DESCRIPTION

[Title of Invention] RECORDING MEDIUM, DATA ACQUISITION METHOD, AND DATA ACQUISITION APPARATUS

[Technical Field]

[0001] The present invention relates to a data acquisition program, a data acquisition method, and a data acquisition apparatus.

[Background Art]

[0002] Crawler tools have been known as tools for collecting information published on the Internet. A crawler tool goes round Web pages on the Internet, and stores contents thereof in uniform resource locator (URL) units, that is, in page units. Further, it has been proposed to detect presence or absence of update of a part by: analyzing contents of Web pages and the like; identifying a location specified by a user; and when a new content is received, extracting a corresponding specified location from the new content and comparing the corresponding specified location with the original data.

[0003] Further, it has been proposed to generate a page part including a content of a specified portion that is able to be independent, which is a portion that is able to be processed independently of a Web page, that is, a Web page, when tag information of the part that is able to be independent is extracted and the portion that is able to be independent is specified by a user. Furthermore, it has been proposed to generate a new Web page, based on the generated page part.

[Citation List]

[Non Patent Citation]

[0004] Patent Document 1: Japanese Laid-open Patent Publication No. 2001-202283

Patent Document 2: Japanese Laid-open Patent

Publication No. 2001-109742

[Summary of Invention]

[Technical Problem]

[0005] However, in a case where specific information is collected from various Web pages on the Internet, for example, if a received content or a Web page based on a generated page part is referred to, other information that is not aimed to be collected may be included. Therefore, it is difficult to collect only the specific information and output highly versatile data.

[0006] According to one aspect of the present invention, a data acquisition program, a data acquisition method, and a data acquisition apparatus, which enable data of a targeted part to be extracted and output even if specific tag information is not available, are provided.

[Solution to Problem]

[0007] In one aspect, a computer-readable recording medium stores a data acquisition program that causes a computer to execute a process of identifying a position of a part to be extracted in a document that is associated with a specific URL and includes structural information of tags, the position being on a hierarchical structure of the tags included in the document, and allowing the position on the hierarchical structure to be registered. The computer-readable recording medium stores the data acquisition program that causes the computer to execute a process of accessing periodically or non-periodically the document associated with the specific URL, and extracting and outputting data corresponding to the registered position on the hierarchical structure of the tags.

[Advantageous Effects of Invention]

[0008] Data of a targeted part are able to be extracted and output even if specific tag information is not available.

[Brief Description of Drawings]

CLAIMS

1. A computer-readable recording medium storing a data acquisition program that causes a computer to execute a process comprising:

identifying a position of a part to be extracted in a document that is associated with a specific URL and includes structural information of tags, the position being on a hierarchical structure of the tags included in the document, and allowing the position on the hierarchical structure to be registered; and

accessing periodically or non-periodically the document associated with the specific URL, and extracting and outputting data corresponding to the registered position on the hierarchical structure of the tags.

2. The computer-readable recording medium according to claim 1, wherein the position of the part to be extracted is further identified by use of a combination of: a name of a tag or an ordering sequence number of the tag in the document; and a hierarchical structure of tags.

3. The computer-readable recording medium according to claim 1, wherein when data corresponding to a position on the hierarchical structure of the tags, the data having been extracted and registered in the past, are different from data corresponding to the position on the hierarchical structure of the tags, the data having been extracted and registered this time, information indicating that the data have changed is output.

4. The computer-readable recording medium according to claim 1, wherein when a plurality of positions of the part to be extracted have been registered for the document, output according to the number or rate of data matching past data,

in data corresponding to the plurality of positions is performed.

5. The computer-readable recording medium according to claim 1, wherein

the document described in HTML format or a source of the document is displayed;

selection of a part to be extracted that is included in the displayed document or the displayed source of the document is received;

a hierarchy of a tag corresponding to the received part to be extracted is identified; and

the identified hierarchy is registered as information identifying a position of the part to be extracted.

6. A data acquisition method, in which a computer executes a process comprising:

identifying a position of a part to be extracted in a document that is associated with a specific URL and includes structural information of tags, the position being on a hierarchical structure of the tags included in the document, and allowing the position on the hierarchical structure to be registered; and

accessing periodically or non-periodically the document associated with the specific URL, and extracting and outputting data corresponding to the registered position on the hierarchical structure of the tags.

7. The data acquisition method according to claim 6, wherein the position of the part to be extracted is identified by use of a combination of: a name of a tag or an ordering sequence number of the tag in the document; and a hierarchical structure of tags.