

CLAIMS

1. 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;

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; and

when a plurality of positions of the part to be extracted have been registered for the document and having performed an access for an unknown document, performing output, for the unknown document, according to the number or rate of data matching data that have been acquired already and correspond to the plurality of positions.

2. The data acquisition program 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 data acquisition program 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 data acquisition program 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 data acquisition program 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: receiving a selection of a part to be extracted in a document that is associated with a specific URL and includes structural information of tags, and identifying a position corresponding to the received part to be extracted and being on a hierarchical structure of the tags included in the document;

registering the identified position on the hierarchical structure of the tags in a storage unit;

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

when a plurality of positions on the hierarchical structure of the tags have been registered for the document and having performed an access for an unknown document, performing output, for the unknown document, according to the number or rate of data matching data that have been acquired already and correspond to the plurality of positions.

7. A data acquisition apparatus comprising:

a registration unit configured to: identify 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 allow the position on the hierarchical structure to be registered;

a first output control unit configured to periodically or non-periodically access the document associated with the specific URL, and extract and output data corresponding to the registered position on the hierarchical structure of the tags; and

a second output control unit configured to, when a plurality of positions of the part to be extracted have been registered for the document and an access for an unknown document has been performed, perform output, for the unknown document, according to the number or rate of data matching data that have been acquired already and correspond to the plurality of positions.