

SEGMENTATION METHOD, SEGMENTATION SYSTEM AND NON-TRANSITORY COMPUTER-READABLE MEDIUM

BACKGROUND

Technical Field

[0001] The present disclosure relates to a segmentation method, a segmentation system and non-transitory computer-readable medium. More particularly, the present application relates to a segmentation method and a segmentation system for captioning and non-transitory computer-readable medium.

Description of Related Art

[0002] "Online Learning Platform" is a network service that stores a large amount of learning content data in a server, allowing users to connect to the server through the Internet to view the learning content data at any time. At present, the types of learning content data provided by various online learning platforms include videos, audio, presentation, documents or discussion forums, and the like.

[0003] Since the amount of learning content data stored in the online learning platform is very large, in order to be convenient for using, it is necessary to automatically segment the text of the learning content data and create paragraph keywords. Therefore, a system for having a function of segmentation the similar topics and mark the keywords in the learning video is required.

SUMMARY

[0004] An aspect of the present disclosure is a segmentation method. The segmentation method includes operations of: receiving a captioning information, wherein the captioning information comprises a plurality of captioning sentences; selecting the captioning sentences according to a default value and dividing the selected captioning sentences into a first paragraph; performing a common segmentation vocabulary determination for a first captioning sentence, wherein the first captioning sentence is one of the plurality of captioning sentences; and generating a second paragraph or merging the first captioning sentence into the first paragraph according to a determination result of the common segmentation vocabulary determination.

[0005] Another aspect of the present disclosure is a segmentation system. The segmentation system includes a storage unit and a processor. The storage unit is configured to store a captioning information, a segmentation result, a common segmentation vocabulary database, a comment corresponding to first paragraph and a comment corresponding to second paragraph. The processor is electrically connected to the storage unit, and is configured to receive the captioning information, wherein the captioning information comprises a plurality of captioning sentences. The processor further includes a segmentation unit, a common vocabulary detecting unit and a paragraph generating unit. The segmentation unit is configured to select the captioning sentences according to a default value and divide the selected captioning sentences into a first paragraph. The common

vocabulary detecting unit is electrically connected to the segmentation unit, and is configured to perform a common segmentation vocabulary determination for a first captioning sentence, wherein the first captioning sentence is one of the plurality of captioning sentences. The paragraph generating unit is electrically connected to the common vocabulary detecting unit, and is configured to generate a second paragraph or merge the first captioning sentence into the first paragraph according to a determination result of the common segmentation vocabulary determination.

[0006] Another aspect of the disclosure is to provide a non-transitory computer-readable medium including one or more sequences of instructions to be executed by a processor for performing a segmentation method, wherein the segmentation method includes operations of: receiving a captioning information, wherein the captioning information comprises a plurality of captioning sentences; selecting the captioning sentences according to a default value and dividing the selected captioning sentences into a first paragraph; performing a common segmentation vocabulary determination for a first captioning sentence, wherein the first captioning sentence is one of the plurality of captioning sentences; and generating a second paragraph or merging the first captioning sentence into the first paragraph according to a determination result of the common segmentation vocabulary determination.

[0007] Accordingly, since the segmentation method, the segmentation system, and non-transitory computer-readable medium for calculating the keywords corresponding to each captioning sentence, and performing a

common segmentation vocabulary determination for the captioning sentence, and then generating a second paragraph or merging the captioning sentence into the first paragraph according to a determination result of the common segmentation vocabulary determination to generate a segmentation result. Therefore, the segmentation system is able to have a function of segmentation the similar topics and mark the keywords in the learning video.

[0008] It is to be understood that both the foregoing general description and the following detailed description are by examples, and are intended to provide further explanation of the disclosure as claimed.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] The present disclosure can be more fully understood by reading the following detailed description of the embodiment, with reference made to the accompanying drawings as follows:

[0010] FIG. 1 is a functional block diagram illustrating a segmentation system according to an embodiment of the disclosure.

[0011] FIG. 2 is a flow diagram illustrating a segmentation method according to an embodiment of this disclosure.

[0012] FIG. 3 is a flow diagram illustrating step S240 according to an embodiment of this disclosure.

[0013] FIG. 4 is a flow diagram illustrating step S241 according to an embodiment of this disclosure.

[0014] FIG. 5 is a flow diagram illustrating step S242 according to an embodiment of this disclosure.

DETAILED DESCRIPTION

[0015] It will be understood that, in the description herein and throughout the claims that follow, when an element is referred to as being "connected" or "coupled" to another element, it can be directly connected or coupled to the other element or intervening elements may be present. In contrast, when an element is referred to as being "directly connected" or "directly coupled" to another element, there are no intervening elements present. Moreover, "electrically connect" or "connect" can further refer to the interoperation or interaction between two or more elements.

[0016] It will be understood that, in the description herein and throughout the claims that follow, although the terms "first," "second," etc. may be used to describe various elements, these elements should not be limited by these terms. These terms are only used to distinguish one element from another. For example, a first element could be termed a second element, and, similarly, a second element could be termed a first element, without departing from the scope of the embodiments.

[0017] It will be understood that, in the description herein and throughout the claims that follow, the terms "comprise" or "comprising," "include" or "including," "have" or "having," "contain" or "containing" and the like used herein are to be understood to be open-ended, i.e., to mean including but not limited to.

[0018] It will be understood that, in the description herein and throughout the claims that follow, the phrase "and/or" includes any and all combinations of one or more of the associated listed items.

[0019] It will be understood that, in the description herein and throughout the claims that follow, words indicating direction used in the description of the following embodiments, such as "above," "below," "left," "right," "front" and "back," are directions as they relate to the accompanying drawings. Therefore, such words indicating direction are used for illustration and do not limit the present disclosure.

[0020] It will be understood that, in the description herein and throughout the claims that follow, unless otherwise defined, all terms (including technical and scientific terms) have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. It will be further understood that terms, such as those defined in commonly used dictionaries, should be interpreted as having a meaning that is consistent with their meaning in the context of the relevant art and will not be interpreted in an idealized or overly formal sense unless expressly so defined herein.

[0021] Any element in a claim that does not explicitly state "means for" performing a specified function, or "step for" performing a specific function, is not to be interpreted as a "means" or "step" clause as specified in 35 U.S.C. § 112(f). In particular, the use of "step of" in the claims herein is not intended to invoke the provisions of 35 U.S.C. § 112(f).

[0022] Reference is made to FIG. 1, which is a functional block diagram illustrating a segmentation system 100 according to an embodiment of the

disclosure. As shown in FIG. 1, the segmentation system 100 includes a storage unit 110 and a processor 130. The storage unit 110 is electrically connected to the processor 130, and the storage unit 110 is configured to store a captioning information, a segmentation result, a common segmentation vocabulary database DB1, a course database DB2, a comment corresponding to first paragraph and a comment corresponding to second paragraph.

[0023] Afterwards, the processor 130 includes a keyword extraction unit 131, a segmentation unit 132, a common vocabulary detecting unit 133, a paragraph generating unit 134 and a comment generating unit 135. The segmentation unit 132 is electrically connected to the keyword extraction unit 131 and the common vocabulary detecting unit 133. The paragraph generating unit 134 is electrically connected to the common vocabulary detecting unit 133 and the comment generating unit 135. The common vocabulary detecting unit 133 is electrically connected to the comment generating unit 135.

[0024] In the embodiment, the storage unit 110 can be implemented by a memory, hardware, flash drive, memory card and etc. The processor 130 can be implemented by a micro controller, a microprocessor, a digital signal processor, an application specific integrated circuit (ASIC), a logical circuitry or any equivalent circuits of the aforesaid components.

[0025] Reference is made to FIG. 2, which is a flow diagram illustrating a segmentation method 200 according to an embodiment of this disclosure. In the embodiment, the segmentation method 200 shown in FIG. 2 can be

applied to the segmentation system 100 of FIG. 1, and the processor 130 is configured to segment the captioning information to generate a segmentation result and the comment corresponding to each paragraph according to the steps described in the following segmentation method 200. As shown in FIG. 2, the segmentation method 200 firstly executes step S210 to receive a captioning information. In the embodiment, the captioning information includes a plurality of captioning sentences. For example, the captioning information can be realized as a captioning file of video. The captioning file of video has been divided into multiple captioning sentences according to the playing time of the video, and the captioning sentences can be sort by the playing time.

[0026] Afterwards, the segmentation method 200 executes step S220 to select the captioning sentences according to a default value and divide the selected captioning sentences into a current paragraph. In the embodiment, the default value can be realized as any positive integer. The default value is set to three as an example, therefore, in this step three captioning sentences have been selected to compose the current paragraph according to the playing time of video. For example, if there are a total of N captioning sentences, the first captioning sentence to the third captioning sentence have been selected to compose the current paragraph.

[0027] Afterwards, the segmentation method 200 executes step S230 to perform a common segmentation vocabulary determination for the current captioning sentence. In the embodiment, the common segmentation vocabularies are storaged in the common segmentation vocabulary database

DB1, and the common vocabulary detecting unit 133 is configured to detect whether the common segmentation vocabularies are appeared in the current captioning sentence. The common segmentation vocabularies include the opening segmentation vocabularies and the ending segmentation vocabularies. For example, the opening segmentation vocabularies can be realized as “then”, “start explaining”, and etc., and the ending segmentation vocabularies cab can be realized as “the description is end of here”, “take a break”, and etc. In this step, it will detect whether the common segmentation vocabularies are appeared in the current captioning sentence and the type of the the common segmentation vocabularies(the opening segmentation vocabularies or the ending segmentation vocabularies).

[0028] Afterwards, the segmentation method 200 executes step S240 to generate a next paragraph or merge the current captioning sentence into the current paragraph according to a determination result of the common segmentation vocabulary determination. In the embodiment, it will determine to generate a new paragraph or merge the current captioning sentence into the current paragraph according to the result detected by the common vocabulary detecting unit 133. For example, the current paragraph is composed by the the first captioning sentence to the third captioning sentence, and the current captioning sentence is the fourth captioning sentence, and then, according to the determination result the fourth captioning sentence can be merged into the current paragraph or the fourth captioning sentence can be utilized as the opening of the new paragraph.

[0029] Afterwards, when the step S240 is performed to merge the current caption sentence into the current paragraph, the common segment vocabulary determination of the next captioning sentence is executed, and thus the determination of step S230 is performed again. For example, if the fourth captioning sentence is merged into the current paragraph, then the common segment vocabulary determination of the fifth captioning sentence will be executed. If the step S240 is performed to generate a next paragraph, it will be utilized the default value to select the captioning sentences according to the specific order and selected captioning sentence will be divided into the next paragraph, and thus the operation of step S220 is performed again. For example, if the fourth captioning sentence is divided into the next paragraph, the fifth captioning sentence, the sixth captioning sentence and the seventh captioning sentence will be reselected to merge into the next paragraph. Therefore, the operation of segmentation is repeated until all of the captioning sentences is divided, and the segmentation result is generated.

[0030] Afterwards, the step S240 further includes steps S241~S242, reference is made to FIG. 3, which is a flow diagram illustrating step S240 according to an embodiment of this disclosure. As the embodiment shown in FIG. 3, the segmentation method 200 executes step S241 if the current captioning sentence is associated with the common segmentation vocabulary, performing a segmentation processing to generate the next paragraph and utilizing the default value to select captioning sentences according to the a specific order, and merging the selected captioning sentences into the next paragraph. Wherein the step S241 further includes

steps S2411~S2413, reference is made to FIG. 4, which is a flow diagram illustrating step S241 according to an embodiment of this disclosure. As shown in FIG. 4, the segmentation method 200 executes step S2411 to determine whether the current captioning sentence is associated with one of an opening segmentation vocabulary and an ending segmentation vocabulary according to the determination result. Based on aforesaid embodiment, according to the determination result of the step S230, it may be determined whether the current captioning sentence is associated with the opening segmentation vocabulary or the ending segmentation vocabulary.

[0031] Afterwards, the segmentation method 200 executes step S2412 if the current captioning sentence is associated with the opening segmentation vocabulary, utilizing the current captioning sentence as opening sentence of the next paragraph. For example, if the vocabulary of "then" in the fourth captioning sentence is detected in the aforesaid determination operation, the fourth captioning sentence is utilized as the opening sentence of the next paragraph.

[0032] Afterwards, the segmentation method 200 executes step S2413 if the current captioning sentence is associated with the ending segmentation vocabulary, utilizing the current captioning sentence as ending sentence of the current paragraph. For example, if the vocabulary of "the description is end of here" in the fourth captioning sentence is detected in the aforesaid determination operation, the fourth captioning sentence is utilized as the ending sentence of the current paragraph. After the step S241 is performing, it will be utilized the default value to select the captioning

sentences according to the specific order and selected captioning sentence will be divided into the next paragraph, and thus the operation of step S220 is performed again. For the sake of brevity, those descriptions will not be repeated here.

[0033] Afterwards, the segmentation method 200 executes step S242 if the current captioning sentence is not associated with the common segmentation vocabulary, performing a similarity calculation between the current captioning sentence and the current paragraph, if the current captioning sentence is similar with the current paragraph, merging the current captioning sentence into the current paragraph. Wherein the step S242 further includes steps S2421~S2423, reference is made to FIG. 5, which is a flow diagram illustrating step S242 according to an embodiment of this disclosure. As shown in FIG. 5, the segmentation method 200 further executes step S2421 to compare whether a difference value between at least one feature corresponding to the current captioning sentence and at least one feature corresponding to the current paragraph is greater than a threshold value.

[0034] Afterwards, in the embodiment, the at least one feature corresponding to the current paragraph is generated by the plurality of keywords extracted from the captioning sentences. It is utilized the term frequency-inverse document frequency (TF-IDF) to calculate the keywords corresponding to the current captioning sentence. The term frequency-inverse document frequency is intended to evaluate how important a vocabulary is to a document in the database. The importance of a vocabulary increases proportionally with the number of times it appears in the file, but it also

decreases inversely with the frequency it appears in the database. In the embodiment, the term frequency–inverse document frequency can be utilized to calculate the keywords of the current captioning sentence. Afterwards, it will calculate a similarity value between at least one feature (keywords) corresponding to the current captioning sentence and at least one feature (keywords) corresponding to the current paragraph. If the calculated similarity value is higher, it can be determined that the content of the current captioning sentence is closer to the content of the current paragraph.

[0035] Afterwards, the segmentation method 200 executes step S2422 if the difference value is less than the threshold value, merging the current captioning sentence into the current paragraph. In the embodiment, it is utilized the threshold value to filter the similarity value, when the similarity value is not less than the threshold value, it will respresented that the content of the current captioning sentence is similar with the content of the current paragraph, and thus the current captioning sentence is merged into the current paragraph. For example, if the similarity value of the fourth captioning sentence and the similarity value of the current paragraph is not less than the threshold value, it is respresented that the content of the fourth captioning sentence is similar with the content of the current paragraph, and thus the fourth captioning sentence can be merged into the current paragraph.

[0036] Afterwards, the segmentation method 200 executes step S2423 if the difference value is not less than the threshold value, utilizing the current captioning sentence as opening sentence of the next paragraph, and utilizing

the default value to select the captioning sentences according to the specific order and merging the selected captioning sentences into the next paragraph. For example, if the similarity value of the fourth captioning sentence and the similarity value of the current paragraph is less than the threshold value, it is represented that the content of the fourth captioning sentence is different from the content of the current paragraph, and thus the fourth captioning sentence can be utilized as opening sentence of the next paragraph. After the step S242 is performed, it will be utilized the default value to select the captioning sentences according to the specific order and selected captioning sentence will be divided into the next paragraph, and thus the operation of step S220 is performed again. For the sake of brevity, those descriptions will not be repeated here.

[0037] According to the aforesaid segmentation operation, the operation of segmentation is repeated until all of the captioning sentences is divided. If the number of the remaining captioning sentences is less than the default value, the segmentation calculation may not be performed for the remaining captioning sentences, and then the remaining captioning sentences may be directly merged into the current paragraph. For example, if the number of the remaining captioning sentences is two, it is less than the aforesaid default value (the aforesaid default value is three), and thus the remaining captioning sentences will be merged into the current paragraph.

[0038] Afterwards, after the aforesaid step is performed, the segmentation method 200 executes step S250 to generate the comment corresponding to each paragraph. For example, if it is divided into three paragraphs after

executing all of the captioning sentences, it will be calculated the comment corresponding to three paragraph, respectively. Wherein, the comment can be generated according to the keywords corresponding to the captioning sentences in paragraph. Finally, the paragraph and the common corresponding the paragraph are stored in the course database DB2 of the storage unit 110. For example, if the difference value is less than the threshold value, it can be realized that the current captioning sentence is similar with the current paragraph, and thus the keywords of the current captioning sentence can be utilized as the at least one feature corresponding to the current paragraph. If the difference value is not less than the threshold value, it can be realized that the current captioning sentence is difference from the current paragraph, and thus the keywords of the current captioning sentence can be utilized as the at least one feature corresponding to the next paragraph.

[0039] Based on aforesaid embodiments, the segmentation method, the segmentation system, and non-transitory computer-readable medium are capable of calculating the keywords corresponding to each captioning sentence; and performing a common segmentation vocabulary determination for the captioning sentence, and then generating a next paragraph or merging the current captioning sentence into the current paragraph according to a determination result of the common segmentation vocabulary determination to generate a segmentation result. In some embodiments, the segmentation system is able to have a function of segmentation the similar topics and mark the keywords in the learning video.

[0040] The foregoing outlines features of several embodiments so that those skilled in the art may better understand the aspects of the present disclosure. Those skilled in the art should appreciate that they may readily use the present disclosure as a basis for designing or modifying other processes and structures for carrying out the same purposes and/or achieving the same advantages of the embodiments introduced herein. Those skilled in the art should also realize that such equivalent constructions do not depart from the spirit and scope of the present disclosure, and that they may make various changes, substitutions, and alterations herein without departing from the spirit and scope of the present disclosure.

WHAT IS CLAIMED IS:

1. A segmentation method, comprising:

receiving a captioning information, wherein the captioning information comprises a plurality of captioning sentences;

selecting the plurality of captioning sentences according to a default value and dividing the selected captioning sentences into a first paragraph;

performing a common segmentation vocabulary determination for a first captioning sentence to detect whether a common segmentation vocabulary is appeared in the first captioning sentence, wherein the first captioning sentence is one of the plurality of captioning sentences; and

generating a second paragraph or merging the first captioning sentence into the first paragraph according to a determination result of the common segmentation vocabulary determination,

wherein the common segmentation vocabulary comprises an opening segmentation vocabulary and an ending segmentation vocabulary.

2. The segmentation method of claim 1, wherein when the first captioning sentence is merged into the first paragraph, performing the common segmentation vocabulary determination for a second captioning sentence, wherein the second captioning sentence is followed by the first captioning sentence in a specific order.

3. The segmentation method of claim 1, wherein when the second paragraph is generated, utilizing the default value to select the plurality of captioning sentences according to a specific order, and merging the selected captioning sentences into the second paragraph.

4. The segmentation method of claim 1, wherein generating the second paragraph or merging the first captioning sentence into the first paragraph according to the determination result of the common segmentation vocabulary determination, and the segmentation method further comprising:

if the first captioning sentence is associated with the common segmentation vocabulary, performing a segmentation processing to generate the second paragraph and utilizing the default value to select the plurality of captioning sentences according to a specific order, and merging the selected captioning sentences into the second paragraph; and

if the first captioning sentence is not associated with the common segmentation vocabulary, performing a similarity calculation between the first captioning sentence and the first paragraph, if the first captioning sentence is similar with the first paragraph, merging the first captioning sentence into the first paragraph.

5. The segmentation method of claim 4, wherein the segmentation processing further comprising:

determining whether the first captioning sentence is associated with the opening segmentation vocabulary and the ending segmentation vocabulary according to the determination result;

if the first captioning sentence is associated with the opening segmentation vocabulary, utilizing the first captioning sentence as opening sentence of the second paragraph; and

if the first captioning sentence is associated with the ending segmentation vocabulary, utilizing the first captioning sentence as ending sentence of the first paragraph.

6. The segmentation method of claim 4, wherein the similarity calculation further comprising:

comparing whether a difference value between at least one feature corresponding to the first captioning sentence and at least one feature corresponding to the first paragraph is greater than a threshold value;

if the difference value is less than the threshold value, merging the first captioning sentence into the first paragraph; and

if the difference value is larger than the threshold value, utilizing the first captioning sentence as opening sentence of the second paragraph, and utilizing the default value to select the plurality of captioning sentences according to the specific order.

7. The segmentation method of claim 6, wherein extracting a plurality of keywords from the plurality of captioning sentences, and utilizing the plurality of keywords as the at least one feature corresponding to the first captioning sentence.

8. The segmentation method of claim 7, wherein the at least one feature corresponding to the first paragraph is generated by the plurality of keywords extracted from the plurality of captioning sentences.

9. A segmentation system, comprising:

a storage unit, configured to store a captioning information, a segmentation result, a common segmentation vocabulary database, a comment corresponding to a first paragraph and a comment corresponding to a second paragraph; and

a processor, electrically connected to the storage unit, and configured to receive the captioning information, wherein the captioning information comprises a plurality of captioning sentences, the processor further comprises:

a segmentation unit, configured to select the plurality of captioning sentences according to a default value and divide the selected captioning sentences into the first paragraph;

a common vocabulary detecting unit, electrically connected to the segmentation unit, and configured to perform a common segmentation vocabulary determination for a first captioning sentence to detect whether a common segmentation vocabulary is appeared in the first captioning sentence, wherein the first captioning sentence is one of the plurality of captioning sentences; and

a paragraph generating unit, electrically connected to the common vocabulary detecting unit, and configured to generate the second paragraph or merge the first captioning sentence into the first paragraph according to a determination result of the common segmentation vocabulary determination,

wherein the common segmentation vocabulary comprises an opening segmentation vocabulary and an ending segmentation vocabulary.

10. The segmentation system of claim 9, wherein when the first captioning sentence is merged into the first paragraph, performing the common segmentation vocabulary determination for a second captioning sentence,

wherein the second captioning sentence is followed by the first captioning sentence in a specific order.

11. The segmentation system of claim 9, wherein when the second paragraph is generated, utilizing the default value to select the plurality of captioning sentences according to a specific order, and merging the selected captioning sentences into the second paragraph.

12. The segmentation system of claim 9, wherein the paragraph generating unit is further configured for executing following steps:

if the first captioning sentence is associated with the common segmentation vocabulary, performing a segmentation processing to generate the second paragraph and utilizing the default value to select the plurality of captioning sentences according to a specific order, and merging the selected captioning sentences into the second paragraph; and

if the first captioning sentence is not associated with the common segmentation vocabulary, performing a similarity calculation between the first captioning sentence and the first paragraph, if the first captioning sentence is similar with the first paragraph, merging the first captioning sentence into the first paragraph.

13. The segmentation system of claim 12, wherein the segmentation processing further comprising:

determining whether the first captioning sentence is associated with the opening segmentation vocabulary and the ending segmentation vocabulary according to the determination result;

if the first captioning sentence is associated with the opening segmentation vocabulary, utilizing the first captioning sentence as opening sentence of the second paragraph; and

if the first captioning sentence is associated with the ending segmentation vocabulary, utilizing the first captioning sentence as ending sentence of the first paragraph.

14. The segmentation system of claim 12, wherein the similarity calculation further comprising:

comparing whether a difference value between at least one feature corresponding to the first captioning sentence and at least one feature corresponding to the first paragraph is greater than a threshold value;

if the difference value is less than the threshold value, merging the first captioning sentence into the first paragraph; and

if the difference value is larger than the threshold value, utilizing the first captioning sentence as opening sentence of the second paragraph, and utilizing the default value to select the plurality of captioning sentences according to the specific order.

15. The segmentation system of claim 14, further comprising:

a keyword extraction unit, electrically connected to the segmentation unit, and configured to extract a plurality of keywords from the plurality of captioning

sentences, and utilize the plurality of keywords as the at least one feature corresponding to the first captioning sentence.

16. The segmentation system of claim 15, wherein the at least one feature corresponding to the first paragraph is generated by the the plurality of keywords extracted from the plurality of captioning sentences.

17. A non-transitory computer-readable medium including one or more sequences of instructions to be executed by a processor for performing a segmentation method, wherein the segmentation method comprises:

receiving a captioning information, wherein the captioning information comprises a plurality of captioning sentences;

selecting the plurality of captioning sentences according to a default value and dividing the selected captioning sentences into a first paragraph;

performing a common segmentation vocabulary determination for a first captioning sentence to detect whether a common segmentation vocabulary is appeared in the first captioning sentence, wherein the first captioning sentence is one of the plurality of captioning sentences; and

generating a second paragraph or merging the first captioning sentence into the first paragraph according to a determination result of the common segmentation vocabulary determination,

wherein the common segmentation vocabulary comprises an opening segmentation vocabulary and an ending segmentation vocabulary.