



(19) **United States**

(12) **Patent Application Publication**  
**LEE et al.**

(10) **Pub. No.: US 2015/0326750 A1**  
(43) **Pub. Date: Nov. 12, 2015**

(54) **DATA HIDING METHOD VIA REVISION RECORDS ON A COLLABORATION PLATFORM**

(52) **U.S. Cl.**  
CPC ..... *H04N 1/32149* (2013.01); *G06F 17/24* (2013.01); *H04N 1/32352* (2013.01)

(71) Applicant: **National Chiao Tung University,**  
Hsinchu City (TW)

(57) **ABSTRACT**

(72) Inventors: **Ya-Lin LEE,** Changhua City (TW);  
**Wen-Hsiang TSAI,** Hsinchu City (TW)

(21) Appl. No.: **14/522,033**

(22) Filed: **Oct. 23, 2014**

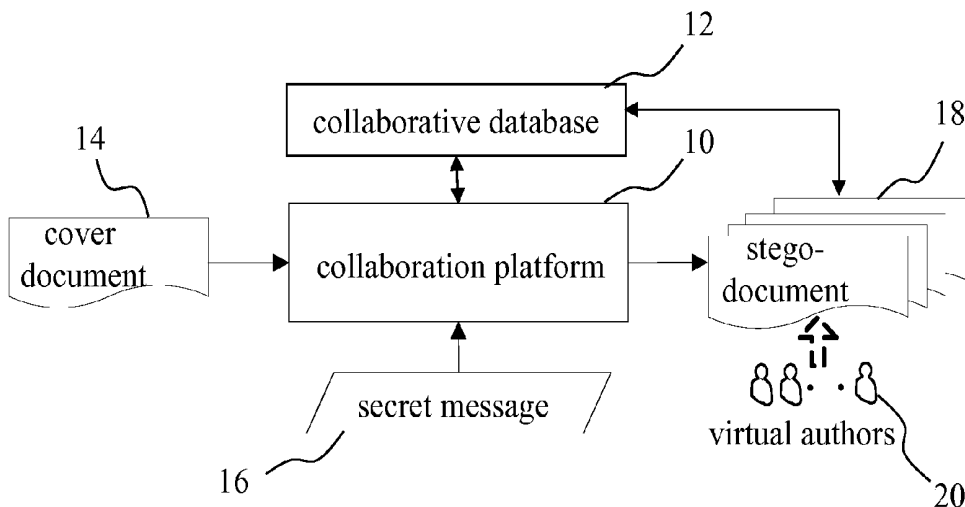
(30) **Foreign Application Priority Data**

May 9, 2014 (TW) ..... 103116542

**Publication Classification**

(51) **Int. Cl.**  
*H04N 1/32* (2006.01)  
*G06F 17/24* (2006.01)

The present invention provides a data hiding method via revision records on a collaboration platform, which first creates a collaborative database including a plurality of articles and revision records. A user puts as input a cover document, a secret message, and a key on a collaboration platform. Based on four characteristics of multi-user collaborative-writing processing, the collaborative-writing platform is used, together with a key, to hide a secret message into the cover document automatically while simulating a collaborative-writing process and generate a stego-document where the secret message is hidden. Only authorized users with the key can extract the right secret message from the stego-document i.e. the message-hidden document successfully.



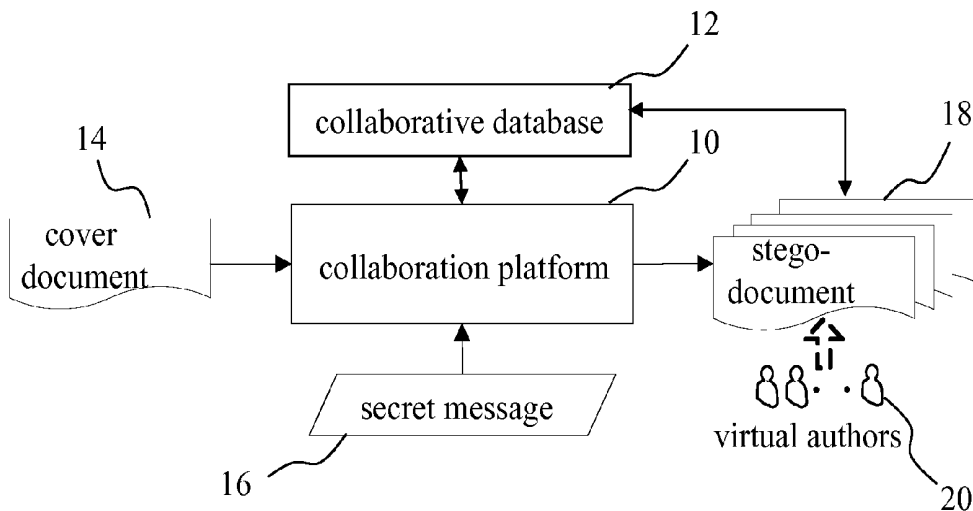


Fig.1

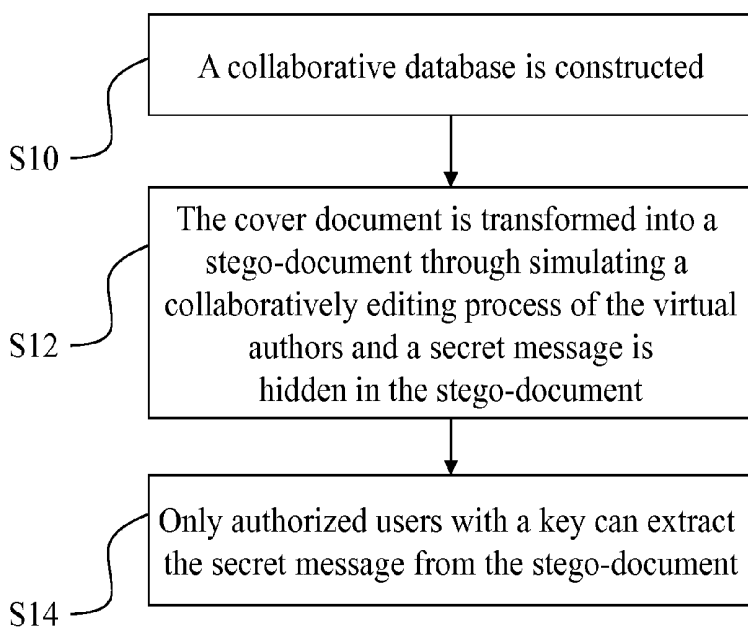


Fig.2

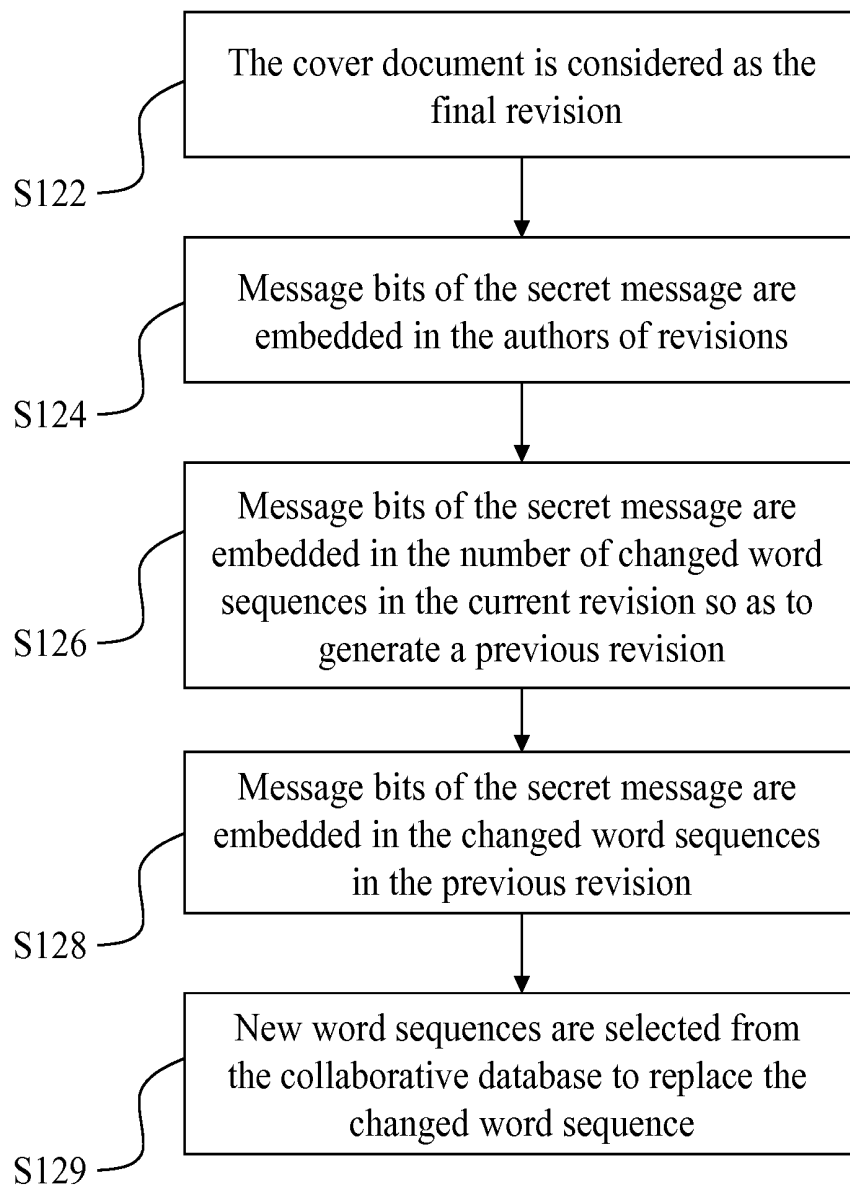


Fig.3

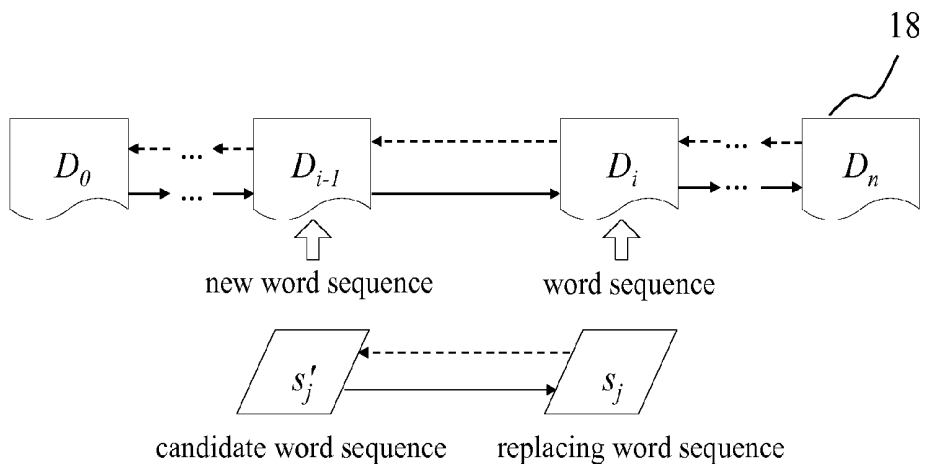


Fig.4

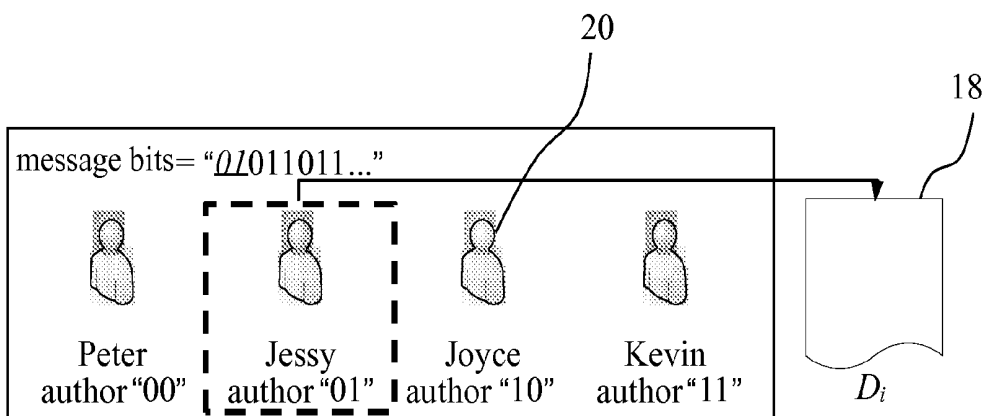
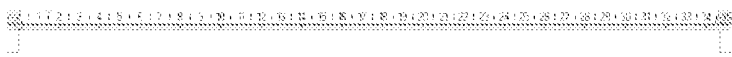


Fig.5



A new tool, called signal-rich-art image, for automatic identification and data capture applications is proposed, which is created from an artistic target image for use as a carrier of a given message. The created image is visually similar to the target image, achieving the effect of so-called signal rich art. With its function similar to those of barcodes or QR codes, such a type of image is produced by fragmenting the composing characters of the message and injecting them into the target image by a novel image-block luminance modulation scheme. Skillful techniques are also proposed for message extraction from a mobile phone-captured version of the signal-rich-art image printed on paper or displayed on a screen. Good experimental results show the feasibility of the proposed method.

Fig.6A

- ◆ (cur | prev) ● 15:40, 13 May 2013 Natalie (Talk)
- ◆ (cur | prev) ● 12:49, 13 May 2013 David (Talk)
- ◆ (cur | prev) ○ 10:08, 13 May 2013 Abigail (Talk)
- ◆ (cur | prev) ○ 08:24, 13 May 2013 Mason (Talk)
- ◆ (cur | prev) ○ 05:48, 13 May 2013 Samantha

Fig.6B

A new tool, called signal-rich-art image, for automatic identification and data capture applications is proposed, which is created from an artistic target image for use as a carrier of a given message. The created image is visually similar to the target image, achieving the effect of so-called signal rich art. With its function similar to those of barcodes or QR codes, such a type of image is produced by fragmenting the composing characters of the message and injecting them into the target image by a novel image-block luminance modulation scheme. Skillful techniques are also proposed for message extraction from a mobile phone-captured version of the signal-rich-art image printed on paper or displayed on a screen. Good experimental results show the feasibility of the proposed method. Signal rich art, as defined by Davis [1], is the art that communicates its identity to context-aware devices. A type of image of this nature, called signal-rich-art image, is proposed in this paper. The image may be printed as a hardcopy for use of any purpose, which is then "re-imaged" by a mobile-phone camera and "understood" by some automatic

Fig.6C

A new tool, called signal-rich-art image, for automatic identification and information capture applications is proposed, which is created from an artistic target image for use as a carrier of a given message. The created image is visually similar to the target image, achieving the effect of so-called signal goods art. With its function similar to those of barcodes or QR codes, such a type of image is produced by fragmenting the composing characters of the message and injecting them into the target image by a novel image-block luminance modulation scheme. Skillful techniques are also proposed for message extraction from a mobile phone-captured version of the signal-rich-art image printed on paper or displayed on a screen. Good experimental results show the feasibility of the proposesd method.

Fig.6D

A new tool, called signal-rich-art image, for automatic identification and data capture applications is proposed, which is created from an artistic target image for use as a carrier of a given message. The created image is visually similar to the target image, achieving the effect of so-called signal rich art. With its function similar to those of barcodes or QR codes, such a type of image is produced by fragmenting the composing characters of the message and injecting them into the target image by a novel image-block luminance modulation scheme. Skillful techniques are also proposed for message extraction from a mobile phone-captured version of the signal-rich-art image printed on paper or displayed on a screen. Good experimental results show the feasibility of the proposed method.

Fig.6E

Art is long, life is short

Fig.6F

境大製 嶺迥 豐嶽 嶺豐揮鐵 購統 艱陵 混 錫龍 畎 迥亞佳姪

Fig.6G

## DATA HIDING METHOD VIA REVISION RECORDS ON A COLLABORATION PLATFORM

[0001] This application claims priority for Taiwan patent application no. 103116542 filed at May 9, 2014, the content of which is incorporated by reference in its entirety.

### BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] The present invention relates to a data hiding method, and more particularly to a data hiding method via revision records on a collaboration platform.

[0004] 2. Description of the Prior Art

[0005] As the cloud systems developed, a variety of collaboration platforms are provided which allow more than one author to collaborate in editing one document, and revision records of the editing process can be stored. Since all of the files and revision records of the document will be uploaded to the Clouds, to protect these files from being attacked and to ensure their safety become a main concern. As a result, professionals in the field are pursuing to search on a new data hiding method to be developed, especially for collaboration platforms used.

[0006] In general, a data hiding method is to embed a secret message into a cover media so as to provide a resulting stego-document as a normal output that attackers or hackers cannot realize. Therefore, the data hiding methodology is the art being able applied to various fields comprising covert communications, secret data keeping, access control, database protection, and so on. Conventional types of cover media usually include image, video and audio, etc., because they are more difficult for human-eyes to realize. On the contrary, data hiding techniques using text-type cover media are much less proposed.

[0007] For example, only three major data hiding techniques using text-type cover media are commonly used in the prior art, which are (1) format-based method, (2) random and statistical method, and (3) linguistic method. Format-based methods use the physical formats of documents to hide messages, for example, the inter-word spaces without affecting the contents. Random and statistical methods generate directly camouflage texts with hidden messages to prevent the attack of comparison with a known plaintext. Alternatively, duplication patterns such as inputting more spaces, using abbreviation instead, or changing priority of parameters in the program may also be applied to conceal the secret message.

[0008] Linguistic methods use written natural languages to conceal secret messages. For instance, a synonym replacement method that generates a cover text according to a secret message using sentence models and synonym dictionary was proposed. Another synonym replacement method that hides data in a text by substituting the words which have different terms in the UK and the US was also proposed as one of the conventional linguistic methods. Alternatively, modifying an original document to a stego-document based on its data-hiding function and revision database, and then tracking the changes of the document so as to get back the original document was also known as another methodology of the conventional linguistic methods used in the prior arts.

[0009] Generally speaking, compared to (1) format-based method and (2) random and statistical method, the linguistic methods are believed to show more resistance when being attacked. Recently, more and more collaborative writing plat-

forms, such as Google Drive, Office Web Apps, Wikipedia, and so on are available. On these platforms, a plurality of authors to collaborate in editing one document is allowed, and a large number of revisions generated during the collaborative writing process are recorded. Furthermore, many people working collaboratively on these platforms make it quite necessary for data hiding applications, such as covert communication or secret data keeping, etc. However, the aforementioned methods can only be applied to documents with single author and single revision version, meaning that these conventional methods are not perfect for hiding data on collaborative writing platforms nowadays.

[0010] Therefore, on account of above, it should be obvious that there is indeed an urgent need for people having ordinary skills in the art to develop a new data hiding method that can effectively solve those above mentioned problems occurring in the prior design and ensure their safety while collaboration writing process.

### SUMMARY OF THE INVENTION

[0011] In order to overcome the above-mentioned disadvantages, one major objective in accordance with the present invention is provided for a data hiding method via revision records on a collaboration platform. The proposed method is aimed to generate a plurality of revisions of an article or document through simulating the article or document with a multi-user collaborative writing process. Then, for every two consecutive revisions, all correction pairs are found are recorded into a collaborative database. As such, the collaborative database is well constructed.

[0012] For achieving the above mentioned objectives, the proposed data hiding method via revision records on the collaboration platform utilizes four characteristics of revisions, which comprises: (1) the author of every revision, (2) the number of changed word sequences in every revision, (3) the at least one changed word sequence in every revision, and (4) the new word sequences selected from the collaborative database to replace the changed word sequence, i.e. the replacing word sequences so as to "hide" the secret message into the revisions sequentially.

[0013] Moreover, when embedding the secret message into the revisions, a key is involved. By employing such key, only authorized authors with the right key can extract the correct secret message from the revision where it is embedded.

[0014] Therefore, the data hiding method via revision records on the collaboration platform of the present invention comprises the following steps: (1) constructing a collaborative database which comprises a plurality of articles and revision records; (2) inputting a cover document, a secret message and a key on the collaboration platform; (3) automatically and artificially transforming the cover document into a stego-document, where the secret message is embedded; and (4) extracting the secret message from the stego-document by at least one authorized user with the key.

[0015] These and other objectives of the present invention will become obvious to those of ordinary skill in the art after reading the following detailed description of preferred embodiments.

[0016] It is to be understood that both the foregoing general description and the following detailed description are exemplary, and are intended to provide further explanation of the invention as claimed.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0017] The accompanying drawings are included to provide a further understanding of the invention, and are incorporated in and constitute a part of this specification. The drawings illustrate embodiments of the invention and, together with the description, serve to explain the principles of the invention in the drawings:

[0018] FIG. 1 shows a basic idea of proposed method that generates a revision history of a stego-document as a camouflage for data hiding in accordance with one embodiment of the present invention.

[0019] FIG. 2 shows a flow chart of the data hiding method proposed in accordance with one embodiment of the present invention.

[0020] FIG. 3 shows a detailed flow chart of the step S12 in FIG. 2.

[0021] FIG. 4 shows an illustrative diagram of construction order of collaborative writing database and revision generation order.

[0022] FIG. 5 shows an illustration of encoding authors of revisions for data hiding in accordance with one embodiment of the present invention.

[0023] FIGS. 6A-6G show an example of generated stego-document with input secret message "Art is long, life is short" according to one embodiment of the present invention.

## DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0024] Reference will now be made in detail to the preferred embodiments of the present invention, examples of which are illustrated in the accompanying drawings. Whenever possible, the same reference numbers are used in the drawings and the description to refer to the same or like parts. The embodiments described below are illustrated to demonstrate the technical contents and characteristics of the present invention and to enable the persons skilled in the art to understand, make, and use the present invention. However, it shall be noticed that, it is not intended to limit the scope of the present invention. Therefore, any equivalent modification or variation according to the spirit of the present invention is to be also included within the scope of the present invention.

[0025] The present invention discloses a data hiding method via revision records on a collaboration platform. The basic idea of proposed method is shown as FIG. 1. As input of a plurality of articles or documents, a user selects one of them as a cover document 14 to hide a secret message 16 into it. A collaboration platform 10 is used to simulate a multi-user collaborative-writing process, which utilizes multiple virtual authors 20 to collaboratively revise the cover document 14 into various different versions and conceal the secret message 16 into the collaborative-writing process. Therefore, a stego-document 18 which includes revision records and seems like being collaboratively edited by the plurality of virtual authors 20 is generated. The revision records and articles are stored in a collaborative database 12.

[0026] FIG. 2 shows a flow chart of the data hiding method proposed according to one embodiment of the present invention. As shown in the step of S10, a collaborative database is constructed, which comprises articles and revision records. For articles, they can be collected from Wikipedia since there were about 4.2 million articles in the English Wikipedia, which is a very large knowledge repository and suitable as a source for constructing the database. Revision records com-

prise word sequence corrections which occur between every 2 consecutive revision version of the article. FIG. 4 shows an illustration of used terms and notations according to one embodiment of the present invention.

[0027] As illustrated in FIG. 4, an article downloaded from Wikipedia has a set of revisions  $\{D_0, D_1, \dots, D_n\}$  in its revision history, where a newer revision  $D_i$  has a smaller index  $i$  with  $D_0$  being the latest version of the article. In this FIG. 4, the solid lines represents revision generation order, and the dash line represents construction order of collaborative writing database. For every two consecutive revisions  $D_i$  and  $D_{i-1}$ , all the correction pairs between  $D_i$  and  $D_{i-1}$  are found, each denoted as  $\langle s_j, s_j' \rangle$ , where  $s_j$  is a word sequence in revision  $D_i$  and was corrected to become another, namely,  $s_{j-1}$ , by the author of revision  $D_{i-1}$ . Then, all correction pairs will be found and recorded so as to construct the collaborative database. For example, assume  $D_i$ ="National Chia Tang University" and  $D_{i-1}$ ="National Chiao Tung University." Then, the correction pair  $\langle s_1, s_1' \rangle$ ="Chia Tang", "Chiao Tung" is generated and included into the collaborative database. Furthermore, according to another embodiment of the present invention, a novel algorithm can also be used for finding automatically all of the correction pairs between every two consecutive revisions for inclusion in the collaborative database. The algorithm is an extension of the longest common subsequence (LCS) algorithm.

[0028] Next, as shown in the step of S12, a secret message is embedded. The user inputs a cover document, the secret message to be embedded and a key on the collaboration platform, and the collaboration platform automatically and artificially makes the cover document become a stego-document which comprises the collaboratively editing process of the virtual authors and the secret message hidden in the document.

[0029] For the details of step S12, please refer to FIG. 3. As the step of S122, in the phase of message embedding with a cover document as the input, the proposed method is designed to provide the cover document as the final revision  $D_n$ , and provide consecutive revisions  $\{D_{n-1}, D_{n-2}, \dots, D_1, D_0\}$  by producing a previous revision a from the current revision  $D_{n-1}$  repeatedly until the entire message is embedded as shown in FIG. 4, where the direction of revision generation order is indicated by the solid lines and the direction of construction order of collaborative writing database is indicated by dash lines. The stego-document  $D_n$  including the revision history  $\{D_{n-1}, D_{n-2}, \dots, D_1, D_0\}$  then is kept on the collaborative writing platform, which may be Wikipedia or others. To simulate a collaborative writing process more realistically, the present invention utilizes four characteristics of revisions to "hide" the message bits into the revisions sequentially: (1) the author of every revision, (2) the number of changed word sequences in every revision, (3) the at least one changed word sequence in every revision, and (4) the new word sequences selected from the collaborative database to replace the changed word sequence, i.e. the replacing word sequences, as shown in steps of S124~S129, respectively. As shown in S124, the authors of revisions are encoded to hide message bits in the proposed method. For this, at first a group of simulated authors are selected, with each author being assigned a unique code  $a$ , called author  $a$ . Then, if the message bits to be embedded form a code  $a_j$ , then the author  $a_j$  will be assigned to the revision  $D_j$  as its author to achieve embedding of message bits  $a_j$  into  $D_j$ . For example, assume that four authors are selected and each is assigned a unique code  $a$ , as



shown in FIG. 5, respectively. If the message bits  $a_j$  to be embedded is "01," then Jessy with author code "01" is selected to be the author of the revision  $D_j$ . Moreover, every revision of  $D_0$  through  $D_n$  will be assigned an author according to the corresponding message bits, and so an author can be assigned to conduct more than one revision or reversely no revision in the generated revisions, which in turns fits the real situation of multi-user collaborating process.

**[0030]** Next, the step of S126 uses the number of changed word sequences for data hiding and generates the previous revision  $D_i$  from the current one  $D_{i-1}$ . In this process, some word sequences in  $D_{i-1}$  are selected and changed into other ones in  $D_i$ . It is desired to use as well the number of word sequences changed in this process  $N_g$  as a message-bit carrier. To implement this aim, at first the present invention sets on the magnitude of  $N_g$  a limit  $N_c$  taken to be the maximum allowed number of word sequences in  $D_{i-1}$  that can be changed to yield  $D_i$ . This limitation makes the simulated step of revising  $D_{i-1}$  to become  $D_i$  look more realistic because usually not very many words are corrected in a single revision. Next, the proposed method scans the word sequences in the text of the current revision  $D_{i-1}$  sequentially and search the database to find all the correction pairs  $\langle s_j, s_j' \rangle$  with  $s_j'$  in  $D_{i-1}$ . Then, collect all  $s_j'$  in these pairs as a set  $Q_r$ , which is called as the candidate set of word sequences for changes in  $D_{i-1}$ . Finally,  $N_g$  word sequences will be selected out of  $Q_r$  to form a set such that the binary version of the number  $N_g$  is just the current message bits to be embedded. In one embodiment, if the number of candidate word sequences for changes is 3 and the binary version of the number 3 is 11, then the secret message bits to be embedded will be "11".

**[0031]** In the step of S128, the secret message bits will be embedded in the changed word sequence in the previous revision  $D_j$ , and the candidate set of word sequences for changes in  $Q_r$  will be divided into  $N_g$  groups. In each group, at least one changed word sequence  $s_j'$  will be selected as for secret message to be embedded in.

**[0032]** As for step S129, certain new word sequences, i.e. the replacing word sequences are selected from the collaborative database to replace the changed word sequence  $s_j'$  in S128. A number  $N_g$  of changed word sequence  $s_j'$  are selected from the previous revision  $D_i$  which are the new word sequence in S126. Since the new word sequences are re-selected in the step of S128 to form a set, the candidate set of word sequences for changes will accordingly be the same as the new word sequences. Among the number  $N_g$  of changed word sequence  $s_j'$  being selected and the revision times each  $s_j'$  replacing  $s_j$ , a Huffman coding technique based on the collaborative writing database is adopted to provide specific codes for every new word sequence which will be selected. As such, every new word sequence will be characterized with a relative code, and the replacing  $s_j$  can be decided based on the secret message. After using the changed word sequence  $s_j'$  to replace  $s_j$ , the current version of revision  $D_{i-1}$  is successfully formed.

**[0033]** At last, as shown in the step of S14 in FIG. 2, only authorized users with the right key can extract the correct secret message from the stego-document, since only they have the access to obtain the information of correction pairs, relative codes for each new word sequence, and so on.

**[0034]** FIGS. 6A-6G show an example of generated stego-document according to one embodiment of the present invention. In FIG. 6A, an article is selected as cover document where the secret message "Art is long, life is short" will be

embedded. After simulating the multi-user collaboratively writing process on the platform is performed, five different revision records are shown as FIG. 6B, which includes the revision date, time, and author name and "Natalie" is the author of the latest version of revision. FIG. 6C shows the stego-document which have exactly the same contents as the cover document shown in FIG. 6A. FIG. 6E is the latest version of revision with contents same as the cover document in FIG. 6A. FIG. 6D is the previous version of FIG. 6E, with words as indicated being corrected to be new ones in FIG. 6E. The revision records are inclusive of the secret message. As shown in FIG. 6F, a user with a right key can extract the correct secret message from the version of FIG. 6E, while compared to FIG. 6G, which shows a wrong extracted secret message with a wrong key, the wrong extracted message becomes a bunch of gibberish. Therefore, it is believed that the data hiding method proposed in the present invention is beneficial and effective to secure safety for secret messages to be embedded in any type of documents.

**[0035]** To sum up, the present invention provides a novel data hiding method via revision records on a collaboration platform. The proposed method first analyzes an existing writing platform on the internet, and obtain useful information from the at least one existing platform so as to construct a collaborative database. An article is then selected from the database as a cover document for the secret message to be embedded in. As such, a stego-document which seems exactly the same as the original cover document but in fact comprising the secret message and revision records of virtual authors is created. The revision records are together with the document to be stored in the database. To embed the secret message and simulate a collaborative writing process, the proposed method utilizes four characteristics of revisions to "hide" the message bits into the revisions sequentially. Moreover, based on the number of times the word sequence in the article is revised, a Huffman coding technique is further adopted to encode this value, i.e. the number of times of the revisions such that the whole simulating process seems more realistically. By employing the proposed method of the present invention, it can be effectively applied to documents with more than one author and revision versions, meaning that the proposed method of the present invention is not only perfect for hiding data on collaborative writing platforms but also useful for convert communications, secret data keeping, access control, database protection, and so on.

**[0036]** It will be apparent to those skilled in the art that various modifications and variations can be made to the present invention without departing from the scope or spirit of the invention. In view of the foregoing, it is intended that the present invention cover modifications and variations of this invention provided they fall within the scope of the invention and its equivalent.

What is claimed is:

1. A data hiding method via revision records on a collaboration platform, comprising steps of:

constructing a collaborative database which comprises a plurality of articles and revision records;

inputting a cover document, a secret message and a key on said collaboration platform, in which said cover document is automatically and artificially transformed into a stego-document, comprising a collaboratively editing process of virtual authors and said secret message is hidden in said stego-document, and

extracting said secret message from said stego-document by at least one authorized user with said key.

2. The data hiding method of claim 1, wherein said secret message is hidden in said stego-document, and a plurality of characteristics of said collaboratively editing process are utilized, comprising: author of every revision, a number of changed word sequence in every revision, at least one changed word sequence in every revision, and at least one new word sequence selected from said collaborative database to replace said changed word sequence.

3. The data hiding method of claim 1, further comprising using an extension of the longest common subsequence (LCS) algorithm to compare every two consecutive revisions of said articles so as to find all correction pairs and to obtain said revision records; and storing said revision records in said collaborative database.

4. The data hiding method of claim 2, in said step of creating said stego-document further comprising:

considering said cover document as a final revision of said article; and

providing consecutive revisions according to said characteristics of said collaboratively editing process by producing a previous revision from a current revision repeatedly until said entire secret message is embedded so as to create said stego-document.

5. The data hiding method of claim 4, wherein when said secret message is hidden in said stego-document according to said author of every revision, said virtual authors on said collaboration platform are selected with each being assigned a unique code, and message bits of said secret message are the same as said unique code of said at least one virtual author, said at least one virtual author will be selected as author of said current revision so that said message bits of said secret message are successfully embedded into said at least one virtual author.

6. The data hiding method of claim 4, wherein when said secret message is hidden in said stego-document according to said number of changed word sequence in every revision, a limit taken to be maximum allowed number of word sequences that can be changed is set; word sequences in text of said current revision is scanned sequentially with searching said database such that all correction pairs can be found; said new word sequence is compared to said changed word sequence in said previous revision and collected to become a set; out of said set a plurality of candidate word sequences for changes is chosen; and a binary version of said candidate word sequences for changes is calculated such that message bits of said secret message can be embedded into said binary version of said candidate word sequences for changes.

7. The data hiding method of claim 6, wherein when said secret message is hidden in said stego-document according to said changed word sequence in every revision, said candidate word sequences for changes will be divided into a plurality of groups; and at least one of said candidate word sequences for changes in each group will be selected as for said secret message to be embedded in.

8. The data hiding method of claim 7, in said step of selecting said new word sequence from said collaborative database to replace said changed word sequence further comprising: choosing a plurality of new word sequences from said previous revision and assigning specific code to every new word sequence; deciding at least one changed word sequence based on said secret message; and replacing said changed word sequence with said new word sequence to form said current revision.

9. The data hiding method of claim 8, wherein said specific code is analyzed through a number of times of revisions, and a Huffman coding technique is adopted to provide said specific code to every new word sequence based on said number of times of revisions.

\* \* \* \* \*