that, for this and other processes and methods disclosed herein, the functions performed in the processes and methods may be implemented in differing order. Furthermore, the outlined steps and operations are only provided as examples, and some of the steps and operations may be optional, combined into fewer steps and operations, or expanded into additional steps and operations without detracting from the essence of the disclosed embodiments.

[078] The present disclosure is not to be limited in terms of the particular embodiments described in this application, which are intended as illustrations of various aspects. Many modifications and variations can be made without departing from its spirit and scope, as will be apparent to those skilled in the art. Functionally equivalent methods and apparatuses within the scope of the disclosure, in addition to those enumerated herein, will be apparent to those skilled in the art from the foregoing descriptions. Such modifications and variations are intended to fall within the scope of the appended claims. The present disclosure is to be limited only by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled. It is to be understood that this disclosure is not limited to particular methods, reagents, compounds compositions or biological systems, which can, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting.

[079] In an illustrative embodiment, any of the operations, processes, etc. described herein can be implemented as computer-readable instructions stored on a computer-readable medium. The computer-readable instructions can be executed by a processor of a mobile unit, a network element, and/or any other computing device.

[080] There is little distinction left between hardware and software implementations of aspects of systems; the use of hardware or software is generally (but not always, in that in certain contexts the choice between hardware

and software can become significant) a design choice representing cost vs. efficiency tradeoffs. There are various vehicles by which processes and/or systems and/or other technologies described herein can be effected (e.g., hardware, software, and/or firmware), and that the preferred vehicle will vary with the context in which the processes and/or systems and/or other technologies are deployed. For example, if an implementer determines that speed and accuracy are paramount, the implementer may opt for a mainly hardware and/or firmware vehicle; if flexibility is paramount, the implementer may opt for a mainly software implementation; or, yet again alternatively, the implementer may opt for some combination of hardware, software, and/or firmware.

[081] The foregoing detailed description has set forth various embodiments of the devices and/or processes via the use of block diagrams, flowcharts, and/or examples. Insofar as such block diagrams, flowcharts, and/or examples contain one or more functions and/or operations, it will be understood by those within the art that each function and/or operation within such block diagrams, flowcharts, or examples can be implemented, individually and/or collectively, by a wide range of hardware, software, firmware, or virtually any combination thereof. In one embodiment, several portions of the subject matter described herein may be implemented via Application Specific Integrated Circuits (ASICs), Field Programmable Gate Arrays (FPGAs), digital signal processors (DSPs), or other integrated formats. However, those skilled in the art will recognize that some aspects of the embodiments disclosed herein, in whole or in part, can be equivalently implemented in integrated circuits, as one or more computer programs running on one or more computers (e.g., as one or more programs running on one or more computer systems), as one or more programs running on one or more processors (e.g., as one or more programs running on one or more microprocessors), as firmware, or as virtually any combination thereof, and that designing the circuitry and/or writing the code for the software and or firmware would be well within the skill of one of skill in the art in light of this disclosure. In addition, those skilled in the art will appreciate that the mechanisms of the subject matter described herein are capable of being distributed as a program product in a variety of forms, and that an illustrative embodiment of the subject matter described herein applies regardless of the particular type of signal bearing medium used to actually carry out the distribution. Examples of a signal bearing

medium include, but are not limited to, the following: a recordable type medium such as a floppy disk, a hard disk drive, a CD, a DVD, a digital tape, a computer memory, etc.; and a transmission type medium such as a digital and/or an analog communication medium (e.g., a fiber optic cable, a waveguide, a wired communications link, a wireless communication link, etc.).

[082] Those skilled in the art will recognize that it is common within the art to describe devices and/or processes in the fashion set forth herein, and thereafter use engineering practices to integrate such described devices and/or processes into data processing systems. That is, at least a portion of the devices and/or processes described herein can be integrated into a data processing system via a reasonable amount of experimentation. Those having skill in the art will recognize that a typical data processing system generally includes one or more of a system unit housing, a video display device, a memory such as volatile and non-volatile memory, processors such as microprocessors and digital signal processors, computational entities such as operating systems, drivers, graphical user interfaces, and applications programs, one or more interaction devices, such as a touch pad or screen, and/or control systems including feedback loops and control motors (e.g., feedback for sensing position and/or velocity; control motors for moving and/or adjusting components and/or quantities). A typical data processing system may be implemented utilizing any suitable commercially available components, such as those typically found in data computing/communication and/or network computing/communication systems.

[083] Figure 6 is a block diagram of an illustrative embodiment of a computing device 600 that is arranged for expanding an ontology. In a very basic configuration 602, computing device 600 typically includes one or more processors 604 and a system memory 606. A memory bus 608 may be used for communicating between processor 604 and system memory 606.

[084] Depending on the desired configuration, processor 604 may be of any type including but not limited to a microprocessor (μ P), a microcontroller (μ C), a digital signal processor (DSP), or any combination thereof. Processor 604 may include one more levels of caching, such as a level one cache 610 and a level two cache 612, a processor core 614, and registers 616. An example processor core 614 may include an arithmetic logic unit (ALU), a floating point unit (FPU), a digital signal processing core (DSP Core), or any combination thereof. An

example memory controller 618 may also be used with processor 604, or in some implementations memory controller 618 may be an internal part of processor 604.

[085] Depending on the desired configuration, system memory 606 may be of any type including but not limited to volatile memory (such as RAM), non-volatile memory (such as ROM, flash memory, etc.) or any combination thereof. System memory 606 may include an operating system 620, one or more applications 622, and program data 624. Application 622 may include an ontology expansion application 626 that is arranged or configured to expand an ontology. Program data 624 may include expansion data 628 for analyzing an ontology and/or for expanding the ontology. In some embodiments, application 622 may be arranged to operate with program data 624 on operating system 620 such that an expanded ontology is generated. This described basic configuration 602 is illustrated in Figure 6 by those components within the inner dashed line.

[086] Computing device 600 may have additional features or functionality, and additional interfaces to facilitate communications between basic configuration 602 and any required devices and interfaces. For example, a bus/interface controller 630 may be used to facilitate communications between basic configuration 602 and one or more data storage devices 632 via a storage interface bus 634. Data storage devices 632 may be removable storage devices 636, non-removable storage devices 638, or a combination thereof. Examples of removable storage and non-removable storage devices include magnetic disk devices such as flexible disk drives and hard-disk drives (HDD), optical disk drives such as compact disk (CD) drives or digital versatile disk (DVD) drives, solid state drives (SSD), and tape drives to name a few. Example computer storage media may include volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information, such as computer readable instructions, data structures, program modules, or other data.

[087] System memory 606, removable storage devices 636 and non-removable storage devices 638 are examples of computer storage media. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which may be used to store the

desired information and which may be accessed by computing device 600. Any such computer storage media may be part of computing device 600.

[088] Computing device 600 may also include an interface bus 640 for facilitating communication from various interface devices (e.g., output devices 642, peripheral interfaces 644, and communication devices 646) to basic configuration 602 via bus/interface controller 630. Example output devices 642 include a graphics processing unit 648 and an audio processing unit 650, which may be configured to communicate to various external devices such as a display or speakers via one or more A/V ports 652. Example peripheral interfaces 644 include a serial interface controller 654 or a parallel interface controller 656, which may be configured to communicate with external devices such as input devices (e.g., keyboard, mouse, pen, voice input device, touch input device, etc.) or other peripheral devices (e.g., printer, scanner, etc.) via one or more I/O ports 658. An example communication device 646 includes a network controller 660, which may be arranged to facilitate communications with one or more other computing devices 662 over a network communication link via one or more communication ports 664.

[089] The network communication link may be one example of a communication media. Communication media may typically be embodied by computer readable instructions, data structures, program modules, or other data in a modulated data signal, such as a carrier wave or other transport mechanism, and may include any information delivery media. A “modulated data signal” may be a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal. By way of example, and not limitation, communication media may include wired media such as a wired network or direct-wired connection, and wireless media such as acoustic, radio frequency (RF), microwave, infrared (IR) and other wireless media. The term computer readable media as used herein may include both storage media and communication media.

[090] Computing device 600 may be implemented as a portion of a small-form factor portable (or mobile) electronic device such as a cell phone, a personal data assistant (PDA), a personal media player device, a wireless web-watch device, a personal headset device, an application specific device, or a hybrid device that include any of the above functions. Computing device 600 may also be

implemented as a personal computer including both laptop computer and non-laptop computer configurations.

[091] With respect to the use of substantially any plural and/or singular terms herein, those having skill in the art can translate from the plural to the singular and/or from the singular to the plural as is appropriate to the context and/or application. The various singular/plural permutations may be expressly set forth herein for sake of clarity.

[092] The herein described subject matter sometimes illustrates different components contained within, or connected with, different other components. It is to be understood that such depicted architectures are merely exemplary, and that in fact many other architectures can be implemented which achieve the same functionality. In a conceptual sense, any arrangement of components to achieve the same functionality is effectively "associated" such that the desired functionality is achieved. Hence, any two components herein combined to achieve a particular functionality can be seen as "associated with" each other such that the desired functionality is achieved, irrespective of architectures or intermedial components. Likewise, any two components so associated can also be viewed as being "operably connected", or "operably coupled", to each other to achieve the desired functionality, and any two components capable of being so associated can also be viewed as being "operably couplable", to each other to achieve the desired functionality. Specific examples of operably couplable include but are not limited to physically mateable and/or physically interacting components and/or wirelessly interactable and/or wirelessly interacting components and/or logically interacting and/or logically interactable components.

[093] With respect to the use of substantially any plural and/or singular terms herein, those having skill in the art can translate from the plural to the singular and/or from the singular to the plural as is appropriate to the context and/or application. The various singular/plural permutations may be expressly set forth herein for sake of clarity.

[094] It will be understood by those within the art that, in general, terms used herein, and especially in the appended claims (*e.g.*, bodies of the appended claims) are generally intended as "open" terms (*e.g.*, the term "including" should be interpreted as "including but not limited to," the term "having" should be interpreted as "having at least," the term "includes" should be interpreted as

“includes but is not limited to,” etc.). It will be further understood by those within the art that if a specific number of an introduced claim recitation is intended, such an intent will be explicitly recited in the claim, and in the absence of such recitation no such intent is present. For example, as an aid to understanding, the following appended claims may contain usage of the introductory phrases “at least one” and “one or more” to introduce claim recitations. However, the use of such phrases should not be construed to imply that the introduction of a claim recitation by the indefinite articles “a” or “an” limits any particular claim containing such introduced claim recitation to embodiments containing only one such recitation, even when the same claim includes the introductory phrases “one or more” or “at least one” and indefinite articles such as “a” or “an” (*e.g.*, “a” and/or “an” should be interpreted to mean “at least one” or “one or more”); the same holds true for the use of definite articles used to introduce claim recitations. In addition, even if a specific number of an introduced claim recitation is explicitly recited, those skilled in the art will recognize that such recitation should be interpreted to mean at least the recited number (*e.g.*, the bare recitation of “two recitations,” without other modifiers, means at least two recitations, or two or more recitations). Furthermore, in those instances where a convention analogous to “at least one of A, B, and C, etc.” is used, in general such a construction is intended in the sense one having skill in the art would understand the convention (*e.g.*, “a system having at least one of A, B, and C” would include but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc.). In those instances where a convention analogous to “at least one of A, B, or C, etc.” is used, in general such a construction is intended in the sense one having skill in the art would understand the convention (*e.g.*, “a system having at least one of A, B, or C” would include but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc.). It will be further understood by those within the art that virtually any disjunctive word and/or phrase presenting two or more alternative terms, whether in the description, claims, or drawings, should be understood to contemplate the possibilities of including one of the terms, either of the terms, or both terms. For example, the phrase “A or B” will be understood to include the possibilities of “A” or “B” or “A and B.”

[095] In addition, where features or aspects of the disclosure are described in terms of Markush groups, those skilled in the art will recognize that the disclosure is also thereby described in terms of any individual member or subgroup of members of the Markush group.

5 [096] From the foregoing, it will be appreciated that various embodiments of the present disclosure have been described herein for purposes of illustration, and that various modifications may be made without departing from the scope and spirit of the present disclosure. Accordingly, the various embodiments disclosed herein are not intended to be limiting, with the true scope and spirit being indicated by
10 the following claims.

CLAIMS

1. A method for expanding an ontology, the method comprising:
identifying the ontology, the ontology including a plurality of terms including
5 concept terms and role terms;
disambiguating the terms in the ontology using a lexical database to identify
additional senses; and
expanding the ontology with the additional senses.
2. The method of claim 1, wherein identifying the ontology further comprises
10 identifying a portion of the ontology.
3. The method of claim 1, wherein disambiguating the terms in the ontology
further comprises identifying a term sense for each of the plurality of terms.
4. The method of claim 3, wherein identifying a term sense for each of the
plurality of terms further comprises identifying a nearest word set for each of the plurality
15 of terms.
5. The method of claim 3, wherein identifying a term sense for each of the
plurality of terms further comprises identifying a context for each of the plurality of terms.
6. The method of claim 5, further comprising determining a plurality of sense
sets from the lexical database, wherein each sense set includes at least one element.
- 20 7. The method of claim 6, wherein the plurality of sense sets include one or
more of:
a hypernym sense set;
a hyponym sense set;
a holonym sense set;
25 a meronym sense set;
a troponym sense set;
a participle sense set; or
a root adjective sense set.
8. The method of claim 5, further comprising identifying a relatedness score
30 between each term in the ontology and each element in the plurality of sense sets, wherein
elements in the plurality of sense sets whose relatedness score exceeds a predetermined
relatedness score are included in the additional senses.
9. The method of claim 7, wherein each relatedness score is determined using
a sense of the term to provide context for the elements in the plurality of sense sets.

10. The method of claim 1, wherein each additional sense is added to the ontology as a superclass, as subclass, or an equivalent class.

11. A method for expanding an ontology, the method comprising:
identifying a context for each term in the ontology;
5 determining a term sense for each term using the context of each term;
disambiguating the term sense of each term using at least a plurality of sense sets and the context to identify additional senses; and
expanding the ontology with the additional senses.

12. The method of claim 11, wherein other terms in the ontology provide the
10 context.

13. The method of claim 11, wherein the plurality of sense sets include one or more of:

a hypernym sense set;
a hyponym sense set;
15 a holonym sense set;
a meronym sense set;
a troponym sense set;
a participle sense set; or
a root adjective sense set.

14. The method of claim 13, further comprising:
determining a relatedness score for at least some elements in at least some
20 of the sense sets; and
identifying first elements in the plurality of sense sets whose relatedness score exceeds a threshold are added to the ontology.

15. The method of claim 14, further comprising adding the first elements to the ontology to generate an expanded ontology.

16. The method of claim 15, wherein the first elements are added to the ontology as one of: a superclass to the associated term sense; a subclass to the associated term sense; or an equivalent class to the associated term sense.

17. The method of claim 14, further comprising determining a gloss for each term, wherein the gloss and the relatedness score are used to augment the ontology.

18. The method of claim 17, wherein determining a gloss includes a pattern, wherein pairs matching the pattern are disambiguated using the sense as context, wherein

each pair is appended to the term sense where relatedness scores are greater than a gloss value.

19. A system for generating an expanded ontology from an ontology, the system comprising:

5 a lexical database that stores a plurality of elements in one or more sense sets, each element including a term sense;

a disambiguation component that expands the ontology using the one or more sense sets in the lexical database to generate the expanded ontology, wherein the disambiguation component determines a context for each term sense in the ontology and determines a relatedness score between each term sense in the ontology and each of the elements in the one or more sense sets, wherein the disambiguation module generates the expanded ontology by adding elements in the one or more sense sets whose relatedness score exceeds a threshold relatedness score.

15 20. The system of claim 19 wherein the one or more sense sets include one or more of:

a hypernym sense set;
a hyponym sense set;
a holonym sense set;
20 a meronym sense set;
a troponym sense set;
a participle sense set; or
a root adjective sense set.

21. The system of claim 19 wherein the disambiguation component determines the relatedness score using a context of the term senses in the ontology, wherein the context is provided by other term senses in the ontology.

22. The system of claim 19, wherein the disambiguation component adds each element whose relatedness score exceed the threshold relatedness score as one of a superclass, a subclass, or an equivalent class.

30 23. The system of claim 19, wherein the disambiguation component generates the expanded ontology automatically without human assistance.

24. The system of claim 19, further comprising determining a gloss for each term sense, wherein the gloss includes a pattern, wherein pairs matching the pattern are

disambiguated and appended to the ontology when relatedness scores associated with the pair are greater than a gloss value

25. A method for augmenting an ontology, the method comprising:
identifying an ontology to augment;
5 disambiguating the ontology to generate a disambiguated ontology;
augmenting the disambiguated ontology to generate an augmented
ontology; and
refining the augmented ontology.
26. The method of claim 25, wherein disambiguating the ontology to generate
10 a disambiguated ontology comprises:
assigning sense sets to each term in the ontology; and
determining relatedness between each term in the ontology and words in
the sense sets, wherein words in the sense sets with sufficient relatedness are used
to identify a term sense for each term in the ontology.
- 15 27. The method of claim 25, wherein augmenting the disambiguated ontology
to generate an augmented ontology further comprises:
augmenting the disambiguated ontology using relations; and
augmenting the disambiguated ontology using glosses.
28. The method of claim 27, wherein
20 augmenting the disambiguated ontology using relations comprises:
determining a similar sense set for each term sense and
adding new term senses to the augmented ontology based on each
similar sense set; and
including term senses that are sufficiently similar in the
25 augmented ontology, wherein term senses included in the
augmented ontology have a relatedness score greater than a
predetermined value; and
augmenting the disambiguated ontology using glosses comprises:
dropping stopwords from a gloss for each term sense; and
30 determining gloss values for pairs in the gloss, wherein
term senses are added to the augmented ontology when the gloss
values exceed a threshold gloss value.
29. The method of claim 25, wherein refining the ontology further comprises:
merging term senses in the augmented ontology that are the same; and

inferring relationships between term senses, wherein the relationships are added as relations in the augmented ontology, the relations including one of a superclass, a subclass, or an equivalent class.

30. The method of claim 25, wherein disambiguating the ontology to generate
5 a disambiguated ontology further comprises:

generating disambiguation candidates from the terms in the ontology;

determining a relatedness score for each candidate, wherein at least the candidate with a best relatedness score is a desired sense of the terms in the ontology.

10 31. The method of claim 30, wherein the relatedness score for each candidate includes a sum of relatedness scores among all pairs of terms in each candidate.

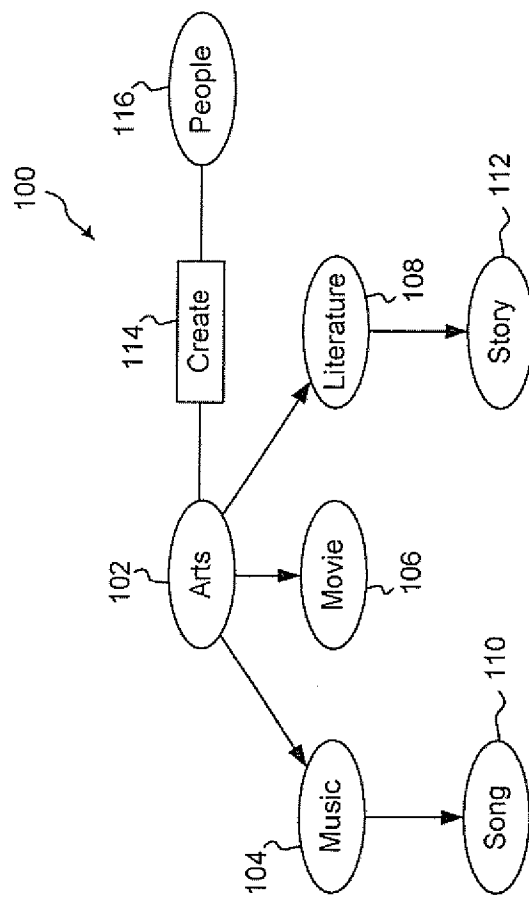


Figure 1

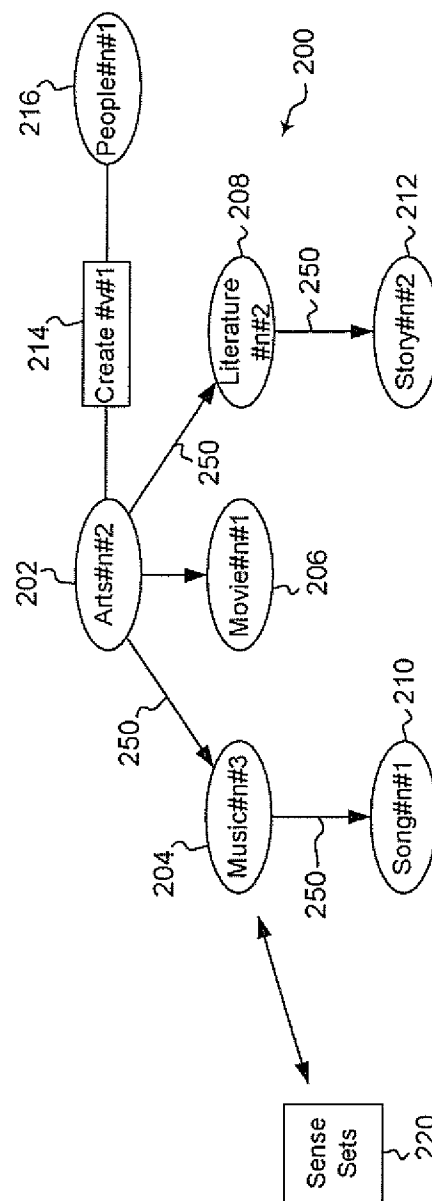


Figure 2

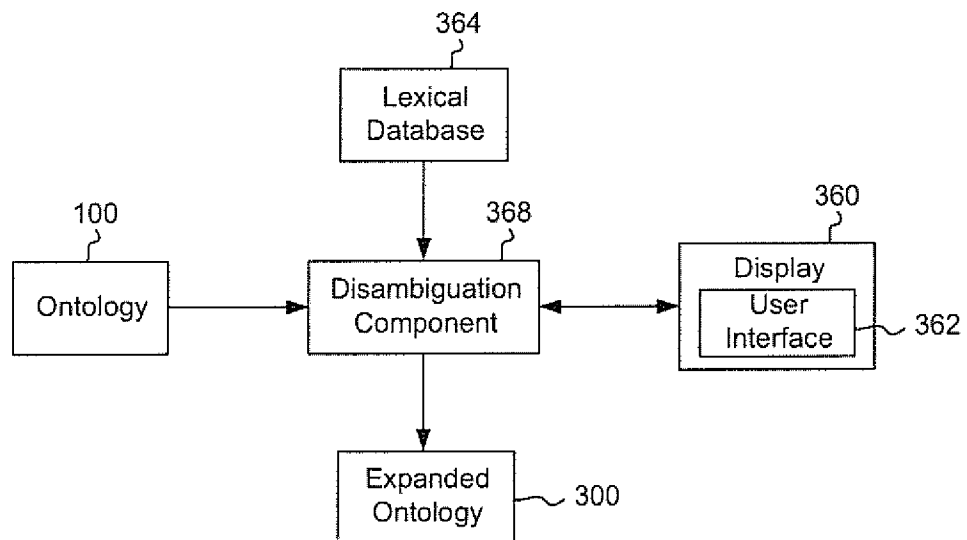


Figure 3A

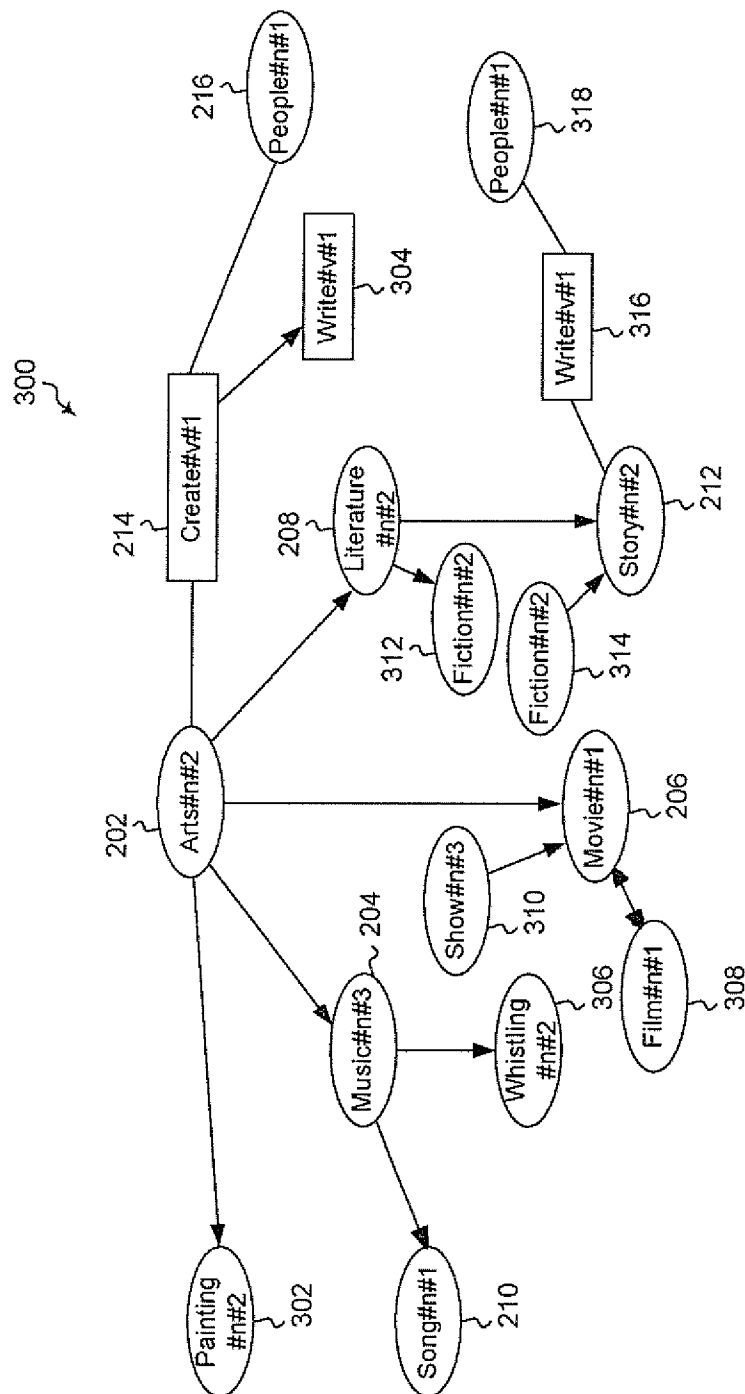


Figure 3B

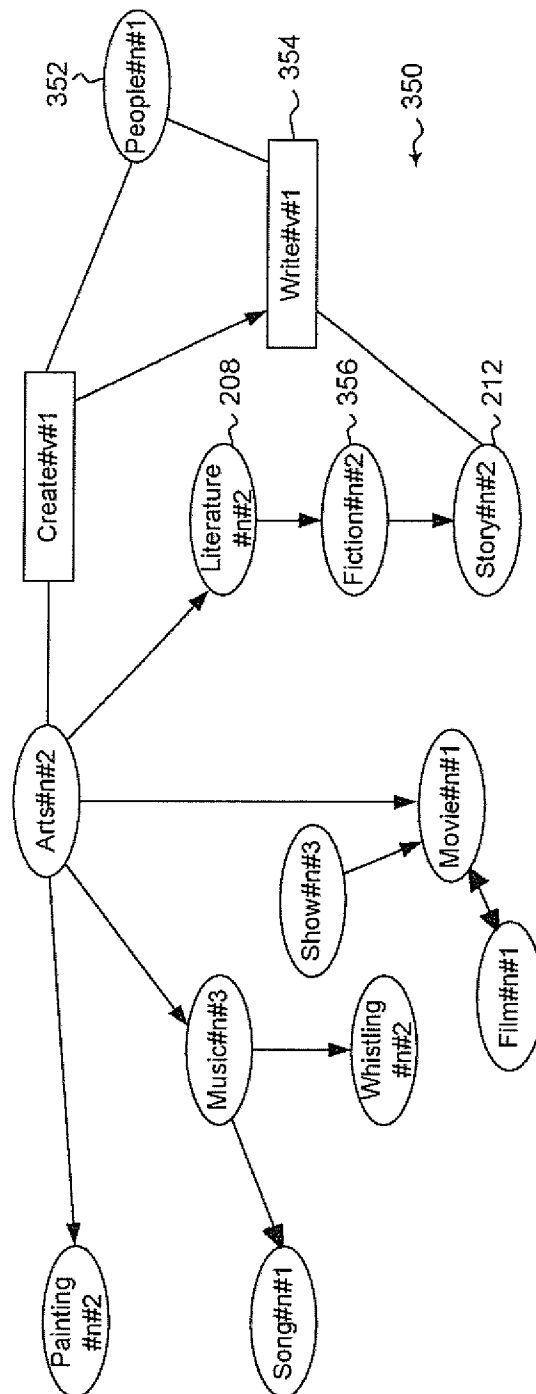
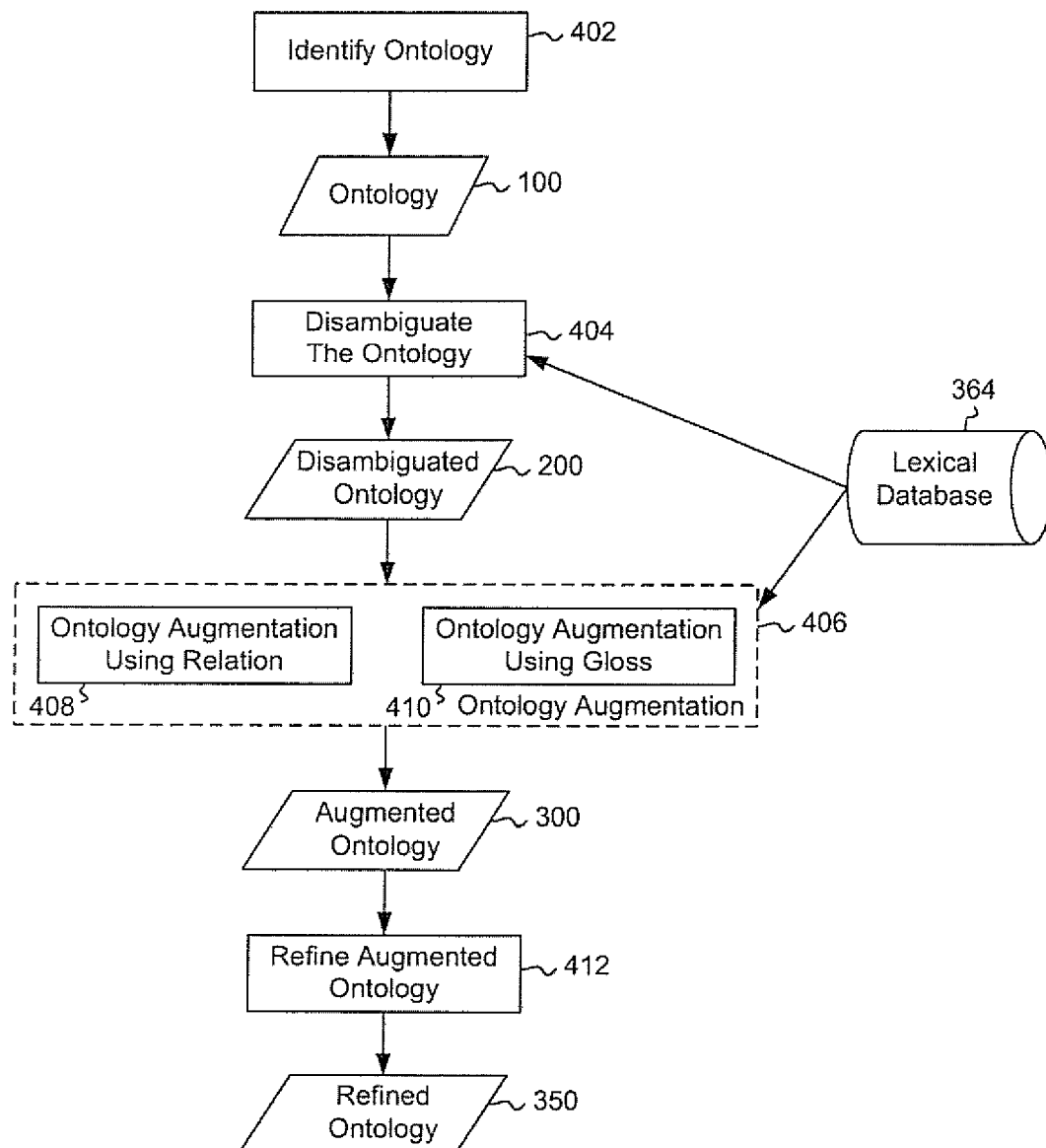


Figure 3C

**Figure 4**

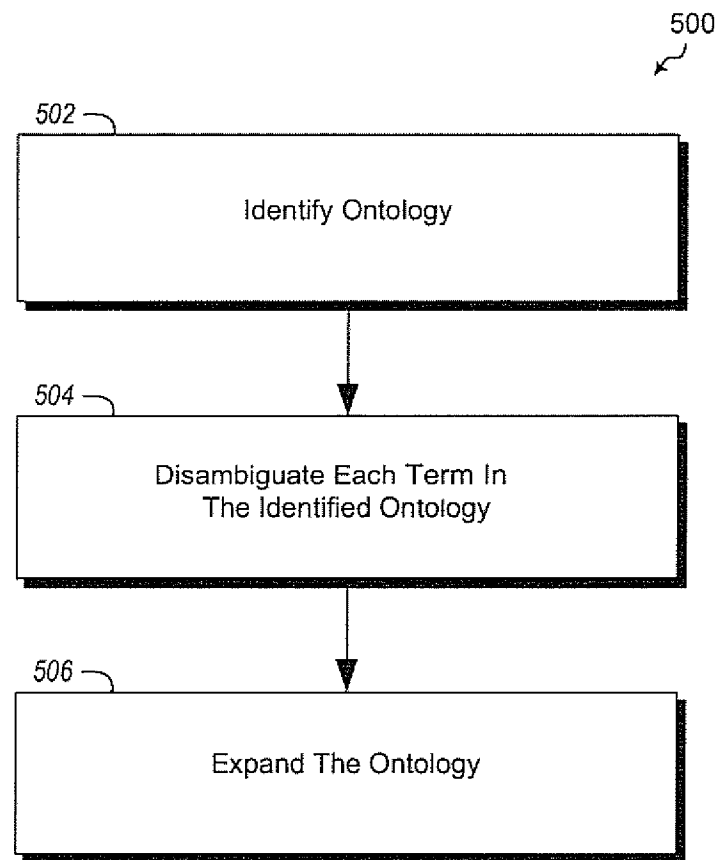
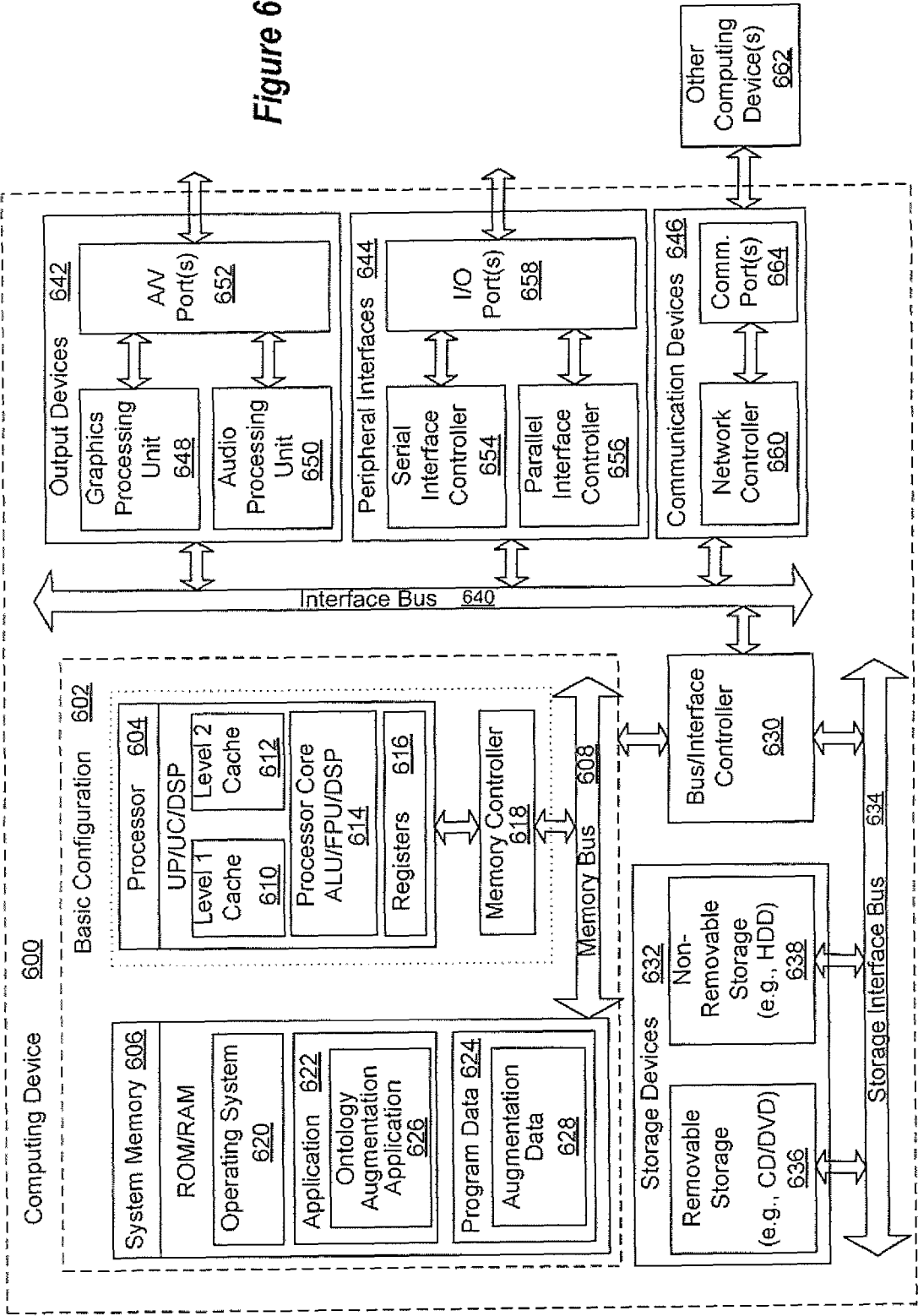


Figure 5

Figure 6



INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2011/071298

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/30 (2006.01)i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC:G06F

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

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

CPRASBS SIPOABS DWPI CNKI(ontology term sense expansion augment identify disambiguate)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 1877566 A(IBM CORP)13 Dec. 2006(13.12.2006)the whole document	1-31
A	CN 101819583 A(FUJITSU LTD)01 Sep. 2010(01.09.2010) the whole document	1-31
A	KR 20090080823 A(KTFREETEL CO LTD)27 Jul. 2009(27.07.2009)the whole document	1-31

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

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

Date of the actual completion of the international search 07 Sep. 2011(07.09.2011)	Date of mailing of the international search report 03 Nov. 2011 (03.11.2011)
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451	Authorized officer WANG Yankun Telephone No. (86-10)62411680

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2011/071298

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN1877566A	13.12.2006	CN1877566B	16.06.2010
		US2008301083A1	04.12.2008
		US7685088B2	23.03.2010
CN101819583A	01.09.2010	US2010217742A1	26.08.2010
		EP2224361A1	01.09.2010
		JP2010211793A	24.09.2010
KR20090080823A	27.07.2009	NONE	