CLAIMS:

1. A multimedia processing method comprising:

obtaining a first multimedia resource;

determining an initial text content corresponding to the first multimedia resource by performing speech recognition on audio data of the first multimedia resource, wherein the audio data of the first multimedia resource comprises speech data of the initial text content;

determining an invalid text content in the initial text content, wherein the invalid text content is semantically non-informative;

determining a first playing position of the speech data of the invalid text content in the first multimedia resource; and

cropping the first multimedia resource based on the first playing position to obtain a second multimedia resource, wherein audio data of the second multimedia resource comprises speech data of a target text content but does not comprise speech data of the invalid text content, wherein the target text content is other text content in the initial text content than the invalid text content,

wherein determining an initial text content corresponding to the first multimedia resource by performing speech recognition on audio data of the first multimedia resource comprises:

performing, according to an automatic speech recognition ASR technique, a speech recognition on the audio data of the first multimedia resource to obtain a plurality of speech words and a timestamp corresponding to each of the speech words, the timestamp denoting a playing position of audio data corresponding to the speech word in the first multimedia resource; and

generating the initial text content according to the speech words; and

wherein determining the invalid text content in the initial text content comprises:

obtaining a duration of each of the speech words according to a timestamp corresponding to the speech word; and

determining, according to a preset standard duration and a duration of the speech word, a speech word with a duration greater than the standard duration by a first duration threshold or a speech word with a duration less than the standard duration by a second duration threshold as the invalid text content in the initial text content.

2. The method of claim 1, wherein determining an invalid text content in the initial text content comprises:

performing a semantic analysis on the initial text content to obtain semantic information corresponding to the initial text content; and

determining the invalid text content in the initial text content according to the semantic information.

3. The method of claim 2, wherein the initial text content comprises a plurality of segment content, and determining the invalid text content in the initial text content according to the semantic information comprises:

determining a credibility coefficient of at least one segment content of the initial text content according to the semantic information corresponding to the initial text content, wherein the credibility coefficient is used for denoting credibility that the segment content is the invalid text content;

determining at least one invalid segment content from the at least one segment content according to the credibility coefficient of the segment content and a preset credibility threshold; and

determining the invalid text content in the initial text content according to the at least one invalid segment content.

4. The method of claim 3, wherein determining a first playing position of the speech data of the invalid text content in the first multimedia resource comprises:

determining a start point and an end point of the speech data of each invalid segment content in the audio data of the first multimedia resource; and

determining the first playing position of the speech data of the invalid text content in the first multimedia resource based on the start point and the end point corresponding to each invalid segment content.

5. The method of claim 3, wherein cropping the first multimedia resource based on the first playing position to obtain a second multimedia resource comprises:

obtaining other text content in the initial text content than the invalid segment content based on the first playing position, wherein the other text content comprises at least one target segment content;

adding a fade-in effect at a start point of speech data corresponding to the at least one target segment content, and/or adding a fade-out effect at an end point of the speech data corresponding to the at least one target segment content, to generate transition speech data corresponding to the target segment content; and

spliceing the transition speech data according to the first playing position to generate the second multimedia resource.

6. The method of claim 2, wherein before performing the semantic analysis on the initial text content to obtain the semantic information corresponding to the initial text content, the method further comprises:

determining the invalid text content in the initial text content based on a preset invalid text content library;

performing a semantic analysis on the initial text content to obtain semantic information corresponding to the initial text content comprises:

performing a semantic analysis on the initial text content to obtain the semantic information corresponding to the initial text content if the invalid text content does not exist in the invalid text content library;

after determining the invalid text content in the initial text content based on the semantic information, the method further comprises:

adding the invalid text content determined based on the semantic information to the invalid text content library.

7. The method of claim 1, wherein before cropping the first multimedia resource based on the first playing position to obtain the second multimedia resource, further comprises:

displaying the invalid text content in the initial text content; and

playing an audio segment corresponding to the invalid text content in response to an operation instruction for the invalid text content.

8. The method of claim 1, wherein determining an initial text content corresponding to the first multimedia resource by performing speech recognition on audio data of the first multimedia resource comprises:

recognizing speech data and non-speech data in the audio data of the first multimedia resource through a voice activity detection VAD algorithm; and

determining the initial text content corresponding to the first multimedia resource by performing a speech recognition on the speech data in the audio data of the first multimedia resource.

9. The method of claim 8, wherein the method further comprises:

determining a second playing position of the non-speech data in the first multimedia resource according to a start point and end point of the non-speech data; and

cropping the first multimedia resource based on the first playing position to obtain a second multimedia resource comprises:

cropping the first multimedia resource based on the first playing position and the second playing position to obtain the second multimedia resource, wherein the second multimedia resource does not comprise the non-speech data.

10. The method of claim 1, wherein determining a first playing position of the speech data of the invalid text content in the first multimedia resource comprises:

determining the first playing position of the speech data of the invalid text content in the first multimedia resource according to a timestamp corresponding to each of the speech words in the invalid text content.

11. The method of claim 1, wherein the first multimedia resource further comprises video data, and cropping the first multimedia resource based on the first playing position to obtain a second multimedia resource comprises:

cropping the audio data and the video data of the first multimedia resource based on the first playing position to obtain the second multimedia resource.

12. An electronic device, comprising: at least one processor and memory;

the memory storing computer-executable instructions; and

the at least one processor executing the computer-executable instructions stored in the memory to cause the at least one processor to implement a multimedia processing method comprising:

obtaining a first multimedia resource;

determining an initial text content corresponding to the first multimedia resource by performing speech recognition on audio data of the first multimedia resource, wherein the audio data of the first multimedia resource comprises speech data of the initial text content;

determining an invalid text content in the initial text content, wherein the invalid text content is semantically non-informative;

determining a first playing position of the speech data of the invalid text content in the first multimedia resource; and

cropping the first multimedia resource based on the first playing position to obtain a second multimedia resource, wherein audio data of the second multimedia resource comprises speech data of a target text content but does not comprise speech data of the

invalid text content, wherein the target text content is other text content in the initial text content than the invalid text content,

wherein determining an initial text content corresponding to the first multimedia resource by performing speech recognition on audio data of the first multimedia resource comprises:

performing, according to an automatic speech recognition ASR technique, a speech recognition on the audio data of the first multimedia resource to obtain a plurality of speech words and a timestamp corresponding to each of the speech words, the timestamp denoting a playing position of audio data corresponding to the speech word in the first multimedia resource; and

generating the initial text content according to the speech words; and

wherein determining the invalid text content in the initial text content comprises:

obtaining a duration of each of the speech words according to a timestamp corresponding to the speech word; and

determining, according to a preset standard duration and a duration of the speech word, a speech word with a duration greater than the standard duration by a first duration threshold or a speech word with a duration less than the standard duration by a second duration threshold as the invalid text content in the initial text content.

13. The electronic device of claim 12, wherein determining an invalid text content in the initial text content comprises:

performing a semantic analysis on the initial text content to obtain semantic information corresponding to the initial text content; and

determining the invalid text content in the initial text content according to the semantic information.

14. The electronic device of claim 13, wherein the initial text content comprises a plurality of segment content, and determining the invalid text content in the initial text content according to the semantic information comprises:

determining a credibility coefficient of at least one segment content of the initial text content according to the semantic information corresponding to the initial text content, wherein the credibility coefficient is used for denoting credibility that the segment content is the invalid text content;

determining at least one invalid segment content from the at least one segment content according to the credibility coefficient of the segment content and a preset credibility threshold; and

determining the invalid text content in the initial text content according to the at least one invalid segment content.

15. The electronic device of claim 14, wherein determining a first playing position of the speech data of the invalid text content in the first multimedia resource comprises:

determining a start point and an end point of the speech data of each invalid segment content in the audio data of the first multimedia resource; and

determining the first playing position of the speech data of the invalid text content in the first multimedia resource based on the start point and the end point corresponding to each invalid segment content.

16. The electronic device of claim 14, wherein cropping the first multimedia resource based on the first playing position to obtain a second multimedia resource comprises:

obtaining other text content in the initial text content than the invalid segment content based on the first playing position, wherein the other text content comprises at least one target segment content;

adding a fade-in effect at a start point of speech data corresponding to the at least one target segment content, and/or adding a fade-out effect at an end point of the speech data corresponding to the at least one target segment content, to generate transition speech data corresponding to the target segment content; and

spliceing the transition speech data according to the first playing position to generate the second multimedia resource.

17. The electronic device of claim 13, wherein before performing the semantic analysis on the initial text content to obtain the semantic information corresponding to the initial text content, the method further comprises:

determining the invalid text content in the initial text content based on a preset invalid text content library;

performing a semantic analysis on the initial text content to obtain semantic information corresponding to the initial text content comprises:

performing a semantic analysis on the initial text content to obtain the semantic information corresponding to the initial text content if the invalid text content does not exist in the invalid text content library;

after determining the invalid text content in the initial text content based on the semantic information, the method further comprises:

adding the invalid text content determined based on the semantic information to the invalid text content library.

18. The electronic device of claim 12, wherein before cropping the first multimedia resource based on the first playing position to obtain the second multimedia resource, the method further comprises:

displaying the invalid text content in the initial text content; and

playing an audio segment corresponding to the invalid text content in response to an operation instruction for the invalid text content.

19. A non-transitory computer-readable storage medium, wherein the computer-readable storage medium stores computer-executable instructions, that, when executed a processor, implements the multimedia processing method comprising:

obtaining a first multimedia resource;

determining an initial text content corresponding to the first multimedia resource by performing speech recognition on audio data of the first multimedia resource, wherein the audio data of the first multimedia resource comprises speech data of the initial text content;

determining an invalid text content in the initial text content, wherein the invalid text content is semantically non-informative;

determining a first playing position of the speech data of the invalid text content in the first multimedia resource; and

cropping the first multimedia resource based on the first playing position to obtain a second multimedia resource, wherein audio data of the second multimedia resource comprises speech data of a target text content but does not comprise speech data of the invalid text content, wherein the target text content is other text content in the initial text content than the invalid text content,

wherein determining an initial text content corresponding to the first multimedia resource by performing speech recognition on audio data of the first multimedia resource comprises:

performing, according to an automatic speech recognition ASR technique, a speech recognition on the audio data of the first multimedia resource to obtain a plurality of speech words

and a timestamp corresponding to each of the speech words, the timestamp denoting a playing position of audio data corresponding to the speech word in the first multimedia resource; and

generating the initial text content according to the speech words; and

wherein determining the invalid text content in the initial text content comprises:

obtaining a duration of each of the speech words according to a timestamp corresponding to the speech word; and

determining, according to a preset standard duration and a duration of the speech word, a speech word with a duration greater than the standard duration by a first duration threshold or a speech word with a duration less than the standard duration by a second duration threshold as the invalid text content in the initial text content.