(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)公開番号
特開2025-21073
(P2025-21073A)
(43)公開日　令和7年2月13日(2025.2.13)

(51)国際特許分類		FI		テーマコード（参考）
G06F	*16/383(2019.01)*	G06F	16/383	5B175
G06Q	*30/06 (2023.01)*	G06Q	30/06	5E555
G06F	*3/0487(2013.01)*	G06F	3/0487	5L030
				5L049

審査請求　未請求　請求項の数　9　ＯＬ　（全25頁）

(21)出願番号　特願2023-124789(P2023-124789)
(22)出願日　令和5年7月31日(2023.7.31)

(71)出願人　000136136
株式会社ＰＦＵ
石川県かほく市宇野気ヌ９８番地の２
(74)代理人　100099759
弁理士　青木 篤
(74)代理人　100123582
弁理士　三橋 真二
(74)代理人　100114018
弁理士　南山 知広
(74)代理人　100180806
弁理士　三浦 剛
(72)発明者　深澤 貴彦
石川県かほく市宇野気ヌ９８番地の２
株式会社ＰＦＵ内

Ｆターム（参考）　5B175　DA01 FA01 FB03 GC03

最終頁に続く

(54)【発明の名称】　画像処理装置、画像処理システム、画像処理方法及び制御プログラム

(57)【要約】
【課題】スキャン画像から認識した文字を、利用者の作業負担を軽減させつつ、効率良く管理することが可能な画像処理装置、画像処理システム、画像処理方法及び制御プログラムを提供する。
【解決手段】画像処理装置は、スキャン画像から認識した文字と、利用者から受け付けた、スキャン画像から文字を認識する目的とを取得する第１取得部と、第１取得部が取得した文字及び目的を含む指示構文を生成する生成部と、指示構文を大規模言語モデルに入力し、大規模言語モデルから出力された出力情報を取得する第２取得部と、出力情報に関する情報を出力する出力制御部と、を有する。
【選択図】図４

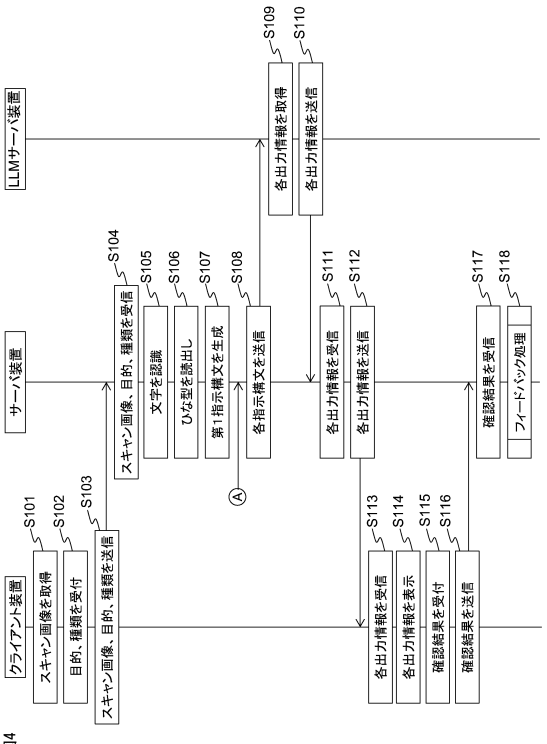


図4

【特許請求の範囲】
【請求項１】
スキャン画像から認識した文字と、利用者から受け付けた、前記スキャン画像から文字を認識する目的とを取得する第１取得部と、
前記第１取得部が取得した文字及び目的を含む指示構文を生成する生成部と、
前記指示構文を大規模言語モデルに入力し、大規模言語モデルから出力された出力情報を取得する第２取得部と、
前記出力情報に関する情報を出力する出力制御部と、
を有することを特徴とする画像処理装置。
【請求項２】
前記第１取得部は、利用者から受け付けた、前記スキャン画像に含まれる媒体の種類をさらに取得し、
前記生成部は、前記第１取得部が取得した媒体の種類をさらに含むように前記指示構文を生成する、請求項１に記載の画像処理装置。
【請求項３】
前記第１取得部は、利用者から受け付けた前記出力情報に対する確認結果を取得し、
前記生成部は、前記第１取得部が取得した確認結果に基づいて、前記指示構文を修正し、
前記第２取得部は、前記修正された指示構文を大規模言語モデルに入力する、請求項１または２に記載の画像処理装置。
【請求項４】
指示構文のひな型を記憶する記憶部をさらに有し、
前記生成部は、前記ひな型に従って前記指示構文を生成し、
前記第１取得部は、利用者から受け付けた前記出力情報に対する確認結果を取得し、
前記取得した確認結果に基づいて、前記ひな型を更新する更新部をさらに有する、請求項１または２に記載の画像処理装置。
【請求項５】
媒体の種類毎に指示構文のひな型を記憶する記憶部をさらに有し、
前記第１取得部は、前記取得した文字に基づいて、前記スキャン画像に含まれる媒体の種類を取得する、請求項１または２に記載の画像処理装置。
【請求項６】
前記生成部は、前記出力情報に基づいて、ファイル名の提案を要求する第２指示構文を生成し、
前記第２取得部は、前記第２指示構文を大規模言語モデルに入力し、大規模言語モデルから出力された第２出力情報を取得し、
前記出力制御部は、前記第２出力情報に関する情報を出力する、請求項１または２に記載の画像処理装置。
【請求項７】
複数の装置を有する画像処理システムであって、
前記複数の装置の何れかが、スキャン画像から認識した文字と、利用者から受け付けた、前記スキャン画像から文字を認識する目的とを取得する第１取得部を有し、
前記複数の装置の何れかが、前記第１取得部が取得した文字及び目的を含む指示構文を生成する生成部を有し、
前記複数の装置の何れかが、前記指示構文を大規模言語モデルに入力し、大規模言語モデルから出力された出力情報を取得する取得部を有し、
前記複数の装置の何れかが、前記出力情報に関する情報を出力する出力制御部を有する、
ことを特徴とする画像処理システム。
【請求項８】
スキャン画像から認識した文字と、利用者から受け付けた、前記スキャン画像から文字

を認識する目的とを取得し、
前記取得した文字及び目的を含む指示構文を生成し、
前記指示構文を大規模言語モデルに入力し、大規模言語モデルから出力された出力情報を取得し、
前記出力情報に関する情報を出力する、
ことを特徴とする画像処理方法。
【請求項９】
コンピュータの制御プログラムであって、
スキャン画像から認識した文字と、利用者から受け付けた、前記スキャン画像から文字を認識する目的とを取得し、
前記取得した文字及び目的を含む指示構文を生成し、
前記指示構文を大規模言語モデルに入力し、大規模言語モデルから出力された出力情報を取得し、
前記出力情報に関する情報を出力する、
ことを前記コンピュータに実行させることを特徴とする制御プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、画像処理装置、画像処理システム、画像処理方法及び制御プログラムに関する。
【背景技術】
【０００２】
近年、帳票等の媒体のデータ入力業務の自動化のために、媒体が撮像されたスキャン画像から文字を認識する画像処理装置が開発されている。このような画像処理装置は、媒体に含まれる文字を適切に管理するために、認識した文字のうち、管理すべき項目（会社名、金額等）を示す文字を特定する必要がある。
【０００３】
言語モデルを効率的に使用するためのプロンプトを開発および最適化するために、プロンプトエンジニアリングが使用されている（非特許文献１を参照）。
【先行技術文献】
【非特許文献】
【０００４】
【非特許文献１】「Prompt Engineering Guide」、インターネット＜URL:https://www.promptingguide.ai/jp＞
【発明の概要】
【発明が解決しようとする課題】
【０００５】
画像処理装置では、スキャン画像から認識した文字を、利用者の作業負担を軽減させつつ、効率良く管理することが望まれている。
【０００６】
本発明の目的は、スキャン画像から認識した文字を、利用者の作業負担を軽減させつつ、効率良く管理することが可能な画像処理装置、画像処理システム、画像処理方法及び制御プログラムを提供することにある。
【課題を解決するための手段】
【０００７】
本発明の一側面に係る画像処理装置は、スキャン画像から認識した文字と、利用者から受け付けた、スキャン画像から文字を認識する目的とを取得する第１取得部と、第１取得部が取得した文字及び目的を含む指示構文を生成する生成部と、指示構文を大規模言語モデルに入力し、大規模言語モデルから出力された出力情報を取得する第２取得部と、出力情報に関する情報を出力する出力制御部と、を有する。

【０００８】
本発明の一側面に係る画像処理システムは、複数の装置を有する画像処理システムであって、複数の装置の何れかが、スキャン画像から認識した文字と、利用者から受け付けた、スキャン画像から文字を認識する目的とを取得する第１取得部を有し、複数の装置の何れかが、第１取得部が取得した文字及び目的を含む指示構文を生成する生成部を有し、複数の装置の何れかが、指示構文を大規模言語モデルに入力し、大規模言語モデルから出力された出力情報を取得する取得部を有し、複数の装置の何れかが、出力情報に関する情報を出力する出力制御部を有する。
【０００９】
本発明の一側面に係る画像処理方法は、スキャン画像から認識した文字と、利用者から受け付けた、スキャン画像から文字を認識する目的とを取得し、取得した文字及び目的を含む指示構文を生成し、指示構文を大規模言語モデルに入力し、大規模言語モデルから出力された出力情報を取得し、出力情報に関する情報を出力する。
【００１０】
本発明の一側面に係る制御プログラムは、コンピュータの制御プログラムであって、スキャン画像から認識した文字と、利用者から受け付けた、スキャン画像から文字を認識する目的とを取得し、取得した文字及び目的を含む指示構文を生成し、指示構文を大規模言語モデルに入力し、大規模言語モデルから出力された出力情報を取得し、出力情報に関する情報を出力することをコンピュータに実行させる。
【発明の効果】
【００１１】
本発明によれば、画像処理装置、画像処理システム、画像処理方法及び制御プログラムは、スキャン画像から認識した文字を、利用者の作業負担を軽減させつつ、効率良く管理することが可能となる。
【図面の簡単な説明】
【００１２】
【図１】実施形態に従った画像処理システム１の概略構成を示す図である。
【図２】第３記憶装置３１０及び第３処理回路３２０の概略構成を示す図である。
【図３】ひな型テーブルのデータ構造について説明するための模式図である。
【図４】ファイル生成処理の動作の例を示すシーケンスである。
【図５】スキャン画像５００の一例を示す模式図である。
【図６】（Ａ）はひな型６００の一例を示す模式図であり、（Ｂ）は第１指示構文６１０の一例を示す模式図である。
【図７】第１出力情報７００の一例を示す模式図である。
【図８】フィードバック処理の動作の例を示すフローチャートである。
【図９】（Ａ）は第２指示構文９００の一例を示す模式図であり、（Ｂ）は第２出力情報９１０の一例を示す模式図である。
【図１０】第１出力情報１０００を示す模式図である。
【図１１】他のファイル生成処理の動作の一部の例を示すシーケンスである。
【図１２】他の第３処理回路５２０の概略構成を示すブロック図である。
【発明を実施するための形態】
【００１３】
以下、本発明の一側面に係る画像処理装置、画像処理システム、画像処理方法及び制御プログラムについて図を参照しつつ説明する。但し、本発明の技術的範囲はそれらの実施の形態に限定されず、特許請求の範囲に記載された発明とその均等物に及ぶ点に留意されたい。
【００１４】
図１は、実施形態に従った画像処理システム１の概略構成を示す図である。
【００１５】
図１に示すように、画像処理システム１は、画像読取装置１００、クライアント装置２

００、サーバ装置３００及びＬＬＭ（Large Language Models）サーバ装置４００等を有する。画像読取装置１００、クライアント装置２００、サーバ装置３００及びＬＬＭサーバ装置４００は、ネットワークＮを介して、相互に通信接続されている。ネットワークＮは、インターネット、イントラネット等である。画像読取装置１００、クライアント装置２００及びサーバ装置３００は、画像処理装置の一例である。ＬＬＭサーバ装置４００は、大規模言語モデルが記憶されたサーバである。大規模言語モデルは、例えばＣｈａｔＧＰＴ（Chat Generative Pre-trained Transformer）で用いられるＧＰＴ－３、ＧＰＴ－３．５、ＧＰＴ－４等である。ＬＬＭサーバ装置４００は、クラウドネットワークに設けられたサーバでもよい。

【００１６】

画像読取装置１００は、例えばスキャナ装置である。スキャナ装置は、原稿である媒体を搬送しながら撮像するＡＤＦ（Auto Document Feeder）タイプのスキャナ装置、又は、媒体を固定して撮像するフラットベッドタイプのスキャナ装置を含む。媒体は、レシート、請求書もしくは納品書等の帳票、又は、名刺、契約書もしくは説明書等の文書等、様々な種類の媒体を含む。画像読取装置１００は、媒体を撮像する携帯電話、スマートフォン、タブレット型コンピュータ、ノート型パーソナルコンピュータ等でもよい。

【００１７】

画像読取装置１００は、第１通信装置１０１と、第１入力装置１０２と、第１表示装置１０３と、撮像装置１０４と、第１記憶装置１１０と、第１処理回路１２０とを有する。

【００１８】

第１通信装置１０１は、有線ＬＡＮ（Local Area Network）等の所定の通信プロトコルに従って有線通信回線を通じて信号の送受信を行うための有線通信インタフェース回路を有する。第１通信装置１０１は、第１処理回路１２０からの指示に従って、クライアント装置２００と画像データ及び各種の情報を送受信する。なお、第１通信装置１０１は、無線信号を送受信するアンテナと、無線ＬＡＮ等の所定の通信プロトコルに従って無線通信回線を通じて信号の送受信を行うための無線通信インタフェース回路とを有してもよい。

【００１９】

第１入力装置１０２は、ボタン等の入力デバイス及び入力デバイスから信号を取得するインタフェース回路を有し、利用者による操作を受け付け、利用者の操作に応じた信号を第１処理回路１２０に出力する。

【００２０】

第１表示装置１０３は、液晶、有機ＥＬ（Electro-Luminescence）等を含むディスプレイ及びディスプレイに画像データを出力するインタフェース回路を有し、第１処理回路１２０からの指示に従って、画像データをディスプレイに表示する。

【００２１】

撮像装置１０４は、媒体を撮像したスキャン画像を生成する。撮像装置１０４は、一次元又は二次元に配列されたＣＣＤ（Charge Coupled Device）による撮像素子を備える縮小光学系タイプの撮像センサを有する。さらに、撮像装置１０４は、光を照射する光源と、撮像素子上に像を結ぶレンズと、撮像素子から出力された電気信号を増幅してアナログ／デジタル（Ａ／Ｄ）変換するＡ／Ｄ変換器とを有する。撮像装置１０４において、撮像センサは、媒体の表面及び／又は裏面を撮像してアナログの画像信号を生成して出力し、Ａ／Ｄ変換器は、このアナログの画像信号をＡ／Ｄ変換してデジタルのスキャン画像を生成して第１処理回路１２０に出力する。なお、ＣＣＤの代わりにＣＭＯＳ（Complementary Metal Oxide Semiconductor）による撮像素子を備える縮小光学系タイプの撮像センサが用いられてもよい。また、ＣＭＯＳ又はＣＣＤによる撮像素子を備える等倍光学系タイプのラインセンサが利用されてもよい。

【００２２】

第１記憶装置１１０は、記憶部の一例であり、ＲＡＭ（Random Access Memory）、ＲＯＭ（Read Only Memory）等のメモリ装置、ハードディスク等の固定ディスク

装置、又はフレキシブルディスク、光ディスク等の可搬用の記憶装置等を有する。また、第１記憶装置１１０には、画像読取装置１００の各種処理に用いられるコンピュータプログラム、データベース、テーブル等が格納される。コンピュータプログラムは、コンピュータ読み取り可能な可搬型記録媒体から公知のセットアッププログラム等を用いて第１記憶装置１１０にインストールされてもよい。可搬型記録媒体は、例えばＣＤ－ＲＯＭ（compact disk read only memory）、ＤＶＤ－ＲＯＭ（digital versatile disk read only memory）等である。また、コンピュータプログラムは、サーバ等から配信されて第１記憶装置１１０にインストールされてもよい。

【００２３】

第１処理回路１２０は、予め第１記憶装置１１０に記憶されているプログラムに基づいて動作する。第１処理回路１２０は、例えばＣＰＵ（Control Processing Unit）である。なお、第１処理回路１２０として、ＤＳＰ（digital signal processor）、ＬＳＩ（large scale integration）、ＡＳＩＣ（Application Specific Integrated Circuit）、ＦＰＧＡ（Field-Programming Gate Array）等が用いられてもよい。

【００２４】

第１処理回路１２０は、第１通信装置１０１、第１入力装置１０２、第１表示装置１０３、撮像装置１０４及び第１記憶装置１１０等と接続され、これらの各部を制御する。第１処理回路１２０は、第１通信装置１０１を介したクライアント装置２００とのデータ送受信制御、第１入力装置１０２の入力制御、第１表示装置１０３の表示制御、撮像装置１０４の撮像制御等を行う。

【００２５】

クライアント装置２００は、パーソナルコンピュータ、ノート型パーソナルコンピュータ、タブレット型コンピュータ、スマートフォン等である。

【００２６】

クライアント装置２００は、第２通信装置２０１と、第２入力装置２０２と、第２表示装置２０３と、第２記憶装置２１０と、第２処理回路２２０とを有する。

【００２７】

第２通信装置２０１は、有線ＬＡＮ等の所定の通信プロトコルに従って有線通信回線を通じて信号の送受信を行うための有線通信インタフェース回路を有する。第２通信装置２０１は、第２処理回路２２０からの指示に従って、画像読取装置１００及びサーバ装置３００と画像データ及び各種の情報を送受信する。なお、第２通信装置２０１は、無線信号を送受信するアンテナと、無線ＬＡＮ等の所定の通信プロトコルに従って無線通信回線を通じて信号の送受信を行うための無線通信インタフェース回路とを有してもよい。

【００２８】

第２入力装置２０２は、キーボード、マウス等の入力装置及び入力装置から信号を取得するインタフェース回路を有し、利用者の操作に応じた信号を第２処理回路２２０に出力する。

【００２９】

第２表示装置２０３は、液晶、有機ＥＬ等を含むディスプレイ及びディスプレイに画像データを出力するインタフェース回路を有し、第２処理回路２２０からの指示に従って、各種の情報をディスプレイに表示する。

【００３０】

第２記憶装置２１０は、記憶部の一例であり、ＲＡＭ、ＲＯＭ等のメモリ装置、ハードディスク等の固定ディスク装置、又はフレキシブルディスク、光ディスク等の可搬用の記憶装置等を有する。第２記憶装置２１０には、クライアント装置２００の各種処理に用いられるコンピュータプログラム、データベース、テーブル等が格納される。コンピュータプログラムは、例えばＣＤ－ＲＯＭ、ＤＶＤ－ＲＯＭ等のコンピュータ読み取り可能な可搬型記録媒体から、公知のセットアッププログラム等を用いて第２記憶装置２１０にインストールされてもよい。また、コンピュータプログラムは、サーバ等から配信されて第２記憶装置２１０にインストールされてもよい。

【００３１】
第２処理回路２２０は、予め第２記憶装置２１０に記憶されているプログラムに基づいて動作する。第２処理回路２２０は、例えばＣＰＵである。なお、第２処理回路２２０として、ＤＳＰ、ＬＳＩ、ＡＳＩＣ、ＦＰＧＡ等が用いられてもよい。
【００３２】
第２処理回路２２０は、第２通信装置２０１、第２入力装置２０２、第２表示装置２０３及び第２記憶装置２１０等と接続され、これらの各部を制御する。第２処理回路２２０は、第２通信装置２０１を介した画像読取装置１００及びサーバ装置３００とのデータ送受信制御、第２入力装置２０２の入力制御、第２表示装置２０３の表示制御等を行う。
【００３３】
サーバ装置３００は、画像処理システム１の管理者により使用されるサーバである。サーバ装置３００は、クラウドネットワークに設けられたサーバでもよい。
【００３４】
サーバ装置３００は、第３通信装置３０１と、第３入力装置３０２と、第３表示装置３０３と、第３記憶装置３１０と、第３処理回路３２０とを有する。
【００３５】
第３通信装置３０１は、有線ＬＡＮ等の所定の通信プロトコルに従って有線通信回線を通じて信号の送受信を行うための有線通信インタフェース回路を有する。第３通信装置３０１は、第３処理回路３２０からの指示に従って、画像読取装置１００及びサーバ装置３００と画像データ及び各種の情報を送受信する。なお、第３通信装置３０１は、無線信号を送受信するアンテナと、無線ＬＡＮ等の所定の通信プロトコルに従って無線通信回線を通じて信号の送受信を行うための無線通信インタフェース回路とを有してもよい。
【００３６】
第３入力装置３０２は、キーボード、マウス等の入力装置及び入力装置から信号を取得するインタフェース回路を有し、利用者の操作に応じた信号を第３処理回路３２０に出力する。
【００３７】
第３表示装置３０３は、液晶、有機ＥＬ等を含むディスプレイ及びディスプレイに画像データを出力するインタフェース回路を有し、第３処理回路３２０からの指示に従って、各種の情報をディスプレイに表示する。
【００３８】
第３記憶装置３１０は、記憶部の一例であり、ＲＡＭ、ＲＯＭ等のメモリ装置、ハードディスク等の固定ディスク装置、又はフレキシブルディスク、光ディスク等の可搬用の記憶装置等を有する。第３記憶装置３１０には、サーバ装置３００の各種処理に用いられるコンピュータプログラム、データベース、テーブル等が格納される。コンピュータプログラムは、例えばＣＤ－ＲＯＭ、ＤＶＤ－ＲＯＭ等のコンピュータ読み取り可能な可搬型記録媒体から、公知のセットアッププログラム等を用いて第３記憶装置３１０にインストールされてもよい。また、コンピュータプログラムは、サーバ等から配信されて第３記憶装置３１０にインストールされてもよい。
【００３９】
第３記憶装置３１０には、データとして、ひな型テーブル等が予め記憶される。ひな型テーブルの詳細については後述する。
【００４０】
第３処理回路３２０は、予め第３記憶装置３１０に記憶されているプログラムに基づいて動作する。第３処理回路３２０は、例えばＣＰＵである。なお、第３処理回路３２０として、ＤＳＰ、ＬＳＩ、ＡＳＩＣ、ＦＰＧＡ等が用いられてもよい。
【００４１】
第３処理回路３２０は、第３通信装置３０１、第３入力装置３０２、第３表示装置３０３及び第３記憶装置３１０等と接続され、これらの各部を制御する。第３処理回路３２０は、第３通信装置３０１を介したクライアント装置２００及びＬＬＭサーバ装置４００と

のデータ送受信制御、第３入力装置３０２の入力制御、第３表示装置３０３の表示制御等を行う。

【００４２】

図２は、第３記憶装置３１０及び第３処理回路３２０の概略構成を示す図である。

【００４３】

図２に示すように、第３記憶装置３１０には、第１取得プログラム３１１、生成プログラム３１２、第２取得プログラム３１３、出力制御プログラム３１４及び更新プログラム３１５等が記憶される。これらの各プログラムは、プロセッサ上で動作するソフトウェアにより実装される機能モジュールである。第３処理回路３２０は、第３記憶装置３１０に記憶された各プログラムを読み取り、読み取った各プログラムに従って動作することにより、第１取得部３２１、生成部３２２、第２取得部３２３、出力制御部３２４及び更新部３２５として機能する。

【００４４】

図３は、ひな型テーブルのデータ構造について説明するための模式図である。

【００４５】

図３に示すように、ひな型テーブルには、媒体の種類毎に、第１指示構文のひな型が記憶される。媒体の種類は、例えばレシート、請求書、納品書、名刺、契約書、説明書等である。第１指示構文は、指示構文の一例である。第１指示構文は、大規模言語モデルに入力する入力構文であり、大規模言語モデルに対する指示又は質問を表現するためのテキストの構文（テキスト文）、即ち大規模言語モデルが回答するために提示する質問又は命令文である。第１指示構文は、例えば、ＣｈａｔＧＰＴのプロンプトである。第１指示構文のひな型は、媒体の種類に応じた、第１指示構文のテンプレートである。

【００４６】

図４は、ファイル生成処理の動作の例を示すシーケンスである。

【００４７】

以下、図４に示したシーケンスを参照しつつ、画像処理システム１のファイル生成処理の動作の例を説明する。なお、以下に説明する動作は、画像処理システム１が有する各装置の各記憶装置に予め記憶されているプログラムに基づき主に各装置の各処理回路により各装置の各要素と協働して実行される。

【００４８】

最初に、クライアント装置２００の第２処理回路２２０は、スキャン画像を取得する（ステップＳ１０１）。第２処理回路２２０は、第２通信装置２０１を介して画像読取装置１００に、スキャン画像の取得を要求する要求信号を送信する。画像読取装置１００の第１処理回路１２０は、第１通信装置１０１を介してクライアント装置２００から要求信号を受信した場合、撮像装置１０４が生成したスキャン画像を、第１通信装置１０１を介してクライアント装置２００に送信する。第２処理回路２２０は、第２通信装置２０１を介して画像読取装置１００からスキャン画像を受信することにより取得する。スキャン画像は、画像読取装置１００により生成された画像に限定されず、媒体を撮像した画像であれば、どのような画像でもよい。第２処理回路２２０は、不図示のインタフェース回路を介して利用者から入力されたスキャン画像を受信することにより取得してもよい。また、第２処理回路２２０は、第２記憶装置２１０から予め記憶されていたスキャン画像を読み出すことにより取得してもよい。

【００４９】

図５は、スキャン画像５００の一例を示す模式図である。

【００５０】

図５に示すスキャン画像５００には、媒体としてレシートが撮像されている。スキャン画像５００には、店舗情報５０１、概要情報５０２、購入情報５０３、付属情報５０４等が含まれている。店舗情報５０１には、そのレシートを発行した店舗の名称、住所、電話番号、営業時間等が含まれている。概要情報５０２には、レシートを発行した日時、レシートの識別番号、購入者による支払方法等が含まれている。購入情報５０３には、購入し

た商品毎の商品名、単価、数量及び小計と、その合計金額と、消費税と、税込金額と、預かり金額と、お釣りとが含まれている。付属情報５０４には、ポイント会員を募集する旨等が含まれている。

【００５１】

次に、第２処理回路２２０は、利用者から、スキャン画像から文字を認識する目的と、スキャン画像に含まれる媒体の種類とを受け付ける（ステップＳ１０２）。第２処理回路２２０は、利用者により第２入力装置２０２を用いて入力された、スキャン画像から文字を認識する目的、及び、スキャン画像に含まれる媒体の種類の指定を受け付ける。スキャン画像から文字を認識する目的として、どのような背景によりデータを取得するかが指定される。スキャン画像から文字を認識する目的は、「経費申請のためのデータ取得」、「請求書処理におけるデータ入力に必要な項目取得」、「住所録の作成」、「納品物の整理」、「名刺管理」、「契約締結のための重要事項の抽出」等、任意の形式を有する。第２処理回路２２０は、スキャン画像から文字を認識する目的及び／又は媒体の種類として、予め定められた複数の候補を選択可能に第２表示装置２０３に表示し、利用者による選択を受け付けてもよい。

【００５２】

次に、第２処理回路２２０は、第２通信装置２０１を介してサーバ装置３００に、ステップＳ１０１で取得したスキャン画像、並びに、ステップＳ１０２で受け付けた目的及び媒体の種類を示す情報を送信する（ステップＳ１０３）。

【００５３】

一方、サーバ装置３００の第１取得部３２１は、第３通信装置３０１を介してクライアント装置２００から、スキャン画像、並びに、スキャン画像から文字を認識する目的、及び、スキャン画像に含まれる媒体の種類を示す情報を受信する（ステップＳ１０４）。これにより、第１取得部３２１は、スキャン画像、並びに、利用者から受け付けた、スキャン画像から文字を認識する目的、及び、スキャン画像に含まれる媒体の種類を取得する。なお、第１取得部３２１は、不図示のインタフェース回路を介して利用者から入力されたスキャン画像を受信することにより取得してもよい。また、第１取得部３２１は、第３記憶装置３１０から予め記憶されていたスキャン画像を読み出すことにより取得してもよい。また、第１取得部３２１は、利用者により第３入力装置３０２を用いて入力された、スキャン画像から文字を認識する目的、及び、スキャン画像に含まれる媒体の種類の指定を受け付けてもよい。

【００５４】

次に、第１取得部３２１は、スキャン画像に対してＯＣＲ（Optical Character Recognition）処理を実行し、スキャン画像から文字を認識する（ステップＳ１０５）。これにより、第１取得部３２１は、スキャン画像から認識した文字を取得する。また、第１取得部３２１は、スキャン画像から認識した各文字（文字列）を所定の順序に従って並べ替える（例えば左上の文字から右側に向かって順次走査していき、改行が検出された場合に次の行を走査していく）。第１取得部３２１は、スキャン画像から認識した文字のうち、予め登録された帳票認識に不要な文言、又は、予め登録された文言以外の帳票認識に不要な文言を削除してもよい。

【００５５】

次に、生成部３２２は、ひな型テーブルから、ステップＳ１０４で取得した媒体の種類に対応するひな型を読み出す（ステップＳ１０６）。

【００５６】

図６（Ａ）は、ひな型６００の一例を示す模式図である。

【００５７】

図６（Ａ）に示すひな型６００は、レシートに対応するひな型である。ひな型６００には、指令文６０１、種類指定文６０２、目的指定文６０３、出力形式指定文６０４、ＯＣＲ結果指定文６０５等が含まれている。指令文６０１は、大規模言語モデルに対する指令であり、指令文６０１として、例えば、指定した種類の媒体に対するＯＣＲ結果から、指

定した目的に応じて、適切な項目を抽出し、指定した出力形式で出力する指令が記述される。種類指定文６０２は、スキャン画像に含まれる媒体の種類を指定する文である。種類指定文６０２には、そのひな型に対応する媒体の種類が予め記述されていてもよい。目的指定文６０３は、スキャン画像から文字を認識する目的を指定する文である。出力形式指定文６０４は、出力形式を指定する文である。出力形式は、ＪＳＯＮ形式、ＣＳＶ形式、ＸＭＬ形式等である。出力形式指定文６０４には、予め定められた形式が予め記述されていてもよい。ＯＣＲ結果指定文６０５は、スキャン画像から認識した文字を指定する文である。

【００５８】

次に、生成部３２２は、第１取得部３２１が取得した文字、目的及び媒体の種類を含む第１指示構文を生成する（ステップＳ１０７）。生成部３２２は、ステップＳ１０６で読み出したひな型に従って、第１指示構文を生成する。

【００５９】

図６（Ｂ）は、第１指示構文６１０の一例を示す模式図である。

【００６０】

図６（Ｂ）に示す第１指示構文６１０は、図６（Ａ）に示したひな型６００に従って、図５に示したスキャン画像５００から認識された文字が含まれるように生成された第１指示構文である。図６（Ｂ）に示すように、指令文６０１は、ひな型６００から変更されていない。種類指定文６０２及び目的指定文６０３には、それぞれ第１取得部３２１が取得した目的及び媒体の種類が記述される。出力形式指定文６０４には特定（所望）の出力形式が記述される。ひな型６００において種類指定文６０２及び／又は出力形式指定文６０４に媒体の種類が予め記述されている場合、種類指定文６０２及び／又は出力形式指定文６０４は変更されなくてもよい。ＯＣＲ結果指定文６０５には、第１取得部３２１が取得した文字が記述される。

【００６１】

次に、第２取得部３２３は、生成部３２２が生成した第１指示構文と、第１指示構文を大規模言語モデルに入力することを要求する要求情報とを、第３通信装置３０１を介してＬＬＭサーバ装置４００に送信する（ステップＳ１０８）。

【００６２】

一方、ＬＬＭサーバ装置４００は、サーバ装置３００から第１指示構文及び要求情報を受信する。ＬＬＭサーバ装置４００は、受信した第１指示構文を、ＬＬＭサーバ装置４００に記憶された大規模言語モデルに入力し、大規模言語モデルから出力された第１出力情報を取得する（ステップＳ１０９）。第１出力情報は、出力情報の一例である。

【００６３】

図７は、第１出力情報７００の一例を示す模式図である。

【００６４】

図７に示す第１出力情報７００は、大規模言語モデルに図６（Ｂ）に示した第１指示構文６１０が入力されたときに大規模言語モデルから出力された情報である。第１出力情報７００には、店舗情報７０１、概要情報７０２、購入情報７０３等が含まれている。店舗情報７０１には、スキャン画像５００の店舗情報５０１のうち、店舗の名称及び住所のみが含まれ、店舗の電話番号及び営業時間は含まれていない。概要情報７０２には、スキャン画像５００の概要情報５０２のうち、レシートの識別番号のみが含まれ、レシートを発行した日時、及び、購入者による支払方法は含まれていない。購入情報７０３には、スキャン画像５００の購入情報５０３のうち、購入した商品毎の商品名、単価、数量及び小計と、その合計金額と、消費税と、税込金額のみが含まれ、預かり金額及びお釣りは含まれていない。また、第１出力情報７００には、スキャン画像５００の付属情報５０４に対応する情報は含まれていない。

【００６５】

このように、第１出力情報には、ＯＣＲ結果から構造化されたデータが含まれる。第１出力情報には、スキャン画像から認識された文字のうち、スキャン画像から文字を認識す

る目的（経費申請のためのデータ抽出）に対応する文字のみが適切に含まれている。また、各項目について事前に設定されていないにもかかわらず、スキャン画像に文字として含まれていなかった「店名」、「住所」等の項目名が推定されて、第１出力情報に記述されている。さらに、文字の羅列でしかなかった商品の情報も、第１出力情報において表形式に整理されて区分されている。また、図７に示す第１出力情報は、ＪＳＯＮ形式で記述されているため、後段の処理において処理しやすい形式を有している。

【００６６】

次に、ＬＬＭサーバ装置４００は、取得した第１出力情報をサーバ装置３００に送信する（ステップＳ１１０）。

【００６７】

一方、サーバ装置３００の第２取得部３２３は、第３通信装置３０１を介してＬＬＭサーバ装置４００から第１出力情報を受信することにより取得する（ステップＳ１１１）。これにより、第２取得部３２３は、生成部３２２が生成した第１指示構文を大規模言語モデルに入力し、大規模言語モデルから出力された第１出力情報を取得する。なお、第２取得部３２３は、取得した第１出力情報を利用者が利用するシステムに適合するように変換又は編集してもよい。

【００６８】

次に、出力制御部３２４は、第２取得部３２３が取得した第１出力情報、又は、第２取得部３２３が変換又は編集した第１出力情報を、第３通信装置３０１を介してクライアント装置２００に送信することにより出力する（ステップＳ１１２）。第２取得部３２３が取得した第１出力情報、又は、第２取得部３２３が変換又は編集した第１出力情報は、出力情報に関する情報の一例である。なお、出力制御部３２４は、第２取得部３２３が取得した第１出力情報、又は、第２取得部３２３が変換又は編集した第１出力情報を、第３表示装置３０３に表示することにより出力してもよい。

【００６９】

一方、クライアント装置２００の第２処理回路２２０は、第２通信装置２０１を介してサーバ装置３００から第１出力情報を受信する（ステップＳ１１３）。

【００７０】

次に、第２処理回路２２０は、受信した第１出力情報を第２表示装置２０３に表示する（ステップＳ１１４）。

【００７１】

次に、第２処理回路２２０は、利用者から、第１出力情報に対する確認結果を受け付ける（ステップＳ１１５）。第２処理回路２２０は、利用者により第２入力装置２０２を用いて入力された確認結果の指定を受け付ける。確認結果には、例えば、第１出力情報が適切であることを示すＯＫ指示と、第１出力情報が適切でないことを示すＮＧ指示と、ファイル生成処理を完了することを指示する終了指示とが含まれる。確認結果がＮＧ指示を示す場合、第２処理回路２２０は、確認結果として、さらに利用者から、第１出力情報に対して修正すべき修正項目を受け付ける。修正項目として、例えば、第１出力情報に含まれる項目のうちの不要項目、及び／又は、第１出力情報に含まれていなかった不足項目が指定される。

【００７２】

次に、第２処理回路２２０は、第２通信装置２０１を介してサーバ装置３００に、利用者から受け付けた第１出力情報に対する確認結果を示す情報を送信する（ステップＳ１１６）。

【００７３】

一方、サーバ装置３００の第１取得部３２１は、第３通信装置３０１を介してクライアント装置２００から確認結果を示す情報を受信する（ステップＳ１１７）。これにより、第１取得部３２１は、利用者から受け付けた第１出力情報に対する確認結果を取得する。なお、第１取得部３２１は、利用者により第３入力装置３０２を用いて入力された確認結果を受信して取得してもよい。

【００７４】
次に、第３処理回路３２０は、第１取得部３２１が取得した確認結果に基づいて、フィードバック処理を実行する（ステップＳ１１８）。以上により、ファイル生成処理は、終了する。
【００７５】
図８は、フィードバック処理の動作の例を示すフローチャートである。フィードバック処理は、図４のステップＳ１１８において実行される。
【００７６】
最初に、第１取得部３２１は、取得した確認結果がＮＧ指示を示すか否かを判定する（ステップＳ２０１）。
【００７７】
確認結果がＮＧ指示を示す場合、生成部３２２は、第１取得部３２１が取得した確認結果に含まれる修正項目に基づいて、第１指示構文を修正し（ステップＳ２０２）、図４のステップＳ１０８へ処理を戻す。生成部３２２は、第１指示構文の指令文６０１に、確認結果に含まれる不要項目を抽出せず且つ／又は確認結果に含まれる不足項目を抽出する旨を追記するように、図４のステップＳ１０７で生成した第１指示構文を修正する。例えば、確認結果に不要項目「Ａ」が含まれる場合、第１指示構文の指令文に「Ａは必要ない項目なので出力には決して含ませないでください。」という一文が追加される。また、確認結果に不足項目「Ｂ」が含まれない場合、第１指示構文の指令文に「Ｂは必要な項目なので、必ず出力してください。」という一文が追加される。この場合、図４のステップＳ１０８では、第２取得部３２３は、修正された第１指示構文と、その第１指示構文を大規模言語モデルに入力することを要求する要求情報とをＬＬＭサーバ装置４００に送信し、大規模言語モデルに入力する。これにより、画像処理システム１は、利用者のフィードバックに基づいて、スキャン画像から必要な文字をより精度良く抽出することができる。
【００７８】
一方、確認結果がＮＧ指示を示していない場合、第１取得部３２１は、確認結果がＯＫ指示を示すか否かを判定する（ステップＳ２０３）。
【００７９】
確認結果がＯＫ指示を示す場合、第１取得部３２１は、現在処理中のスキャン画像について、生成部３２２が、確認結果に含まれる修正項目に基づいて第１指示構文を修正したか否かを判定する（ステップＳ２０４）。生成部３２２が、確認結果に含まれる修正項目に基づいて第１指示構文を修正していない場合、第１取得部３２１は、特に処理を実行せずにステップＳ２０６へ処理を移行する。
【００８０】
一方、生成部３２２が、確認結果に含まれる修正項目に基づいて第１指示構文を修正している場合、更新部３２５は、修正された第１指示構文に基づいて、第３記憶装置３１０に記憶された第１指示構文のひな型を更新する（ステップＳ２０５）。更新部３２５は、対応するひな型、即ち使用したひな型の指令文６０１を、生成部３２２が修正（追記）した指令文に置換することにより、第１指示構文のひな型を更新する。このように、更新部３２５は、第１取得部３２１が取得した確認結果に基づいて、第１指示構文のひな型を更新する。これにより、次回以降、更新されたひな型に従って第１指示構文が生成されるため、画像処理システム１は、利用者によるフィードバックを得る前に、スキャン画像から適切な文字を抽出できる。したがって、画像処理システム１は、利用者の手間を省きつつ、スキャン画像から必要な文字をより精度良く抽出することができる。
【００８１】
次に、生成部３２２は、大規模言語モデルに対してファイル名の提案を要求する第２指示構文を生成し（ステップＳ２０６）、図４のステップＳ１０８へ処理を戻す。生成部３２２は、第２取得部３２３が取得した第１出力情報に基づいて、第２指示構文を生成する。第１指示構文が生成される場合と同様に、サーバ装置３００は、第２指示構文のひな型を予め第３記憶装置３１０に記憶しておき、生成部３２２は、第３記憶装置３１０に記憶

されたひな型に従って、第２指示構文を生成してもよい。
【００８２】
図９（Ａ）は、第２指示構文９００の一例を示す模式図である。
【００８３】
図９（Ａ）に示す第２指示構文９００は、図７に示した第１出力情報７００に基づいて生成された第２指示構文である。図９（Ａ）に示すように、第２指示構文９００には、指令文９０１、ファイル名例指定文９０２、出力結果指定文９０３等が含まれている。指令文９０１は、大規模言語モデルに対する指令であり、指令文９０１として、例えば、第１出力情報と、ファイル名例とから、適切なファイル名を決定し出力する指令が記述される。ファイル名例指定文９０２には、ファイル名の例（フォーマット例）が記述される。指令文９０１及びファイル名例指定文９０２は、第２指示構文のひな型に記述されていてもよい。出力結果指定文９０３には、第２取得部３２３が取得した第１出力情報が記述される。
【００８４】
この場合、図４のステップＳ１０８では、第２取得部３２３は、新たに生成された第２指示構文と、その第２指示構文を大規模言語モデルに入力することを要求する要求情報とをＬＬＭサーバ装置４００に送信し、大規模言語モデルに入力する。図４のステップＳ１０９では、ＬＬＭサーバ装置４００は、サーバ装置３００から第２指示構文及び要求情報を受信する。ＬＬＭサーバ装置４００は、受信した第２指示構文を、ＬＬＭサーバ装置４００に記憶された大規模言語モデルに入力し、大規模言語モデルから出力された第２出力情報を取得する。
【００８５】
図９（Ｂ）は、第２出力情報９１０の一例を示す模式図である。
【００８６】
図９（Ｂ）に示す第２出力情報９１０は、大規模言語モデルに図９（Ａ）に示した第２指示構文９００が入力されたときに大規模言語モデルから出力された第２出力情報である。第２出力情報９１０には、大規模言語モデルが提案するファイル名等が含まれている。大規模言語モデルが提案するファイル名は、第２指示構文で指定されたファイル名に従って提案されている。
【００８７】
図４のステップＳ１１０では、ＬＬＭサーバ装置４００は、取得した第２出力情報をサーバ装置３００に送信する。図４のステップＳ１１１では、サーバ装置３００の第２取得部３２３は、第３通信装置３０１を介してＬＬＭサーバ装置４００から第２出力情報を受信することにより取得する。図４のステップＳ１１２では、出力制御部３２４は、第２取得部３２３が取得した第２出力情報、又は、第２取得部３２３が変換又は編集した第２出力情報を、第３通信装置３０１を介してクライアント装置２００に送信することにより出力する。第２取得部３２３が取得した第２出力情報、又は、第２取得部３２３が変換又は編集した第２出力情報は、第２出力情報に関する情報の一例である。このように、第２取得部３２３は、第２指示構文を大規模言語モデルに入力し、大規模言語モデルから出力された第２出力情報を取得し、出力制御部３２４は、第２出力情報に関する情報を出力する。これにより、画像処理システム１は、第１出力情報を適切に管理するためのファイル名を利用者に提案することができ、利用者の利便性を向上させることができる。
【００８８】
図４のステップＳ１１３では、クライアント装置２００の第２処理回路２２０は、第２通信装置２０１を介してサーバ装置３００から第２出力情報を受信する。図４のステップＳ１１４では、第２処理回路２２０は、受信した第２出力情報を第２表示装置２０３に表示する。図４のステップＳ１１４では、第２処理回路２２０は、利用者から、第２出力情報に対する確認結果として、ファイル生成処理を完了することを指示する終了指示を受け付ける。
【００８９】

図８に戻って、ステップＳ２０３で確認結果がＯＫ指示を示していなかった場合、即ち確認結果がＮＧ指示及びＯＫ指示の何れでもなかった場合、第１取得部３２１は、確認結果が終了指示であると判定する。その場合、生成部３２２は、利用者による確認結果がＯＫ指示であった第１出力情報を含むファイルを生成し、第２出力情報に示されるファイル名を付けて第３記憶装置３１０に保存する。そして、生成部３２２は、生成したファイルを、第３通信装置３０１を介してクライアント装置２００に送信することにより出力し（ステップＳ２０７）、一連のステップを終了する。なお、クライアント装置２００の第２処理回路２２０は、第２入力装置２０２を用いて、利用者から第１出力情報及び又は第２出力情報の修正指示を受け付けてもよい。その場合、生成部３２２は、利用者による修正指示に従って修正した第１出力情報を含むファイルを生成し、且つ／又は、利用者による修正指示に従って修正した第２出力情報に示されるファイル名を付けて第３記憶装置３１０に保存する。

【００９０】

なお、図４のステップＳ１０４において、第１取得部３２１は、利用者から受け付けた、スキャン画像に含まれる媒体の種類を取得しなくてもよい。その場合、例えば、第１取得部３２１は、ステップＳ１０５においてスキャン画像から認識することにより取得した文字に基づいて、スキャン画像に含まれる媒体の種類を特定する。例えば、サーバ装置３００は、媒体の種類毎に、各種類に対応する一又は複数のキーワードを予め第３記憶装置３１０に記憶しておく。例えば、媒体の種類が請求書である場合、キーワードとして「請求書」、「料金」等が設定される。また、媒体の種類が納品書である場合、キーワードとして「納品書」、「物品」等が設定される。また、媒体の種類がレシートである場合、キーワードとして「レシート」、「商品名」等が設定される。第１取得部３２１は、第３記憶装置３１０に記憶されたキーワードを参照し、取得した文字と一致するキーワードに対応する媒体の種類を、スキャン画像に含まれる媒体の種類として特定する。

【００９１】

または、第１取得部３２１は、公知の画像処理技術を利用して、スキャン画像から罫線を抽出する。サーバ装置３００は、媒体の種類毎に、各種類の媒体に対応する罫線が配置された画像を予め第３記憶装置３１０に記憶しておく。第１取得部３２１は、スキャン画像から抽出した罫線のみを配置した画像と、第３記憶装置３１０に記憶された各画像との類似度合いを算出し、類似度合いが最も高い画像に対応する種類をスキャン画像に含まれる画像の種類として特定する。類似度は、例えば正規化相互相関値等である。

【００９２】

または、第１取得部３２１は、文字（文字列）又は画像が入力された場合に、媒体の種類を出力するように事前学習された学習モデルにより、媒体の種類を特定してもよい。学習モデルは、例えばニューラルネットワーク、サポートベクターマシン等の教師あり学習により、様々な種類の媒体に含まれる各文字又は様々な種類の媒体が撮像された画像と各媒体の種類とのセットを用いて事前学習され、予め第３記憶装置３１０に記憶される。第１取得部３２１は、スキャン画像から認識した文字又はスキャン画像を学習モデルに入力し、学習モデルから出力された媒体の種類を、スキャン画像に含まれる媒体の種類として特定する。

【００９３】

これらにより、利用者は、スキャン画像に含まれる媒体の種類を指定する必要がなくなり、画像処理システム１は、利用者の手間を省き、利用者の利便性を向上させることができる。

【００９４】

また、第１取得部３２１は、媒体の種類を取得しなくてもよい。その場合、生成部３２２は、媒体の種類によらずに予め定められたひな型を用いて、媒体の種類を含まない第１指示構文を生成する。媒体の種類を含まない第１指示構文を用いる場合、媒体の種類を含む第１指示構文を用いる場合と比較して、大規模言語モデルによる必要な項目の抽出精度は低下する可能性がある。しかしながら、この場合も、第１指示構文に、スキャン画像か

ら文字を認識する目的が含まれることにより、大規模言語モデルは、必要な項目を精度良く抽出することができる。即ち、画像処理システム１は、第１指示構文に媒体の種類を含ませない場合、利用者の手間を省きつつスキャン画像から必要な項目を抽出でき、第１指示構文に媒体の種類を含ませる場合、スキャン画像から必要な項目をより高精度に抽出できる。

【００９５】

また、スキャン画像がボーンデジタルである場合、ステップＳ１０５の処理は省略されてもよい。その場合、第１取得部３２１は、スキャン画像に含まれる文字と、利用者から受け付けたスキャン画像に含まれる文字を利用する目的と、スキャン画像に含まれる媒体の種類とを取得する。生成部３２２は、第１取得部３２１が取得した文字、目的及び媒体の種類を含む第１指示構文を生成する。

【００９６】

また、図４のステップＳ１０６の処理は省略され、生成部３２２は、ひな型を用いずに第１指示構文を生成してもよい。また、図４のステップＳ１１５～Ｓ１１８の処理は省略され、第１取得部３２１は、利用者からの確認結果を受け付けなくてもよい。また、図８のステップＳ２０１～Ｓ２０２、ステップＳ２０４～Ｓ２０５、及び／又は、ステップＳ２０６の処理は省略されてもよい。ステップＳ２０４～Ｓ２０６の処理が省略される場合、ステップＳ２０３の処理も省略される。

【００９７】

以下、第１指示構文に、スキャン画像から文字を認識する目的を含ませることの技術的意義について説明する。

【００９８】

図１０は、スキャン画像から文字を認識する目的を含まない第１指示構文が大規模言語モデルに入力されたときに大規模言語モデルから第１出力情報１０００を示す模式図である。

【００９９】

図１０に示す第１出力情報１０００は、図６（Ｂ）に示した第１指示構文６１０のうち、目的指定文６０３を含まない第１指示構文が大規模言語モデルに入力されたときに大規模言語モデルから出力された情報である。第１出力情報１０００には、店舗情報１００１、概要情報１００２、購入情報１００３、付属情報１００４等が含まれている。店舗情報１００１には、スキャン画像５００の店舗情報５０１の全て（店舗の名称、住所、電話番号、営業時間）が含まれている。概要情報１００２には、スキャン画像５００の概要情報５０２のうち、レシートを発行した日時が含まれていないが、レシートの識別番号及び購入者による支払方法が含まれている。購入情報１００３には、スキャン画像５００の購入情報５０３の全て（購入した商品毎の商品名、単価、数量及び小計と、その合計金額と、消費税と、税込金額と、預かり金額と、お釣り）が含まれている。このように、目的を指定した場合の第１出力情報７００と比較して、目的を指定していない場合の第１出力情報１０００には、経費申請に不要な多数の項目が含まれている。そのため、利用者は、多数の不要項目を指定して、再度、大規模言語モデルから第１出力情報を取得する必要、又は、第１出力情報から多数の不要項目を削除するように第１出力情報を修正する必要がある。

【０１００】

画像処理システム１は、スキャン画像から文字を認識する目的を含むように第１指示構文を生成し、大規模言語モデルに入力する。これにより、画像処理システム１は、大規模言語モデルから出力される第１出力情報に、目的と合致しない項目が含まれることを抑制でき、その結果、利用者の手間を省くとともに、ファイル生成処理にかかる処理負荷及び処理時間を低減させることができる。

【０１０１】

以上詳述したように、画像処理システム１は、媒体が撮像されたスキャン画像から認識された文字に加えて、スキャン画像から文字を認識する目的を含む第１指示構文を大規模

言語モデルに入力し、抽出すべき項目を取得する。これにより、画像処理システム１は、スキャン画像から認識した文字から目的と合致する項目を高精度に抽出し、利用者に提供することができる。したがって、画像処理システム１は、スキャン画像から認識した文字を、利用者の作業負担を軽減させつつ、効率良く管理することが可能となった。

【０１０２】

また、画像処理システム１は、大規模言語モデルを利用することにより、抽出対象項目を事前に設定する必要がなくなる。そのため、画像処理システム１は、媒体内で抽出対象項目が配置されている位置を設定する必要がなくなり、利用者の利便性を向上させることができる。また、画像処理システム１は、抽出対象項目が配置される位置が定まっていない媒体に対しても、抽出対象項目を適切に抽出することができる。一般に、スキャン画像から文字を認識する目的は、業務内容毎に異なり、抽出対象項目は、スキャン画像から文字を認識する目的毎に異なる。したがって、データ入力業務では、業務内容毎に、又は、スキャン画像から文字を認識する目的毎に、抽出対象項目を事前に設定しておく必要がある。画像処理システム１は、スキャン画像から文字を認識する目的を大規模言語モデルに入力することにより、抽出対象項目を、事前に設定しておくことなく、精度良く抽出することができる。

【０１０３】

図１１は、他の実施形態に係るファイル生成処理の動作の一部の例を示すシーケンスである。

【０１０４】

図１１に示すシーケンスは、図４に示したシーケンスのステップＳ１０２～Ｓ１０５の代わりに実行される。図１１に示すシーケンスでは、画像処理システム１は、スキャン画像から文字を認識する目的と、スキャン画像に含まれる媒体の種類とを、大規模言語モデルを利用して取得する。

【０１０５】

まず、クライアント装置２００の第２処理回路２２０は、第２通信装置２０１を介してサーバ装置３００に、ステップＳ１０１で取得したスキャン画像を送信する（ステップＳ３０１）。

【０１０６】

一方、サーバ装置３００の第１取得部３２１は、第３通信装置３０１を介してクライアント装置２００からスキャン画像を受信し、取得する（ステップＳ３０２）。

【０１０７】

次に、第１取得部３２１は、スキャン画像に対してＯＣＲ処理を実行して、スキャン画像から文字を認識し、スキャン画像から認識した文字を取得する（ステップＳ３０３）。

【０１０８】

次に、生成部３２２は、大規模言語モデルに対して、スキャン画像に含まれる媒体の種類の特定を要求する第３指示構文を生成する（ステップＳ３０４）。生成部３２２は、第１指示構文と同様にして、第１取得部３２１が取得した文字を含む第３指示構文を生成する。但し、第３指示構文における指令文には、指定したＯＣＲ結果から、媒体の種類を特定し、出力する指令が記述される。また、第３指示構文には、種類指定文及び目的指定文が含まれない。第１指示構文が生成される場合と同様に、サーバ装置３００は、第３指示構文のひな型を予め第３記憶装置３１０に記憶しておき、生成部３２２は、第３記憶装置３１０に記憶されたひな型に従って、第３指示構文を生成してもよい。

【０１０９】

次に、第２取得部３２３は、生成部３２２が生成した第３指示構文と、その第３指示構文を大規模言語モデルに入力することを要求する要求情報とを、第３通信装置３０１を介してＬＬＭサーバ装置４００に送信する（ステップＳ３０５）。

【０１１０】

一方、ＬＬＭサーバ装置４００は、サーバ装置３００から第３指示構文及び要求情報を受信する。ＬＬＭサーバ装置４００は、受信した第３指示構文を、ＬＬＭサーバ装置４０

０に記憶された大規模言語モデルに入力し、大規模言語モデルから出力された第３出力情報を取得する（ステップＳ３０６）。この第３出力情報には、大規模言語モデルによって特定された媒体の種類が含まれる。

【０１１１】

次に、ＬＬＭサーバ装置４００は、取得した第３出力情報をサーバ装置３００に送信する（ステップＳ３０７）。

【０１１２】

一方、サーバ装置３００の第２取得部３２３は、第３通信装置３０１を介してＬＬＭサーバ装置４００から第３出力情報を受信することにより、媒体の種類を取得する（ステップＳ３０８）。これにより、第２取得部３２３は、生成部３２２が生成した第３指示構文を大規模言語モデルに入力し、大規模言語モデルから出力された第３出力情報を取得する。

【０１１３】

次に、生成部３２２は、大規模言語モデルに対して、スキャン画像から文字を認識する目的について一又は複数の候補の特定を要求する第４指示構文を生成する（ステップＳ３０９）。生成部３２２は、第１指示構文と同様にして、第１取得部３２１が取得した文字を含む第４指示構文を生成する。但し、第４指示構文における指令文には、指定した種類の媒体に対するＯＣＲ結果から、スキャン画像から文字を認識する目的の候補を特定し、出力する指令が記述される。また、第４指示構文には、第２取得部３２３で取得した媒体の種類を指定する種類指定文が含まれる。一方、第４指示構文には、目的指定文が含まれない。第１指示構文が生成される場合と同様に、サーバ装置３００は、第４指示構文のひな型を予め第３記憶装置３１０に記憶しておき、生成部３２２は、第３記憶装置３１０に記憶されたひな型に従って、第４指示構文を生成してもよい。

【０１１４】

次に、第２取得部３２３は、生成部３２２が生成した第４指示構文と、その第４指示構文を大規模言語モデルに入力することを要求する要求情報とを、第３通信装置３０１を介してＬＬＭサーバ装置４００に送信する（ステップＳ３１０）。

【０１１５】

一方、ＬＬＭサーバ装置４００は、サーバ装置３００から第４指示構文及び要求情報を受信する。ＬＬＭサーバ装置４００は、受信した第４指示構文を、ＬＬＭサーバ装置４００に記憶された大規模言語モデルに入力し、大規模言語モデルから出力された第４出力情報を取得する（ステップＳ３１１）。この第４出力情報には、大規模言語モデルによって特定された、スキャン画像から文字を認識する目的についての一又は複数の候補が含まれる。

【０１１６】

次に、ＬＬＭサーバ装置４００は、取得した第４出力情報をサーバ装置３００に送信する（ステップＳ３１２）。

【０１１７】

一方、サーバ装置３００の第２取得部３２３は、第３通信装置３０１を介してＬＬＭサーバ装置４００から第４出力情報を受信し、受信した第４出力情報から、スキャン画像から文字を認識する目的の候補を抽出する（ステップＳ３１３）。これにより、第２取得部３２３は、生成部３２２が生成した第４指示構文を大規模言語モデルに入力し、大規模言語モデルから出力された第４出力情報を取得する。

【０１１８】

次に、出力制御部３２４は、第２取得部３２３が抽出した、スキャン画像から文字を認識する目的の候補を示す情報を、第３通信装置３０１を介してクライアント装置２００に送信する（ステップＳ３１４）。

【０１１９】

一方、クライアント装置２００の第２処理回路２２０は、第２通信装置２０１を介してサーバ装置３００から、スキャン画像から文字を認識する目的の候補を示す情報を受信す

る（ステップＳ３１５）。
【０１２０】
次に、第２処理回路２２０は、スキャン画像から文字を認識する目的の候補を、選択可能に第２表示装置２０３に表示する（ステップＳ３１６）。
【０１２１】
次に、第２処理回路２２０は、利用者による、選択可能に表示された目的の候補の中から、スキャン画像から文字を認識する目的の選択を受け付ける（ステップＳ３１７）。第２処理回路２２０は、利用者により第２入力装置２０２を用いて選択された目的の選択を受け付ける。
【０１２２】
次に、第２処理回路２２０は、第２通信装置２０１を介してサーバ装置３００に、利用者により選択された目的を示す情報を送信する（ステップＳ３１８）。
【０１２３】
一方、サーバ装置３００の第１取得部３２１は、第３通信装置３０１を介してクライアント装置２００から、利用者により選択された目的を示す情報を受信する（ステップＳ３１９）。これにより、第１取得部３２１は、利用者から受け付けた、スキャン画像から文字を認識する目的を取得する。以降、図４のステップＳ１０６～Ｓ１１８の処理が実行される。ステップＳ１０６において、生成部３２２は、ひな型テーブルから、ステップＳ３０８で取得された媒体の種類に対応するひな型を読み出す。ステップＳ１０７において、生成部３２２は、ステップＳ３０３で取得された文字、ステップＳ３１９で取得されたスキャン画像から文字を認識する目的、及び、ステップＳ３０３で取得された媒体の種類を含む第１指示構文を生成する。
【０１２４】
なお、同一目的及び同一種類の媒体についてファイル生成処理が実行される場合、二回目以降のファイル生成処理においては、ステップＳ３０４～Ｓ３１９の処理が省略され、初回に取得又は修正された目的及び媒体の種類を用いて指示構文が生成されてもよい。
【０１２５】
以上詳述したように、画像処理システム１は、文字を認識する目的及び媒体の種類を、大規模言語モデルを利用して取得する場合も、スキャン画像から認識した文字を、利用者の作業負担を軽減させつつ、効率良く管理することが可能となった。
【０１２６】
特に、画像処理システム１は、利用者がデータ入力業務に不慣れな場合でも、文字を認識する目的及び媒体の種類を適切に設定することができ、その結果、スキャン画像から認識した文字から、必要な項目を精度良く抽出することが可能となった。
【０１２７】
図１２は、他の実施形態に係るサーバ装置における第３処理回路５２０の概略構成を示すブロック図である。
【０１２８】
第３処理回路５２０は、第３処理回路３２０の代わりに用いられ、ファイル生成処理等を実行する。第３処理回路３２０は、第１取得回路５２１、生成回路５２２、第２取得回路５２３、出力制御回路５２４及び更新回路５２５等を有する。
【０１２９】
第１取得回路５２１は、第１取得部の一例であり、第１取得部３２１と同様の機能を有する。第１取得回路５２１は、第３通信装置３０１からスキャン画像、並びに、スキャン画像から文字を認識する目的、及び、スキャン画像に含まれる媒体の種類を示す情報を受信し、第３記憶装置３１０に記憶する。また、第１取得回路５２１は、スキャン画像から文字を認識し、第３記憶装置３１０に記憶する。また、第１取得回路５２１は、第３通信装置３０１から、利用者から受け付けた第１出力情報に対する確認結果を受信し、第３記憶装置３１０に記憶する。
【０１３０】

生成回路５２２は、生成部の一例であり、生成部３２２と同様の機能を有する。生成回路５２２は、第３記憶装置３１０から、スキャン画像から認識された文字、スキャン画像から文字を認識する目的、スキャン画像に含まれる媒体の種類、第１出力情報、及び／又は、第１出力情報に対する確認結果を読み出す。生成回路５２２は、読み出した各情報に基づいて、第１指示構文及び／又は第２指示構文を生成し、第３記憶装置３１０に記憶する。

【０１３１】

第２取得回路５２３は、第２取得部の一例であり、第２取得部３２３と同様の機能を有する。第２取得回路５２３は、第３記憶装置３１０から第１指示構文及び／又は第２指示構文を読み出し、第３通信装置３０１を介してＬＬＭサーバ装置４００に送信する。第２取得回路５２３は、第３通信装置３０１を介してＬＬＭサーバ装置４００から第１出力情報及び／又は第２出力情報等を受信し、第３記憶装置３１０に記憶する。

【０１３２】

出力制御回路５２４は、出力制御部の一例であり、出力制御部３２４と同様の機能を有する。出力制御回路５２４は、第３記憶装置３１０から第１出力情報及び／又は第２出力情報等を読み出し、第３通信装置３０１に出力する。

【０１３３】

更新回路５２５は、更新部の一例であり、更新部３２５と同様の機能を有する。更新回路５２５は、第３記憶装置３１０から、利用者から受け付けた第１出力情報に対する確認結果、及び、第１指示構文のひな型を読み出し、読み出した確認結果に基づいてひな型を更新し、第３記憶装置３１０に記憶する。

【０１３４】

以上詳述したように、サーバ装置は、第３処理回路５２０を用いる場合も、スキャン画像から認識した文字を、利用者の作業負担を軽減させつつ、効率良く管理することが可能となった。

【０１３５】

以上、好適な実施形態について説明してきたが、実施形態はこれらに限定されない。例えば、大規模言語モデルは、ＬＬＭサーバ装置４００に記憶されるのでなく、サーバ装置３００の第３記憶装置３１０に予め記憶されてもよい。その場合、図４のステップＳ１０８～Ｓ１１１の処理は省略され、第２取得部３２３は、第１指示構文を、第３記憶装置３１０に記憶された大規模言語モデルに入力し、大規模言語モデルから出力された第１出力情報を取得する。

【０１３６】

また、ステップＳ１０５～Ｓ１０８、Ｓ１１１、Ｓ１１８の処理は、サーバ装置３００でなく、クライアント装置２００により実行されてもよい。その場合、クライアント装置２００の第２記憶装置２１０には、第３記憶装置３１０に記憶される各プログラム、データベース、テーブル、情報等が記憶される。第２処理回路２２０は、第２記憶装置２１０に記憶された各プログラムを読み取り、読み取った各プログラムに従って動作することにより、第１取得部、生成部、第２取得部、出力制御部及び更新部として機能する。第２処理回路２２０の第１取得部、生成部、第２取得部、出力制御部及び更新部は、第３処理回路３２０の第１取得部３２１、生成部３２２、第２取得部３２３、出力制御部３２４及び更新部３２５と同様の機能を有する。

【０１３７】

ステップＳ１０３～Ｓ１０４の処理は省略され、クライアント装置２００の第１取得部は、ステップＳ１０２において、利用者から受け付けた、スキャン画像から文字を認識する目的、及び、スキャン画像に含まれる媒体の種類を取得する。ステップＳ１０８において、クライアント装置２００の第２取得部は、各指示構文及び要求情報を、第２通信装置２０１を介してＬＬＭサーバ装置４００に送信する。ステップＳ１１１において、クライアント装置２００の第２取得部は、第２通信装置２０１を介してＬＬＭサーバ装置４００から各出力情報を受信する。ステップＳ１１２～Ｓ１１３の処理は省略される。ステップ

Ｓ１１４において、クライアント装置２００の出力制御部は、各出力情報を第２表示装置２０３に表示することにより出力する。ステップＳ１１６～Ｓ１１７の処理は省略され、図６のステップＳ２０７において、クライアント装置２００の生成部は、生成したファイルを第２表示装置２０３に表示することにより出力する。

【０１３８】

同様に、図１１に示すファイル生成処理においても、サーバ装置３００により実行される各処理は、クライアント装置２００により実行されてもよい。

【０１３９】

これらの場合、大規模言語モデルは、ＬＬＭサーバ装置４００に記憶されるのでなく、クライアント装置２００の第２記憶装置２１０に予め記憶されてもよい。その場合、クライアント装置２００の第２取得部は、各指示構文を、第２記憶装置２１０に記憶された大規模言語モデルに入力し、大規模言語モデルから出力された各出力情報を取得する。

【０１４０】

また、ステップＳ１０１～Ｓ１０２、Ｓ１０５～Ｓ１０８、Ｓ１１１、Ｓ１１４～Ｓ１１５、Ｓ１１８の処理は、サーバ装置３００及びクライアント装置２００でなく、画像読取装置１００により実行されてもよい。その場合、画像読取装置１００の第１記憶装置１１０には、第３記憶装置３１０に記憶される各プログラム、データベース、テーブル、情報等が記憶される。第１処理回路１２０は、第１記憶装置１１０に記憶された各プログラムを読み取り、読み取った各プログラムに従って動作することにより、第１取得部、生成部、第２取得部、出力制御部及び更新部として機能する。第２処理回路２２０の第１取得部、生成部、第２取得部、出力制御部及び更新部は、第３処理回路３２０の第１取得部３２１、生成部３２２、第２取得部３２３、出力制御部３２４及び更新部３２５と同様の機能を有する。

【０１４１】

ステップＳ１０１において、画像読取装置１００の第１取得部は、撮像装置１０４が生成したスキャン画像を取得する。ステップＳ１０２において、画像読取装置１００の第１取得部は、第１入力装置１０２を用いて利用者から受け付けた、スキャン画像から文字を認識する目的、及び、スキャン画像に含まれる媒体の種類を取得する。ステップＳ１０３～Ｓ１０４の処理は省略される。ステップＳ１０８において、画像読取装置１００の第２取得部は、各指示構文及び要求情報を、第１通信装置１０１を介してＬＬＭサーバ装置４００に送信する。ステップＳ１１１において、画像読取装置１００の第２取得部は、第１通信装置１０１を介してＬＬＭサーバ装置４００から各出力情報を受信する。ステップＳ１１２～Ｓ１１３の処理は省略される。ステップＳ１１４において、画像読取装置１００の出力制御部は、各出力情報を第１表示装置１０３に表示することにより出力する。ステップＳ１１６～Ｓ１１７の処理は省略され、図６のステップＳ２０７において、画像読取装置１００の生成部は、生成したファイルを第１表示装置１０３に表示することにより出力する。

【０１４２】

同様に、図１１に示すファイル生成処理においても、クライアント装置２００及びサーバ装置３００により実行される各処理は、画像読取装置１００により実行されてもよい。

【０１４３】

これらの場合、大規模言語モデルは、ＬＬＭサーバ装置４００に記憶されるのでなく、画像読取装置１００の第１記憶装置１１０に予め記憶されてもよい。その場合、画像読取装置１００の第２取得部は、各指示構文を、第１記憶装置１１０に記憶された大規模言語モデルに入力し、大規模言語モデルから出力された各出力情報を取得する。

【０１４４】

また、図４又は図１１に示すファイル生成処理の各処理を実行する装置の組合せは、任意に設定可能である。即ち、第１取得部、生成部、第２取得部、出力制御部及び更新部のそれぞれを、サーバ装置３００、クライアント装置２００及び画像読取装置１００の何れに配置するかは任意に設定可能である。

【符号の説明】

【０１４５】

１　画像処理システム、１００　画像読取装置、２００　クライアント装置、３００　サーバ装置、３１０　第３記憶装置、３２１　第１取得部、３２２　生成部、３２３　第２取得部、３２４　出力制御部、３２５　更新部

【図面】

【図１】

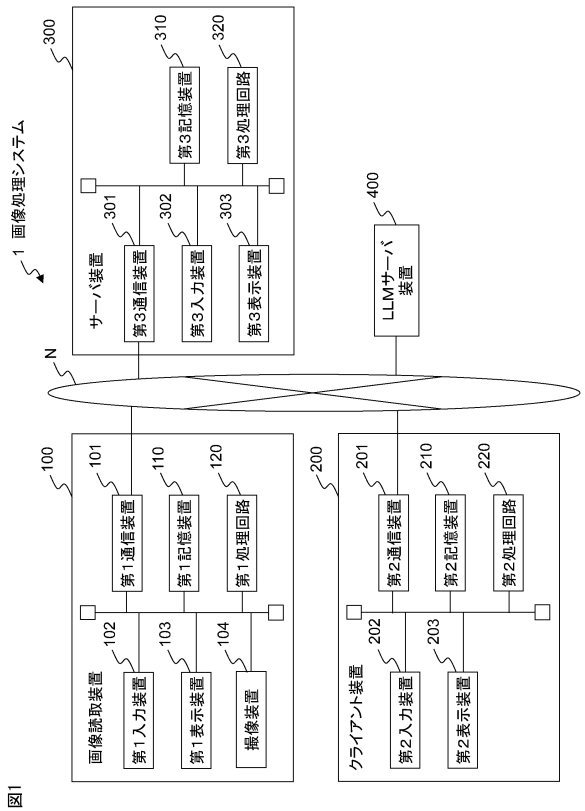
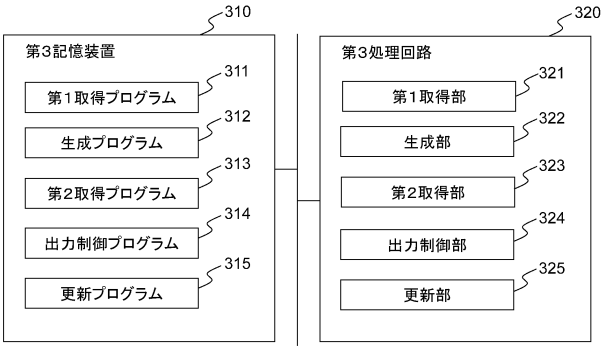


図1

【図２】

図2



【図３】

図3

種類	ひな型
種類A	ひな型A
種類B	ひな型B
種類C	ひな型C
…	…

【図４】

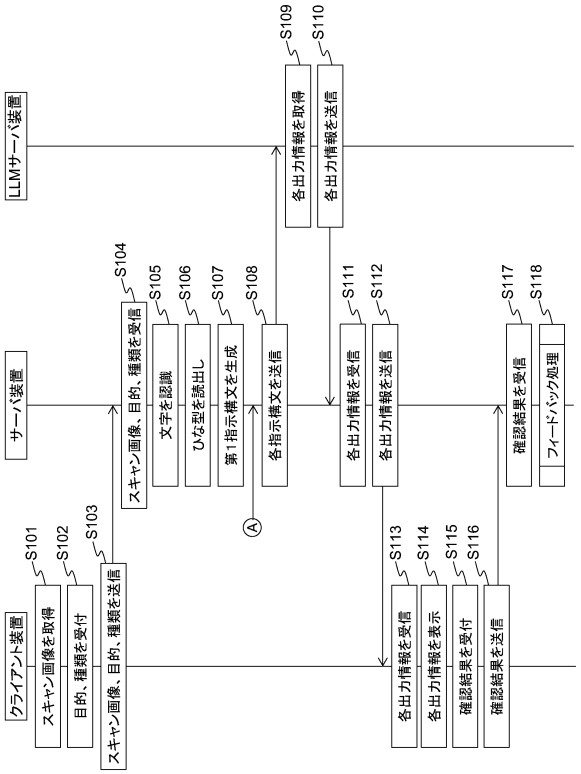
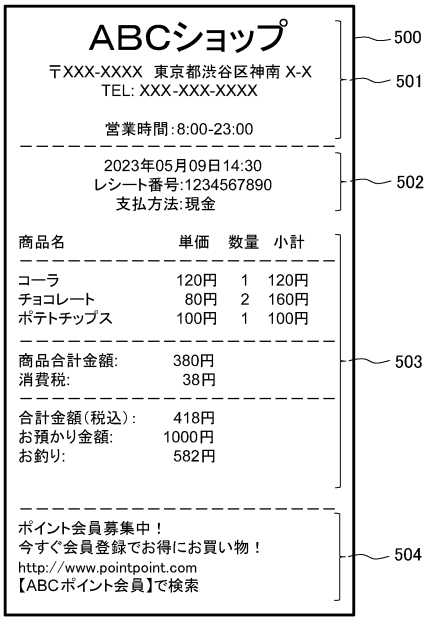


図4

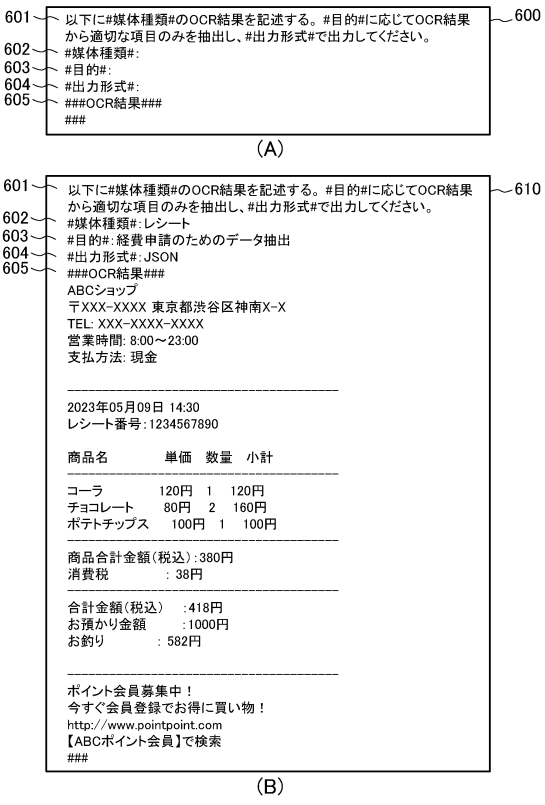
【図５】

図5



【図６】

図6



【図７】

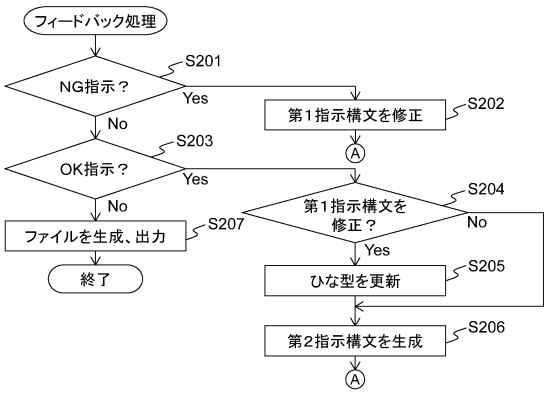
図7


```
{
"店名": "ABCショップ",
"住所": "〒XXX-XXX 東京都渋谷区神南X-X",
"レシート番号": "1234567890",
"商品一覧": [
{
"商品名": "コーラ",
"単価": "120円",
"数量": "1",
"小計": "120円"
},
{
"商品名": "チョコレート",
"単価": "80円",
"数量": "2",
"小計": "160円"
},
{
"商品名": "ポテトチップス",
"単価": "100円",
"数量": "1",
"小計": "100円"
}
],
"商品合計金額": "380円",
"消費税": "38円",
"合計金額(税込)": "418円"
}
```


701 702 703 700

【図８】

図8



【図９】

図9

901 以下のJSON形式のテキストとファイル名例から、最適なファイル名を出力してください。
902 # ファイル名例
XYZレストラン_324234323_領収書.jpg
903 #JSON形式のテキスト


```
{
"店名": "ABCショップ",
"住所": "〒XXX-XXX 東京都渋谷区神南X-X",
"レシート番号": "1234567890",
"商品一覧": [
{
"商品名": "コーラ",
"単価": "120円",
"数量": "1",
"小計": "120円"
},
{
"商品名": "チョコレート",
"単価": "80円",
"数量": "2",
"小計": "160円"
},
{
"商品名": "ポテトチップス",
"単価": "100円",
"数量": "1",
"小計": "100円"
}
],
"商品合計金額": "380円",
"消費税": "38円",
"合計金額(税込)": "418円"
}
```


900

(A)

910 ABCショップ_1234567890_領収書.jpg
「ABCコンビニ」を店名とし、レシート番号「1234567890」を含めたファイル名を提案します。これにより、ファイル名がより具体的で識別性が高くなります。

(B)

【図１０】

図10


```
{
"店名": "ABCコンビニ",
"住所": "〒123-4567 東京都渋谷区神南1-2-3",
"電話番号": "012-3456-7890",
"営業時間": "8:00～23:00",
"支払方法": "現金",
"レシート情報": {
"レシート番号": "1234567890",
"商品リスト": [
{
"商品名": "コーラ",
"単価": "120円",
"数量": 1,
"小計": "120円"
},
{
"商品名": "チョコレート",
"単価": "80円",
"数量": 2,
"小計": "160円"
},
{
"商品名": "ポテトチップス",
"単価": "100円",
"数量": 1,
"小計": "100円"
}
],
"商品合計金額(税込)": "380円",
"消費税": "38円",
"合計金額(税込)": "418円",
"お預かり金額": "1000円",
"お釣り": "582円"
},
"その他": "ポイント会員募集中！¥n今すぐ会員登録でお得に買い物！¥nhttp://www.pointpoint.com¥n【ABCポイント会員】で検索"
}
```


1000 1001 1002 1003 1004

【図１１】

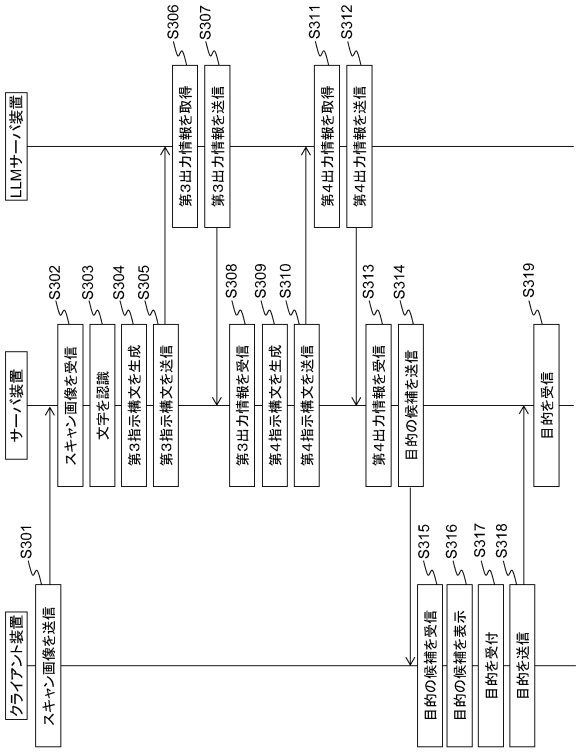


図11

【図１２】

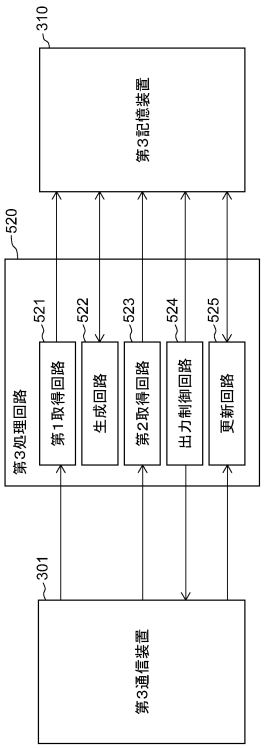


図12

フロントページの続き

Ｆターム（参考） 5E555 AA13 BA10 BA82 BA83 BB10 BC19 BD01 DB49 EA16 EA24
EA27 FA00
5L030 BB72
5L049 BB72