CLAIMS

1. An apparatus for creating an index for conducting a search on a database based on a resource including a text sentence, comprising:

a dictionary that includes a plurality of phrases associated with each other for each specific concept;

an input interface unit that receives an input of the resource;

a text processing unit that performs processing for extracting a plurality of words, as tokens, from the text sentence in the received resource;

a directed graph generation unit that generates a directed graph representing a connection relationship between the extracted plurality of tokens in accordance with the text sentence;

a directed graph search unit that conducts a search on the directed graph based on a search-target-phrase and that, if the search-target-phrase is found in the dictionary, locates a location in the directed graph where the search-target-phrase appeared; and

an index creation unit that creates an index indicating association with the resource based on the plurality of tokens in the directed graph,

wherein the directed graph search unit sets at least one token in the directed graph to be the search-target-phrase and refers to the dictionary based on the set search-target-phrase, and

wherein the directed graph generation unit adds to the directed graph, based on the dictionary, a new token that is based on at least one phrase associated with the search-target-phrase in order to update the directed graph.

2. The apparatus according to claim 1, wherein the directed graph generation unit generates the directed graph in which neighboring tokens, in the plurality of tokens, are connected to each other by a node.

3. The apparatus according to claim 2, wherein the directed graph generation unit connects the at least one phrase associated with the search-target-phrase found in the

dictionary between a node in front of and a node behind the appearance location in the directed graph.

4. The apparatus according to claim 1, wherein the directed graph search unit conducts a search on a directed graph updated by the directed graph generation unit.

5. The apparatus according to claim 4, wherein the directed graph search unit conducts a search on the updated directed graph so as to avoid a re-search of a token that has already been searched in the updated directed graph.

6. The apparatus according to claim 5,

wherein the directed graph includes a plurality of nodes that connect neighboring tokens to each other in the plurality of tokens, and

wherein the directed graph search unit stores a node that is associated with a token that has already been searched as a searched node and conducts a re-search on a token that is connected to a node other than the searched node.

7. The apparatus according to claim 1, further comprising a data arrangement that indicates a tree structure arranged based on the plurality of phrases included in the dictionary, the tree structure having an ordered node corresponding to each of the plurality of phrases,

wherein the directed graph search unit conducts a search on the directed graph with reference to the tree structure.

8. The apparatus according to claim 7, wherein the directed graph search unit determines that the search-target-phrase has appeared in the directed graph, if the search-target-phrase matches at least one phrase corresponding to at least one ordered node in the tree structure.

9. The apparatus according to claim 8, wherein the directed graph generation unit

adds to the directed graph a new token that is based on at least one phrase in the dictionary in order to update the directed graph, the at least one phrase being associated with the matched search-target-phrase.

10. The apparatus according to claim 1, wherein the index creation unit extracts an N-gram based on the plurality of tokens in the updated directed graph and creates the index based on the extracted N-gram.

11. The apparatus according to claim 10, wherein the index creation unit extracts a word-based N-gram as the N-gram.

12. A search apparatus, comprising:

an index created by the apparatus according to claim 1;

a resource information file that is related to the resource associated with the index; and

a query server that conducts a search on the index based on a search query, identifies at least one resource from the resource information file based on the result of the search, and outputs a search result indicating the identified at least one resource,

wherein the query server identifies, based on a directed graph being obtained from the identified at least one resource and including a predetermined related phrase that is related to a phrase in the at least one resource, a ground phrase that served as grounds for the at least one resource to be identified and an original phrase, in the at least one resource, that is associated with the ground phrase, and

wherein the search result includes the ground phrase and the original phrase.

13. A method executed by a computing device for creating an index for conducting a search on a database based on a resource including a text sentence, the method comprising, under control of a processor of the computing device:

loading a dictionary including a plurality of phrases associated with each other for each specific concept into a memory accessed by the processor;

receiving an input of the resource via an input interface unit;

performing text-processing for extracting a plurality of words, as tokens, from the text sentence in the received resource;

generating a directed graph representing a connection relationship between the extracted plurality of tokens in accordance with the text sentence in order to expand the directed graph in a memory accessed by the processor;

searching the directed graph with reference to the dictionary based on a search-target-phrase; and

creating an index indicating the association with the resource based on the plurality of tokens in the directed graph in order to register the index in the database,

wherein the searching the directed graph includes:

setting at least one token in the directed graph to be the search-target-phrase;

referring to the dictionary based on the set search-target-phrase; and

if the search-target-phrase is found in the dictionary, locating a location in the directed graph where the search-target-phrase appeared, and

wherein the generation of the directed graph includes:

adding to the directed graph a new token that is based on at least one phrase associated with the search-target-phrase found in the dictionary in order to update the directed graph.

14. The method according to claim 13, wherein the generating the directed graph includes generating the directed graph in which neighboring tokens, in the plurality of tokens, are connected to each other by a node.

15. The method according to claim 14, wherein the generating the directed graph includes connecting the at least one phrase associated with the search-target-phrase found in the dictionary between a node in front of and a node behind the appearance location in the directed graph.

16. The method according to claim 13, wherein the searching the directed graph includes conducting a search on the updated directed graph as a new directed graph to be searched.

17. The method according to claim 13, wherein the creation of the index includes extracting an N-gram based on a plurality of words in the updated directed graph and creating the index based on the extracted N-gram.

18. The method according to claim 17, wherein the creation of the index includes extracting a word-based N-gram as the N-gram.

19. An information search method executed by a search apparatus including a search engine, the method comprising, under control of a processor of the search apparatus:

conducting a search on an index in a database of the search apparatus based on a search query received from a client computer index in order to identify at least one resource from a resource information file based on the result of the search;

expanding a directed graph being obtained from the identified at least one resource and including a predetermined related phrase that is related to a phrase in the at least one resource in a memory accessed by the processor and identifying, based on the directed graph expanded in the memory, a ground phrase that served as grounds for the at least one resource to be identified and an original phrase, in the at least one resource, that is associated with the ground phrase; and

outputting information, as a search result, indicating the at least one resource, the ground phrase and the original phrase in order to transmit the search result to the client computer.

20. A computer program for causing a computing device to implement a method for creating an index for conducting a search on a database based on a resource including a text sentence, the method comprising:

loading a dictionary including a plurality of phrases associated with each other for each specific concept in a memory accessed by the processor of the apparatus;

receiving an input of the resource via an input interface unit;

performing text-processing for extracting a plurality of words, as tokens, from the text sentence in the received resource;

generating a directed graph representing a connection relationship between the extracted plurality of tokens in accordance with the text sentence;

conducting a search on the directed graph with reference to the dictionary based on a search-target-phrase; and

creating an index indicating the association with the resource based on the plurality of tokens in the directed graph,

wherein the conducting the search on the directed graph includes:

setting at least one token in the directed graph to be the search-target-phrase;

referring to the dictionary based on the set search-target-phrase; and

if the search-target-phrase is found in the dictionary, locating a location in the directed graph where the search-target-phrase appeared, and

wherein the generating the directed graph includes:

adding to the directed graph a new token that is based on at least one phrase associated with the search-target-phrase found in the dictionary in order to update the directed graph.