

**Method and Apparatus for Voice Annotation and Retrieval of Multimedia Data****FIELD OF THE INVENTION**

The present invention relates to databases, and in particular to systems for
5 conveniently creating, indexing and retrieving media content including audio, image
and video data and other time-sequence data, from a repository of media content.

BACKGROUND

With the advent of the Internet and the proliferation of digital multimedia technology,
10 vast amounts of digital media content are readily available. The digital media content
can be time-sequence data including audio and video data. Databases of such digital
media content have grown to voluminous proportions. However, tools for
conveniently and effectively storing such data for later retrieval and retrieving the
data have not kept abreast of the development in the volume of such data.

15 Attempts have been made to manage databases of video data. However, such systems
are characterised by being difficult to achieve automatic and convenient indexing and
retrieval of media information. Further, such systems typically have a low level of
retrieval accuracy. Therefore, a need clearly exists for an improved system of
20 indexing and retrieving media content.

**SUMMARY**

166166

In accordance with a first aspect of the invention, there is disclosed a method of voice
annotating digital media data. The method includes the steps of: speech annotating
25 one or more portions of the digital media data; and indexing the digital media data
and speech annotation based on the speech annotation to provide indexed media
content.

Preferably, the method also includes the step of creating a word lattice using the
30 speech annotation. It may also include the step of recording the speech annotation
separately from the digital media data. Optionally, the speech annotation is generated
using a formal language. Further, the annotating step can be dependent upon at least
one of a customised vocabulary and Backus-Naur Form grammar. Still further, the

-2-

step of creating the word lattice may be dependent upon at least one of acoustic and linguistic knowledge.

Preferably, the method includes the step of reverse indexing the word lattice to
5 provide a reverse index table. It may also include the step of content addressing the reverse index table.

In accordance with a second aspect of the invention, there is disclosed an apparatus
10 for voice annotating digital media data. The apparatus includes: a device for speech annotating one or more portions of the digital media data; and a device for indexing the digital media data and speech annotation based on the speech annotation to provide indexed media content.

In accordance with a third aspect of the invention, there is disclosed a computer
15 program product having a computer readable medium having a computer program recorded therein for voice annotating digital media data. The computer program product includes: a module for speech annotating one or more portions of the digital media data; and a module for indexing the digital media data and speech annotation based on the speech annotation to provide indexed media content.

20 In accordance with a fourth aspect of the invention, there is disclosed a method of voice retrieving digital media data annotated with speech. The method includes the steps of: providing indexed digital media data, the index of the indexed digital media data being derived from speech annotation of the digital media data; generating a
25 speech query; and retrieving one or more portions of the indexed digital media data dependent upon the speech query.

Preferably, the indexed digital media data is derived from a word lattice created from the speech annotation. Preferably, the method further includes the step of creating a
30 word lattice from the speech query. The word lattice may be created dependent upon at least one of acoustic and linguistic knowledge. The method may also include the step of searching the indexed media data dependent upon the speech query by matching the word lattice created from the speech query with word lattices of the

-3-

indexed media data. It may also include the step of confidence filtering the lattice created from the speech query to produce a short-list for the searching step.

5 Optionally, the method further includes the step of searching the indexed digital media data dependent upon a text query. Further, the speech query can generated dependent upon at least one of a customised vocabulary and Backus-Naur Form grammar.

10 In accordance with a fifth aspect of the invention, there is disclosed an apparatus for voice retrieving digital media data annotated with speech. The apparatus includes: a device for indexed digital media data, the index of the indexed digital media data being derived from a word lattice created from speech annotation of the digital media data; a device for generating a speech query; and a device for retrieving one or more portions of the indexed digital media data dependent upon the speech query.

15 In accordance with a sixth aspect of the invention, there is disclosed a computer program product having a computer readable medium having a computer program recorded therein for voice retrieving digital media data annotated with speech. The computer program product includes: a module for providing indexed digital media data, the index of the indexed digital media data being derived from a word lattice created from speech annotation of the digital media data; a module for generating a speech query; and a module for retrieving one or more portions of the indexed digital media data dependent upon the speech query.

20 In accordance with a seventh aspect of the invention, there is disclosed a system for voice annotating and retrieving digital media data. The system includes: a device for speech annotating at least one segment of the digital media data; a device for indexing the speech-annotated digital media data based on the speech annotation to provide indexed digital media data; a device for generating a speech or voice query; and a device for retrieving one or more portions of the indexed digital media data dependent upon the speech query.

25

30

path, an electronic network, or the like. The network may include a local area network (LAN), a wide area network (WAN), an Intranet, and/or the Internet.

The computer 502 includes: a central processing unit(s) (simply referred to as a processor hereinafter) 504, a memory 506 that may include random access memory (RAM) and read-only memory (ROM), input/output (IO) interfaces 508A and 508B, a video interface 510, and one or more storage devices generally represented by a block 512 in Fig. 5. The storage device(s) 512 can consist of one or more of the following: a floppy disc, a hard disc drive, a magneto-optical disc drive, CD-ROM, magnetic tape or any other of a number of non-volatile storage devices well known to those skilled in the art.

Each of the components 504 to 512 is typically connected to one or more of the other devices via a bus 514 that in turn can consist of data, address, and control buses. Numerous other devices can be employed as part of the computer system 500 including video capture cards, scanners, sound cards, for example. Such devices can be used to obtain video, audio and image data for use by the system of Fig. 1. The video interface 510 is connected to the video display 516 and provides video signals from the computer 502 for display on the video display 516. User input to operate the computer 502 can be provided by one or more input devices via the interface 508B. For example, an operator can use the keyboard 518 and/or a pointing device such as the mouse 520 to provide input to the computer 502.

The system 500 is simply provided for illustrative purposes and other configurations can be employed without departing from the scope and spirit of the invention. Computers with which the embodiment can be practised include IBM-PC/ATs (RTM) or compatibles, one of the Macintosh (RTM) family of PCs, Sun Sparcstation (RTM), a workstation or the like. Many such computers use graphical operating systems such as Microsoft Windows 95 and 98 (RTM), for example. The foregoing is merely exemplary of the types of computers with which the embodiments of the invention may be practised. Typically, the processes of the embodiments are resident as software or a program recorded on a hard disk drive (generally depicted as block 512 in Fig. 5) as

Claims:

1. A method of voice annotating digital media data, said method including the steps of:
5 speech annotating one or more portions of said digital media data; and
indexing said digital media data and speech annotation based on the speech annotation to provide indexed media content.
2. The method according to claim 1, further including the step of creating
10 a word lattice using said speech annotation.
3. The method according to claim 1 or 2, further including the step of recording said speech annotation separately from said digital media data.
- 15 4. The method according to any one of claims 1 to 3, wherein said speech annotation is generated using a formal language.
5. The method according to any one of claims 2 to 4, wherein said step of creating said word lattice is dependent upon at least one of acoustic and linguistic
20 knowledge.
6. The method according to any one of claims 2 to 5, further including the step of reverse indexing said word lattice to provide a reverse index table.
- 25 7. The method according to claim 6, further including the step of content addressing said reverse index table.
8. The method according to any one of claims 1 to 7, wherein said annotating step is dependent upon at least one of a customised vocabulary and
30 Backus-Naur Form grammar.
9. An apparatus for voice annotating digital media data, said apparatus including:

-18-

means for speech annotating one or more portions of said digital media data;
and

means for indexing said digital media data and speech annotation based on the
speech annotation to provide indexed media content.

5

10. The apparatus according to claim 9, further including means for
creating a word lattice using said speech annotation.

11. The apparatus according to claim 9 or 10, further including means for
10 recording said speech annotation separately from said digital media data.

12. The apparatus according to any one of claims 9 to 11, wherein said
speech annotation is generated using a formal language.

15 13. The apparatus according to any one of claims 10 to 12, wherein said
means for creating said word lattice is dependent upon at least one of acoustic and
linguistic knowledge.

14. The apparatus according to any one of claims 10 to 13, further
20 including means for reverse indexing said word lattice to provide a reverse index
table.

15. The apparatus according to claim 14, further including means for
content addressing said reverse index table.

25

16. The apparatus according to any one of claims 9 to 15, wherein said
annotating means is dependent upon at least one of a customised vocabulary and
Backus-Naur Form grammar.

30 17. A computer program product having a computer readable medium
having a computer program recorded therein for voice annotating digital media data,
said computer program product including:

means for speech annotating one or more portions of said digital media data;
and

means for indexing said digital media data and speech annotation based on the
speech annotation to provide indexed media content.

5

18. The computer program product according to claim 17, further
including means for creating a word lattice using said speech annotation.

19. The computer program product according to claim 17 or 18, further
10 including means for recording said speech annotation separately from said digital
media data.

20. The computer program product according to any one of claims 17 to
19, wherein said speech annotation is generated using a formal language.

15

21. The computer program product according to any one of claims 17 to
20, wherein said means for creating said word lattice is dependent upon at least one of
acoustic and linguistic knowledge.

22. The computer program product according to any one of claims 18 to
20, further including means for reverse indexing said word lattice to provide a reverse
index table.

23. The computer program product according to claim 22, further
25 including means for content addressing said reverse index table.

24. The computer program product according to any one of claims 17 to
23, wherein said annotating means is dependent upon at least one of a customised
vocabulary and Backus-Naur Form grammar.

30

25. A method of voice retrieving digital media data annotated with speech,
said method including the steps of:

-20-

providing indexed digital media data, the index of said indexed digital media data derived from speech annotation of said digital media data;

generating a speech query; and

retrieving one or more portions of said indexed digital media data dependent
5 upon said speech query.

26. The method according to claim 25, wherein said indexed digital media data is derived from a word lattice created from said speech annotation.

10 27. The method according to claim 25 or 26, further including the step of creating a word lattice from said speech query.

28. The method according to claim 27, further including the step of searching said indexed media data dependent upon said speech query by matching
15 said word lattice created from said speech query with word lattices of said indexed media data.

29. The method according to claim 28, further including the step of confidence filtering said lattice created from said speech query to produce a short-list
20 for said searching step.

30. The method according to claim 27, wherein said word lattice is created dependent upon at least one of acoustic and linguistic knowledge.

25 31. The method according to any one of claims 25 to 30, further including the step of searching said indexed digital media data dependent upon a text query.

32. The method according to any one of claims 25 to 31, wherein said speech query is generated dependent upon at least one of a customised vocabulary and
30 Backus-Naur Form grammar.

33. An apparatus for voice retrieving digital media data annotated with speech, said apparatus including:

-21-

means for providing indexed digital media data, the index of said indexed digital media data derived from speech annotation of said digital media data;

means for generating a speech query; and

means for retrieving one or more portions of said indexed digital media data
5 dependent upon said speech query.

34. The apparatus according to claim 33, wherein said indexed digital media data is derived from a word lattice created from said speech annotation.

10 35. The apparatus according to claim 33 or 34, further including means for creating a word lattice from said speech query.

36. The apparatus according to claim 35, further including means for searching said indexed media data dependent upon said speech query by matching
15 said word lattice created from said speech query with word lattices of said indexed media data.

37. The apparatus according to claim 36, further including means for confidence filtering said lattice created from said speech query to produce a short-list
20 for said searching means.

38. The apparatus according to claim 35, wherein said word lattice is created dependent upon at least one of acoustic and linguistic knowledge.

25 39. The apparatus according to any one of claims 33 to 38, further including means for searching said indexed digital media data dependent upon a text query.

30 40. The apparatus according to any one of claims 33 to 39, wherein said speech query is generated dependent upon at least one of a customised vocabulary and Backus-Naur Form grammar.

-22-

41. A computer program product having a computer readable medium having a computer program recorded therein for voice retrieving digital media data annotated with speech, said computer program product including:

- means for providing indexed digital media data, the index of said indexed
5 digital media data derived from speech annotation of said digital media data;
means for generating a speech query; and
means for retrieving one or more portions of said indexed digital media data
dependent upon said speech query.

10 42. The computer program product according to claim 41, wherein said indexed digital media data is derived from a word lattice created from said speech annotation.

43. The computer program product according to claim 41 or 42, further
15 including means for creating a word lattice from said speech query.

44. The computer program product according to claim 43, further
including means for searching said indexed media data dependent upon said speech
query by matching said word lattice created from said speech query with word lattices
20 of said indexed media data.

45. The computer program product according to claim 44, further
including means for confidence filtering said lattice created from said speech query to
produce a short-list for said searching means.

25

46. The computer program product according to claim 43, wherein said
word lattice is created dependent upon at least one of acoustic and linguistic
knowledge.

30 47. The computer program product according to any one of claims 41 to
46, further including means for searching said indexed digital media data dependent
upon a text query.

-23-

48. The computer program product according to any one of claims 41 to 47, wherein said speech query is generated dependent upon at least one of a customised vocabulary and Backus-Naur Form grammar.

5 49. A system for voice annotating and retrieving digital media data, said system including:

 means for speech annotating at least one segment of said digital media data;

 means for indexing said speech-annotated digital media data based on the speech annotation to provide indexed digital media data;

10 means for generating a speech or voice query; and

 means for retrieving one or more portions of said indexed digital media data dependent upon said speech query.

 50. The system according to claim 49, further including means for creating
15 a lattice structure from speech annotation.

 51. The system according to any one of claims 50, wherein said means for creating said lattice structure is dependent upon at least one of acoustic and linguistic knowledge.
20

 52. The system according to any one of claims 49 to 51, wherein said speech-annotating means post-annotates said digital media data.

 53. The system according to any one of claims 49 to 52, wherein speech
25 annotation is generated using a formal language.

 54. The system according to any one of claims 49 to 53, further including means for reverse indexing said lattice structure to provide a reverse index table.

30 55. The system according to claim 54, further including means for content addressing said reverse index table.

56. The system according to any one of claims 49 to 55, further including means for creating a lattice structure from said speech query.

57. The system according to claim 56, further including means for
5 searching said indexed digital media data dependent upon said speech query by
matching said lattice structure created from said speech query with lattice structures
of said indexed digital media data.

58. The system according to claim 57, further including means for
10 confidence filtering said lattice structure created from said speech query to produce a
short-list for said searching means.

59. The system according to claim 56, wherein said lattice structure is
created dependent upon at least one of acoustic and linguistic knowledge.
15

60. The system according to any one of claims 56 to 59, further including
means for searching said indexed digital media data dependent upon a text query.

61. The system according to any one of claims 49 to 60, wherein at least
20 one of said annotating means and said speech query is dependent upon at least one of
a customised vocabulary and Backus-Naur Form grammar.