US Patent & Trademark Office
Patent Public Search | Text View

United States Patent	12394238
Kind Code	B2
Date of Patent	August 19, 2025
Inventor(s)	Kim; Dong Hwan et al.

Method and apparatus for data structuring of text

Abstract

Provided are method and apparatus for data structuring of text. The apparatus for data structuring of text includes a processor; and a memory storing instructions executable by the processor, wherein the processor is configured to execute the instructions to: extract text and location information of the text from an image, set text units for the extracted text, assigning a first tag and a second tag to at least one of the text units, connect text units with related tags among the text units allocated the first tag and the second tag, label the connected text units as first text, second text, and third text respectively corresponding to an item name, an item value, and others based on a natural language processing model, and structure the extracted text by mapping the second text to the first text.

Inventors: **Kim; Dong Hwan (Seoul, KR), Kwon; You Kyung (Seoul, KR), Ko; So Young (Seoul, KR), Roe; Sook Jin (Seoul, KR), Kwon; Ki Beom (Gyeongggi-do, KR), Moon; Da Hea (Seoul, KR)**

Applicant: **42Maru Inc.** (Seoul, KR)

Family ID: **1000008763571**

Assignee: **42Maru Inc. (Seoul, KR)**

Appl. No.: **18/741370**

Filed: **June 12, 2024**

Prior Publication Data

Document Identifier	Publication Date
US 20240331432 A1	Oct. 03, 2024

Foreign Application Priority Data

KR	10-2021-0135569	Oct. 13, 2021

Related U.S. Application Data

continuation parent-doc US 17502017 20211014 US 12033413 child-doc US 18741370

Publication Classification

Int. Cl.: **G06K9/00** (20220101); **G06F16/953** (20190101); **G06F40/20** (20200101); **G06V30/12** (20220101); **G06V30/19** (20220101); **G06V30/412** (20220101); **G06V30/413** (20220101); **G06V30/414** (20220101); **G06V30/416** (20220101)

U.S. Cl.:

CPC **G06V30/413** (20220101); **G06F16/953** (20190101); **G06F40/20** (20200101); **G06V30/12** (20220101); **G06V30/19093** (20220101); **G06V30/412** (20220101); **G06V30/414** (20220101); **G06V30/416** (20220101);

Field of Classification Search

CPC: G06V (30/10); G06V (30/414); G06V (30/40); G06V (30/416); G06V (10/242); G06V (20/69); G06V (30/153); G06V (30/268); G06V (30/413); G06V (10/24); G06V (10/98); G06V (30/12); G06V (30/19107); G06V (30/1914); G06V (30/2504); G06V (30/412); G06V (10/56); G06V (10/758); G06V (10/764); G06V (10/82); G06V (10/945); G06V (20/48); G06V (20/62); G06V (2201/033); G06V (2201/09); G06V (2201/10); G06V (30/1423); G06V (30/1444); G06V (30/1452); G06V (30/162); G06V (30/19147); G06V (30/19173); G06V (30/226); G06V (30/2276); G06V (30/245); G06V (30/287); G06V (30/373); G06V (30/418); G06V (40/161); G06V (10/25); G06V (10/30); G06V (10/462); G06V (10/464); G06V (10/507); G06V (10/70); G06V (10/751); G06V (10/774); G06V (10/987); G06V (20/13); G06V (20/20); G06V (20/30); G06V (20/40); G06V (20/41); G06V (20/46); G06V (20/47); G06V (20/49); G06V (20/635); G06V (20/68); G06V (20/698); G06V (2201/03); G06V (2201/034); G06V (30/1448); G06V (30/148); G06V (30/19007); G06V (30/19067); G06V (30/19093); G06V (30/1912); G06V (30/19127); G06V (30/2264); G06V (30/2272); G06V (30/244); G06V (30/262); G06V (30/274); G06V (30/32); G06V (30/333); G06V (30/347); G06V (30/41); G06V (30/43); G06V (40/172); G06V (40/28); G06V (40/50)

References Cited

U.S. PATENT DOCUMENTS

Patent No.	Issued Date	Patentee Name	U.S. Cl.	CPC
5689585	12/1996	Bloomberg	382/229	G06V 30/262
10957433	12/2020	Lucas et al.	N/A	N/A
2005/0037806	12/2004	Na	379/355.02	H04M 1/2755
2008/0152238	12/2007	Sarkar	N/A	N/A
2011/0280481	12/2010	Radakovic et al.	N/A	N/A
2012/0072859	12/2011	Wang	715/764	G06V 30/40

2015/0310561	12/2014	Sugaya	705/30	G06Q 40/12
2020/0110930	12/2019	Simantov et al.	N/A	N/A
2021/0342544	12/2020	Galle	N/A	G06F 40/30
2022/0044298	12/2021	Oshinaike	N/A	G06V 20/52

FOREIGN PATENT DOCUMENTS

Patent No.	Application Date	Country	CPC
2021-060876	12/2020	JP	N/A
10-2019-0123790	12/2018	KR	N/A
10-2021-0033730	12/2020	KR	N/A
10-2021-0050202	12/2020	KR	N/A

OTHER PUBLICATIONS

Korean Office Action issued Aug. 24, 2023 in Application No. 10-2021-0135569. cited by applicant

Extended European Search Report issed Apr. 5, 2022 in European Application No. 21202576.1. cited by applicant

---

Primary Examiner: Liew; Alex Kok S

Attorney, Agent or Firm: Sughrue Mion, PLLC

Background/Summary

CROSS-REFERENCE TO RELATED APPLICATION (1) This application is a continuation application of U.S. application Ser. No. 17/502,017, filed on Oct. 14, 2021, which claims priority to and the benefit of Korean Patent Application No. 2021-0135569, filed on Oct. 13, 2021, the disclosure of which is incorporated herein by reference in its entirety.

BACKGROUND

1. Field of the Invention

(1) The present invention relates to a method and apparatus for data structuring of text, and more particularly, to a method and apparatus for data structuring of text that is capable of improving the efficiency of operational management of documents by structuring text included in an unstructured image document.

2. Discussion of Related Art

(2) In the industrial field, there has been a continuous demand for processing unstructured documents that are mutually transmitted between individuals, organizations, or different organizations so that the unstructured documents can be effectively managed and operated. However, there is a disadvantage that a large cost is required for a simple iterative input process of information due to different document formats, diversity of terms, and the like.

SUMMARY OF THE INVENTION

(3) The present invention is directed to efficient operational management of an image document through classification of an image document, extraction of text (item name and item value) in the image document, data structuring of text, correction of misrecognition, and the like.

(4) According to an aspect of the present invention, there is provided an apparatus for data structuring of text including a processor; and a memory storing instructions executable by the

processor, wherein the processor is configured to execute the instructions to: extract text and location information of the text from an image, set text units for the extracted text, assigning a first tag and a second tag to at least one of the text units, connect text units with related tags among the text units allocated the first tag and the second tag, label the connected text units as first text, second text, and third text respectively corresponding to an item name, an item value, and others based on a natural language processing model, and structure the extracted text by mapping the second text to the first text.

(5) According to an aspect of the present invention, there is provided a method of data structuring of text including extracting text and location information of the text from an image; setting text units for the extracted text; assigning a first tag and a second tag to at least one of the text units; connecting text units with related tags among the text units allocated the first tag and the second tag; labeling the connected text units as first text, second text, and third text respectively corresponding to an item name, an item value, and others based on a natural language processing model; and structuring the extracted text by mapping the second text to the first text.

Description

BRIEF DESCRIPTION OF THE DRAWINGS

(1) The above and other objects, features and advantages of the present invention will become more apparent to those of ordinary skill in the art by describing in detail exemplary embodiments thereof with reference to the accompanying drawings, in which:

(2) FIG. **1** is a diagram illustrating a schematic structure of an apparatus for data structuring of text according to an embodiment of the present invention;

(3) FIG. **2** is a diagram illustrating an input and output of a process of extracting text from an image according to an embodiment of the present invention;

(4) FIG. **3** is a diagram illustrating a schematic structure of an apparatus for data structuring of text according to another embodiment of the present invention; and

(5) FIG. **4** is a flowchart for describing a method of data structuring of text according to an embodiment of the present invention.

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

(6) The above-described objects, features, and advantages will be described below in detail with reference to the accompanying drawings, and accordingly, those skilled in the art to which the present invention pertains will be able to easily implement the technical idea of the present invention. When it is determined that the detailed description of the known art related to the present invention may unnecessary obscure the gist of the present invention, a detailed description thereof will be omitted.

(7) In the drawings, the same reference numerals are used to designate the same or similar elements, and all combinations described in the specification and claims may be combined in any manner. Unless otherwise specified, it should be understood that references to the singular may include more than one, and references to the singular may also include the plural.

(8) The terms used in this specification are for the purpose of describing specific exemplary embodiments only and are not intended to be limiting. Singular expressions as used herein may also be intended to include plural meanings unless clearly indicated otherwise in the corresponding sentence. The term “and/or” includes all combinations and any of the items listed in connection therewith. The terms “include,” “including,” “being included,” “comprising,” “have,”, “having,” and the like have inclusive meanings, and accordingly, these terms specify features, integers, steps, actions, elements, and/or components described herein, and do not preclude the presence or addition of one or more other features, integers, steps, actions, elements, components, and/or groups thereof. The steps, processes, and operations of the method described herein should not be

construed as necessarily being performed in any specific order as discussed or exemplified unless specifically determined to be the order of performance thereof. It should also be understood that additional or alternative steps may be used.

(9) In addition, each component may be implemented as a hardware processor, the above components may be integrated and implemented as one hardware processor, or the above components may be combined with each other and implemented as a plurality of hardware processors.

(10) Hereinafter, exemplary embodiments of the present invention will be described in detail with reference to the accompanying drawings.

(11) FIG. **1** is a diagram illustrating a schematic structure of an apparatus for data structuring of text according to an embodiment of the present invention.

(12) Referring to FIG. **1**, the apparatus for data structuring of text according to the embodiment of the present invention includes a data extraction unit **100**, a data processing unit **200**, a form classification unit **300**, a labeling unit **400**, a relationship identification unit **500**, a misrecognition correction unit **600**, and a database **700** and may extract user's desired information from a received image.

(13) The data extraction unit **100** may extract text included in an image received by a user terminal (not illustrated) and location information of the text. The data extraction unit **100** may extract the text included in the image and the location information of the text based on an optical character recognition (OCR) technique. For example, the data extraction unit **100** may apply the OCR technique to the image as illustrated in FIG. **2**A and extract an output as illustrated in FIG. **2**B.

(14) The data processing unit **200** may generate a text unit based on the text extracted from the data extraction unit **100** and the location information. The data processing unit **200** may set a text unit based on a preset length threshold value and a distance between keywords included in the text, based on the location information of the text.

(15) The form classification unit **300** may classify the form of the received image based on the text extracted by the data extraction unit **100**. The form classification unit **300** may search for the text extracted by the data extraction unit **100** in the database **700** in which a plurality of previously registered form samples are stored, identify the form most similar to the text, and classify the form of the received image.

(16) For example, the form classification unit **300** includes a first form including text A, B, C, and D and a second form including text A, C, E, and F in the form sample, and classifies the form of the received image into the first form when the text extracted by the data extraction unit **100** includes the text A, B, and C.

(17) The labeling unit **400** may label the text units as first text to third text based on the forms (e.g., items for each form or the like) classified by the form classification unit **300**. The labeling unit **400** may label text corresponding to an item name as the first text, text corresponding to an item value as the second text, and text corresponding to others as the third text among the text units.

(18) The labeling unit **400** may apply the text unit included in the set group to a text classification model, which is a BERT-based natural language processing model, to label the text unit with the item name, the item value, or the like.

(19) The text classification model operates as a pre-trained model by the labeling unit **400** based on training data labeled with the item name, the item value, or others by the user to label the text unit included in a group. The text classification model may modify and train training data for characters with a high misrecognition occurrence rate, such as O-Q to prepare for a case where the text extracted by the OCR technique is misrecognized. For example, when one text includes Orange, the text classification model may further perform keyword training on the Qrange as well as the keyword training on the Orange.

(20) The relationship identification unit **500** may structure the text by mapping the first text and the second text labeled by the labeling unit **400**. The relationship identification unit **500** may map the

second text corresponding to the first text, that is, the item value corresponding to the item name, to each other.

(21) The relationship identification unit **500** may perform mapping based on the location information of the first text and the second text. The relationship identification unit **500** may identify fourth text that is the first text that falls within a preset distance threshold value with the second text.

(22) When a plurality of fourth texts are identified, the relationship identification unit **500** may calculate the vector similarity between the second text and the fourth text through a similarity verification model and map the fourth text with the highest similarity to the second text based on the calculated vector similarity.

(23) In this case, when identifying the fourth text, the relationship identification unit **500** processes the mapping process under the assumption that in the format of the form, the first text has a larger y value or a smaller x value in the location information than the second text.

(24) Also, when a plurality of second texts having the same format are present in one group, the relationship identification unit **500** may map all the second texts to the first text that is mapped to the leftmost second text or the first text mapped to the uppermost second text.

(25) The relationship identification unit **500** may set the first text and the second text mapped to the first text as a group to map texts for the item name-item value. Accordingly, one group may include a text unit for the item name and a text unit for at least one item value.

(26) The misrecognition correction unit **600** may determine whether the first text is misrecognized and correct the first text. The misrecognition correction unit **600** may compare the first text and the representative keyword for the item name in the training data of the natural language processing model used by the labeling unit **400**. Since the keyword for the item name in the training data includes all keywords that are modified to prevent the misrecognition, it is possible to determine whether or not the misrecognition occurs by comparing the first text with an unmodified representative keyword.

(27) The misrecognition correction unit **600** may calculate the similarity between the representative keyword and the first text and may determine that, when the representative keyword and the first text are not identical to each other but the calculated similarity is greater than or equal to a preset similarity threshold, the first text is misrecognized.

(28) The misrecognition correction unit **600** may correct the first text determined to be misrecognized as a representative keyword and provide correction details to the user terminal.

(29) Since the item name serves as an indicator of the data written in the form, as the misrecognized item name acts as a major obstacle in the data structuring process, the misrecognition correction unit **600** corrects the first text corresponding to the item name to improve the quality of data structuring and reduce a workload in the inspection and calibration process.

(30) The apparatus for data structuring of text according to the embodiment of the present invention is somewhat dependent on data because it is based on existing training data in the process of labeling text. To solve the above problem, an apparatus (see FIG. **3**) for data structuring of text according to another embodiment of the present invention may include a data extraction unit **100**, a data processing unit **210**, a form classification unit **300**, a labeling unit **410**, a relationship identification unit **510**, a misrecognition correction unit **600**, and a database **700**.

(31) The apparatus for data structuring of text according to another embodiment of the present invention of FIG. **3** is different from the apparatus for data structuring of text according to the embodiment of the present invention of FIG. **1** in terms of the operations of the data processing unit, the labeling unit, and the relationship identification unit. Hereinafter, the operations of the data processing unit **210**, the labeling unit **410**, and the relationship identification unit **510** will be described.

(32) The data processing unit **210** may extract line information included in an image by using text extracted by the data extraction unit **100**, location information of the text, and an original image

file. The data processing unit **210** may generate a text unit based on the extracted line information. Also, the line information extracted at this time may be used to map text in the relationship identification unit **510**.
(33) The labeling unit **410** may classify text units present in the image into first text to third text by applying the image received from the user terminal to the deep learning model as input data.
(34) The labeling unit **410** may add a tag to each text unit present in the image through the deep learning model. The labeling unit **410** may add tags of <KEY-B>, <KEY-I>, <VAL-B>, and <VAL-I> to each of the keywords included in the text unit through the deep learning model. KEY refers to an item name, VAL refers to an item value, B refers to start, and I refers to middle.
(35) For example, when SHIPPER/EXPORTER and NO. & DATE INVOICE keyword are present in the text, the labeling unit **410** may add tags such as SHIPPER/EXPORTER <KEY-B>, NO. & DATE <KEY-B> INVOICE <KEY-I> through a deep learning model. This will be the basis for possibly determining that SHIPPER/EXPORTER and NO. & DATE INVOICE are item names.
(36) The labeling unit **410** may label the text by adding a tag to each keyword as described above by applying an NER technique through the deep learning model.
(37) The relationship identification unit **510** may structure text by mapping the first text and the second text labeled by the labeling unit **410**, and may further use line information extracted from the data processing unit **210**.
(38) The relationship identification unit **510** may map the first text and the second text based on the line information. Since the line information present in the document classifies text for each item, it is possible to improve the mapping accuracy of the text by further using the line information.
(39) The relationship identification unit **510** will map the first text and the second text based on the vector similarity of the first text and the second text through a similarity verification model like the relationship identification unit **500** when there is no line information in the image.
(40) FIG. **4** is a flowchart for describing a method of data structuring of text according to an embodiment of the present invention. Hereinafter, a method of data structuring of text will be described with reference to FIG. **4**. In the description of the method of data structuring of text, a detailed embodiment overlapping with the apparatus for data structuring of text described above may be omitted.
(41) In operation S**100**, a server may extract text included in the image received from the user terminal and location information of the text. The server may extract the text included in the image and the location information of the text based on the OCR technique.
(42) In operation S**200**, the server may generate a text unit based on the extracted text and location information. In this case, the text unit may be set based on a preset length threshold value and a distance between keywords included in the text, based on the location information of the text.
(43) In operation S**300**, the server may classify a form of an image based on the extracted text. The server may classify the form of the image by identifying a form most similar to the text by searching for the text extracted in the database in which a plurality of previously registered form samples are stored.
(44) In operation S**400**, the server may label the text unit as the first text to third text based on the form. The server may label a text unit corresponding to an item name as the first text, a text unit corresponding to an item value as the second text, and text units corresponding to others as the third text among the text units. The server may apply the text unit included in the set group to a BERT-based natural language processing model and label the text unit as the first text to the third text corresponding to the item name, the item value, or others.
(45) In operation S**500**, the server may structure the text by mapping the labeled first text and second text. The server may perform the mapping based on the location information of the first text and the second text. The server may identify fourth text which is the first text that falls within the preset distance threshold value with the second text to map the fourth text and the second text, and when there are the plurality of pieces of fourth text, the fourth text having the highest similarity

may be mapped to the second text based on the vector similarity.
(46) In operation S**500**, when identifying the fourth text, the server processes the mapping process under the assumption that in the format of the form, the first text has a larger y value or a smaller x value in the location information than the second text.
(47) In operation S**600**, the server may determine whether the first text is misrecognized and correct the first text. The server may compare the first text and the representative keyword for the item name in the training data of the natural language processing model. Since the keyword for the item name in the training data includes all keywords that are modified to prevent the misrecognition, it is possible to determine whether or not the misrecognition occurs by comparing the first text with an unmodified representative keyword.
(48) According to the present invention as described above, it is possible to efficiently operationally manage an image document through classification of an image document, extraction of text (item name and item value) in the image document, data structuring of text, correction of misrecognition, and the like. In addition, according to the present invention, it is possible to effectively respond to training data loading and meta-information that is differently applied according to field conditions.
(49) Embodiments of the present invention described in the present specification and shown in the accompanying drawings are only specific examples provided in order to easily describe technical contents of the present invention and assist in the understanding of the present invention, and are not intended to limit the scope of the present invention. It is obvious to those of ordinary skill in the art to which the present invention pertains that other modifications based on the technical idea of the present invention can be implemented in addition to the embodiments disclosed herein.

Claims

1. An apparatus for data structuring of text, the apparatus comprising: a processor; and a memory storing instructions executable by the processor, wherein the processor is configured to execute the instructions to: extract text and location information of the text from an image, set text units for the extracted text, assign a first tag and a second tag to at least one of the text units, connect text units with related tags among the text units allocated the first tag and the second tag, label the connected text units as first text corresponding to an item name, second text corresponding to an item value, and third text corresponding to others based on a natural language processing model, and structure the extracted text by mapping the second text to the first text, wherein the processor is configured to identify fourth text, which is the first text that falls within a preset distance threshold value from the second text, and maps the first text corresponding to the fourth text and the second text in the process of structuring.
2. The apparatus of claim 1, wherein the processor is configured to: extract line information from the image, and set the text units based on the extracted line information.
3. The apparatus of claim 1, wherein the processor is configured to set text whose distance between keywords in the extracted text is less than or equal to a preset threshold value to one text unit, wherein the distance between the keywords is based on the location information of the text.
4. The apparatus of claim 1, wherein the first tag includes one of tags that indicate the item name and the item value, and the second tag includes one of tags that indicate start and middle.
5. The apparatus of claim 1, wherein the processor is configured to: compare the extracted text and text included in each of form samples pre-stored, identify a form sample including text that is the most similar to the extracted text based on the comparison result, classify a form of the image based on the identified form sample, and use an item name corresponding to the classified form in the process of labeling.
6. The apparatus of claim 1, wherein when a plurality of fourth texts are identified, the processor is configured to calculate vector similarity of the second text and each of the plurality of fourth texts

through a similarity verification model, and the first text corresponding to a fourth text having highest vector similarity among the plurality of fourth texts is mapped to the second text, and wherein each of the plurality of fourth texts is the first text falling within the preset distance threshold value from the second text.

7. The apparatus of claim 1, wherein the natural language processing model is pre-trained based on training data labeled with the item name, the item value, or others.

8. The apparatus of claim 7, wherein the natural language processing model modifies and trains the training data for characters with a high misrecognition occurrence rate, to prepare for a case where the extracted text is misrecognized.

9. The apparatus of claim 8, wherein keyword for the item name in the training data includes representative keyword that is not modified and keywords that are modified to prevent the misrecognition.

10. The apparatus of claim 9, wherein the processor is configured to: calculate similarity between the representative keyword and the first text, determine that, when the representative keyword and the first text are not identical to each other but the calculated similarity is greater than or equal to a preset similarity threshold, the first text is misrecognized, and correct the first text determined to be misrecognized as the representative keyword.

11. A method of data structuring of text, the method comprising: extracting text and location information of the text from an image; setting text units for the extracted text; assigning a first tag and a second tag to at least one of the text units; connecting text units with related tags among the text units allocated the first tag and the second tag; labeling the connected text units as first text corresponding to an item name, second text corresponding to an item value, and third text corresponding to others based on a natural language processing model; and structuring the extracted text by mapping the second text to the first text, wherein the method further comprises identifying fourth text, which is the first text that falls within a preset distance threshold value from the second text, and mapping the first text corresponding to the fourth text and the second text in the process of structuring.

12. A non-transitory computer-readable recording medium containing instructions for causing a computer to execute the method according to claim 11.

13. The method of claim 11, wherein the method comprises, when a plurality of fourth texts are identified, calculating vector similarity of the second text and each of the plurality of fourth texts through a similarity verification model, and the first text corresponding to a fourth text having highest vector similarity among the plurality of fourth texts is mapped to the second text, and wherein each of the plurality of fourth texts is the first text falling within the preset distance threshold value from the second text.