

appearance of an expression such as "Trojan heritage" or "carousel horse" in a negative example group can result in the lower feature degree of an expression such as "Trojan" or "horse" and also in lower detection precision. However, in reality, a negative example group with a sufficient amount is rarely obtained, and such a problem as described above occurs frequently.

[0010]

In PTL 1, a technique in which a word collocating with a feature word is also regarded as a dictionary registration candidate is disclosed; however, an index such as the product of TF (Term Frequency) and IDF (Inverse Document Frequency) is used in determination whether or not to carry out dictionary registration, and it is considered that there is such a problem as described above for plural phrases overlapping each other.

[0011]

As described above, conventional techniques in which a dictionary for monitoring text information is constructed with a feature degree calculated from a positive example group and a negative example group have a problem that lower detection precision is caused.

~~The~~Embodiments of present invention ~~is to solve the problems described above and to provide a dictionary generation device for monitoring text information, a dictionary generation method for monitoring text information, and a dictionary generation program for monitoring text information such that it is possible to achieve high-precision detection compared to the conventional art.~~

[Solution to Problem]

[0013]

~~The present invention which is to solve the problems described above is a dictionary generation device for monitoring text information, which is used in a text information monitoring system and generates a~~

dictionary in which a detection condition is registered, including: a feature degree calculation unit that calculates, for a phrase as a candidate for the detection condition, a feature degree representing a degree by which the phrase matches an information content targeted for monitor; and a phrase usefulness determination unit that determines whether or not the phrase is appropriate for the detection condition based on the feature degree and a usefulness degree representing littleness of ambiguity of meaning defined by the phrase.

[0013a]

According to a first aspect of the present invention, there is provided a dictionary generation device for monitoring text information, that is used in a text information monitoring system and generates a dictionary in which a detection condition is registered, the dictionary generation device comprising: a phrase usefulness determination unit that determines whether or not a phrase as a candidate for the detection condition is appropriate for the detection condition based on a feature degree representing a degree by which the phrase matches an information content targeted for monitor and a usefulness degree representing littleness of ambiguity of meaning defined by the phrase.

[0014]

The present invention which is to solve the problems described above is a method for generating a dictionary used in a text information monitoring system, wherein a dictionary generation device for monitoring text information calculates, for a phrase as a candidate for a detection condition, a feature degree representing a degree by which the phrase matches an information content targeted for monitor; determines whether or not the phrase is appropriate for the detection condition based on the

~~feature degree and a usefulness degree representing littleness of ambiguity of meaning defined by the phrase; and outputs a phrase judged to be appropriate and registers the phrase as the detection condition.~~

[0014a]

5 According to a second aspect of the present invention, there is provided a dictionary generation method for monitoring text information, that is a method for generating a dictionary used in a text information monitoring system, wherein a dictionary generation device for monitoring text information determines whether or not a phrase as a candidate for the
10 detection condition is appropriate for the detection condition based on a feature degree representing a degree by which the phrase matches an information content targeted for monitor and a usefulness degree representing littleness of ambiguity of meaning defined by the phrase; and outputs the phrase judged to be appropriate and registers the phrase as the
15 detection condition.

[0015]

~~The present invention which is to solve the problems described above is a dictionary generation program for monitoring text information, which allows a dictionary generation device for monitoring text~~
20 ~~information to execute a processing of calculating, for a phrase as a candidate for a detection condition, a feature degree representing a degree by which the phrase matches an information content targeted for monitor; a processing of determining whether or not the phrase is appropriate for the detection condition based on the feature degree and a usefulness degree~~
25 ~~representing littleness of ambiguity of meaning defined by the phrase; and~~

~~a processing of outputting a phrase judged to be appropriate and of
registering the phrase as the detection condition.~~

[0015a]

According to a second aspect of the present invention, there is
provided a non-transitory computer-readable storage medium storing a
dictionary generation program for monitoring text information, that causes a
dictionary generation device for monitoring text information to execute a
processing of determining whether or not a phrase as a candidate for the
detection condition is appropriate for the detection condition based on a
feature degree representing a degree by which the phrase matches an
information content targeted for monitor and a usefulness degree
representing littleness of ambiguity of meaning defined by the phrase; and a
processing of outputting the phrase judged to be appropriate and of
registering the phrase as the detection condition.

[Advantageous Effects of Invention]

[0016]

In general, the longer length of a phrase results in the less
ambiguity of meaning and in a higher matching rate for a detection
condition. In embodiments of the present invention, a usefulness degree
is calculated based on the length of a phrase, and a phrase to be registered
in a dictionary is extracted based on the usefulness degree and a feature
degree. In other words, priority is given to a phrase having a longer
length.

[0017]

As a result, there can be generated a dictionary for monitoring text
information such that it is possible to achieve high-precision detection
compared to the conventional art.

[Brief Description of Drawings]

[0018]

[Fig. 1] Fig. 1 is a functional block diagram of a dictionary generation device.

[Fig. 2] Fig. 2 is an operation flow of a dictionary generation device.

[Fig. 3] Fig. 3 is an example of a positive example group and a negative example group (common to the conventional art).

[Fig. 4] Fig. 4 is an example of the frequency and feature degree of each phrase (common to the conventional art).

[Fig. 5] Fig. 5 is an example of the usefulness degree and score of each phrase (Application Example 1).

10 [Fig. 6] Fig. 6 is an example of the usefulness degree and score of each phrase (Application Example 2).

[Fig. 7] Fig. 7 is an example of the usefulness degree and score of each phrase (Application Example 3).