(19) DANMARK (10) DK 2003 01926 L

(12) PATENTANSØGNING

Patent- og
Varemærkestyrelsen

(51) Int.Cl[7].: G 06 F 17/27 G 06 F 17/30

(21) Patentansøgning nr: PA 2003 01926

(22) Indleveringsdag: 2003-12-23

(24) Løbedag: 2003-12-23

(41) Alm. tilgængelig: 2005-06-24

(71) Ansøger: Thomas Eskebæk, Svinget 11, 2. th, 2300 København S, Danmark

(72) Opfinder: Thomas Eskebæk, Svinget 11, 2. th, 2300 København S, Danmark

(54) Benævnelse: Videnstyresystem med ontologibaserede metoder for udledning af viden og søgen efter viden

(57) Sammendrag:

A method of performing knowledge extraction from natural language text documents including the steps of reading an input, using semantic based means for extracting concepts and their interrelations from said input text, transforming said input text into a machine understandable knowledge representation so as to provide knowledge libraries from said documents and optionally storing said libraries The method uses a defined ontology to specify possible semantic relations. Furthermore, the method provides knowledge structures consisting of an arbitrary number of concepts and their interrelations. The method utilises a predetermined mapping between words in the input and concepts and relations in the ontology. These features make a very precise knowledge extraction possible. Thereby, very specific searches can be carried out using the generated knowledge base or library. Additionally, the method is not dependent on the grammatical structures of the sentences in the input text. In that sense, the method for extracting knowledge is language independent.

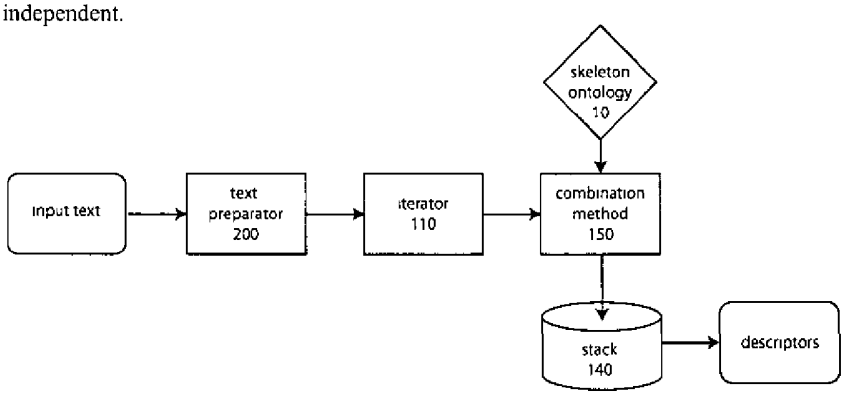

Fig. 1

C L A I M S

---

1. A method of performing knowledge extraction from natural language text documents including the steps of:

- reading an input text;

- using semantic based means for extracting concepts and their interrelations from said input text;

- transforming said input text into a machine understandable knowledge representation so as to provide knowledge libraries from said documents; and

- optionally storing said libraries

characterised by

- using a defined ontology to specify possible semantic relations,

- providing knowledge structures consisting of an arbitrary number of concepts and their interrelations; and

- using a predetermined mapping between words in the input text and concepts and relations in the ontology.

2. A method according to claim 1 **characterised by** using said ontology to control the knowledge extraction process.

3. A method according to claim 1 or 2 **characterised in that** the used ontology being adapted to determine if a concept is a generalisation or a specialisation of another concept.

4. A method according to any of claims 1 to 3 **characterised in that** the used ontology being adapted to specify non-hierarchic relations between concepts.

5. A method according to any of the preceding claims **characterised by** storing of stacks during said knowledge extraction.

6. A method according to claim 5 **characterised by** connecting said stacks to each other.

7. A method according to any of the preceding claims **characterised by** using descriptors to represent the semantic knowledge in pieces of said input text.

8. A method according to claim 7 **characterised in that** a number of descriptors are allowed for given pieces of input text.

9. A method according to claims 7 or 8 **characterised in that** the building of at least one of the descriptors is broken off at the middle of said input text and a new descriptor is created for the remainder of the input text.

10. A method according to any of claims 7 to 9 **characterised by** connecting descriptors from consecutive parts of said input text.

11. A method according to any of claims 7 to 10 **characterised by** defining a pivot point for a descriptor whenever a pivot concept is combined with said descriptor.

12. A method according to any of claims 7 to 11 **characterised by** said descriptors being graded according to a semantic distance scoring.

13 A method according to any of claims 7 to 12 **characterised by** said descriptors being graded according to a sentence distance scoring.

14. A method according to any of claims 7 to 13 **characterised in that** said descriptors being graded according to a linguistic grammar scoring.

15. A method according to any of the preceding claims **characterised by** extracting several adjacent words as a single concept.

16. A method according to any of the preceding claims **characterised by** words and their conceptual meaning being added to the ontology.

17. A method according to claim 16 **characterised by** words and their conceptual meaning automatically being added to the ontology by means of a lexical inquiry.

18. A method according to any of the preceding claims **characterised by** said predetermined mapping allowing for skipping of irrelevant words.

19. A method according to any of the preceding claims **characterised by** using non-word relations.

20 20 A method according to any of the preceding claims **characterised by** skipping at least one sentence element.

21. A method according to any of the preceding claims **characterised by** delaying a sentence element in order to firstly combine with a following sentence element

22. A method according to any of the preceding claims **characterised by** adding a neural network to the knowledge extraction method.

23. A method for iteratively searching knowledge structures **characterised by** performing a conceptual searching by use of said defined ontology.

24. A method according to claim 23 **characterised in that** a knowledge search returns occurrences of a given concept and concepts that it subsumes.

25. A method according to claim 23 or 24 **characterised in that** a knowledge search returns occurrences of a given concept and concepts by which it is subsumed.

26. A method according to any of claims 23 to 25 **characterised in that** a knowledge search for a concept is performed in accordance with, preferably constrained by, a word that instantiates the concept.

27. A method according to any of claims 23 to 26 **characterised in that** a knowledge search includes inverted relations.

28. A method according to any of claims 23 to 27 **characterised in that** a search includes the semantic equivalences of concepts and relations.

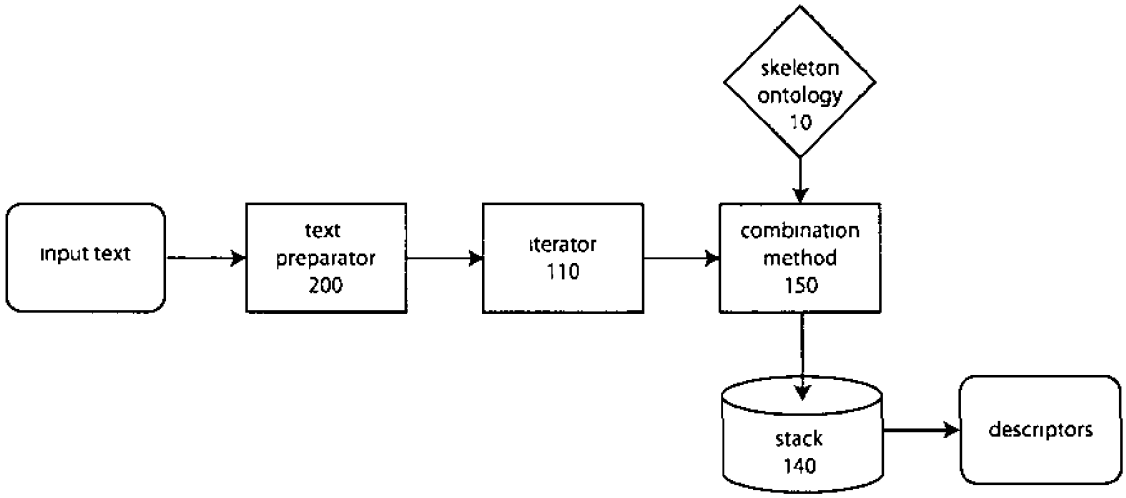
skeleton
ontology
10
input text
text
preparator
200
iterator
110
combination
method
150
stack
140
descriptors

Fig. 1

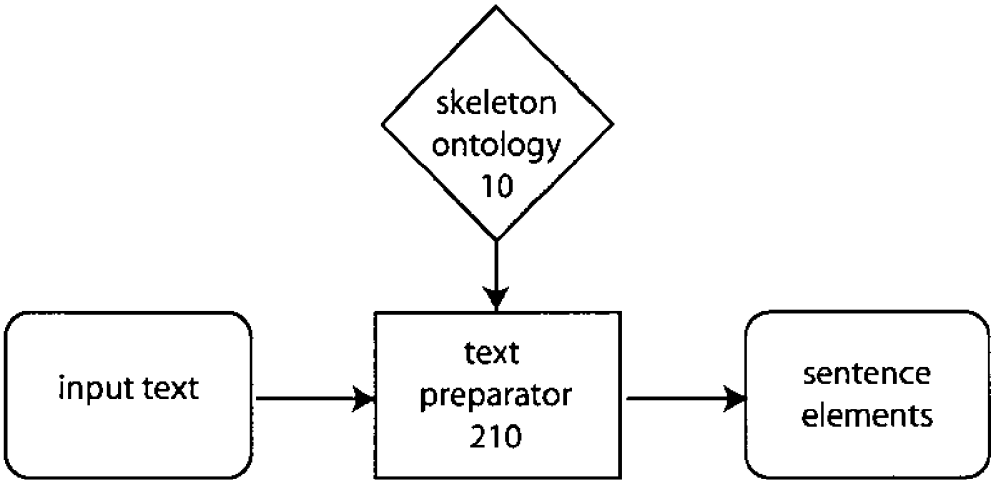
skeleton ontology 10
input text
text preparator 210
sentence elements

Fig. 2

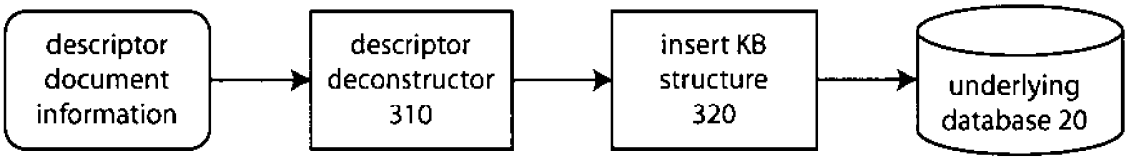
descriptor document information
descriptor deconstructor 310
insert KB structure 320
underlying database 20

Fig. 3

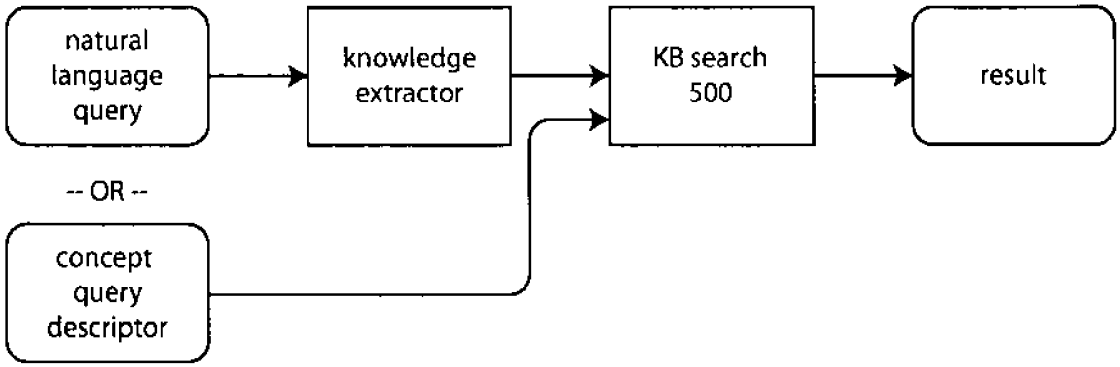
natural language query
knowledge extractor
KB search 500
result
-- OR --
concept query descriptor

Fig. 4

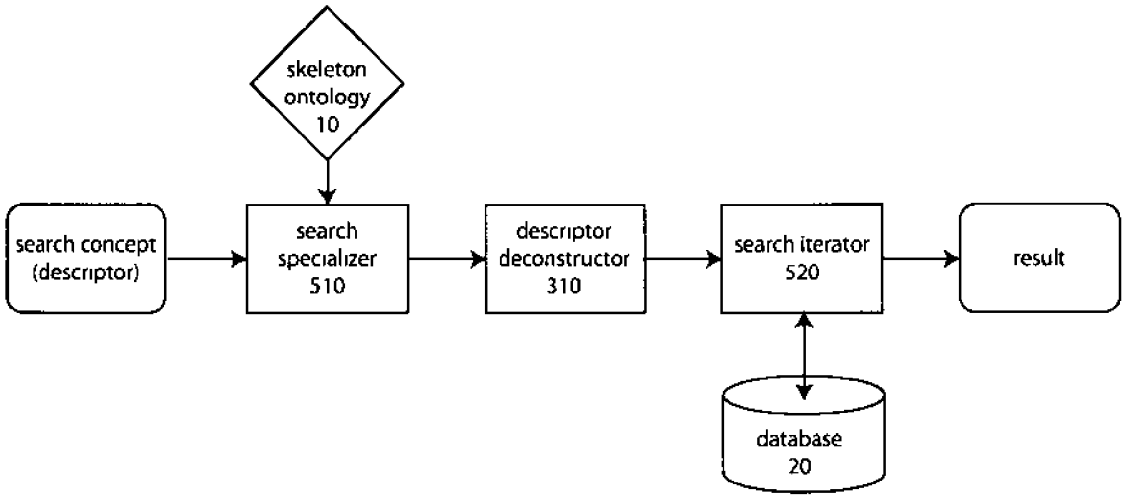
skeleton ontology 10
search concept (descriptor)
search specializer 510
descriptor deconstructor 310
search iterator 520
result
database 20

Fig. 5

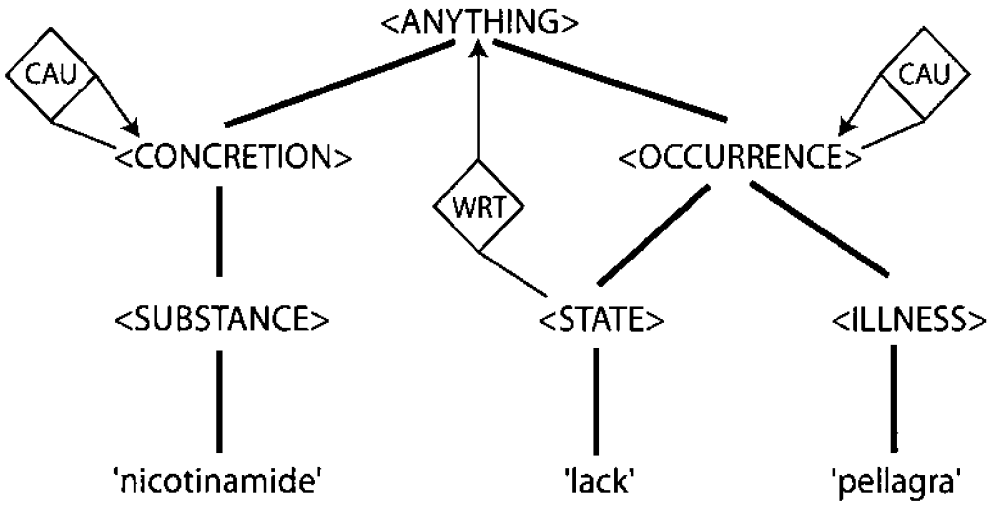
<ANYTHING>
CAU
CAU
<CONCRETION>
WRT
<OCCURRENCE>
<SUBSTANCE>
<STATE>
<ILLNESS>
'nicotinamide'
'lack'
'pellagra'

Fig. 6