



(10) **DE 10 2011 005 579 A1** 2011.11.10

(12) **Offenlegungsschrift**

(21) Aktenzeichen: **10 2011 005 579.7**

(22) Anmeldetag: **15.03.2011**

(43) Offenlegungstag: **10.11.2011**

(51) Int Cl.: **G06K 9/52 (2011.01)**

(30) Unionspriorität:
2010-059160 16.03.2010 JP

(71) Anmelder:
CANON KABUSHIKI KAISHA, Tokio/Tokyo, JP

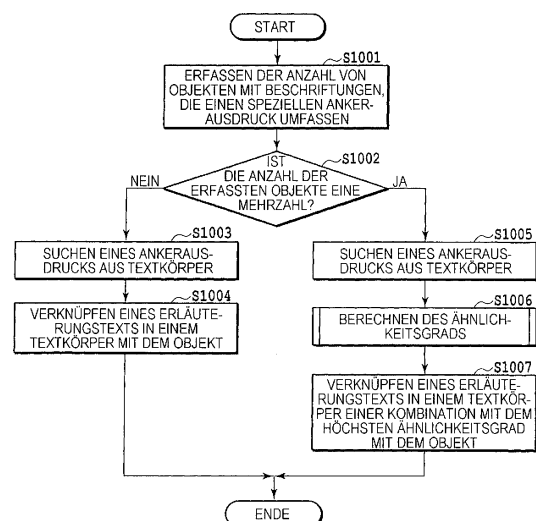
(74) Vertreter:
TBK, 80336, München, DE

(72) Erfinder:
**Sohma, Hidetomo, Ohta-ku, Tokyo, JP; Kanatsu,
Tomotoshi, Ohta-ku, Tokyo, JP; Kosaka, Ryo,
Ohta-ku, Tokyo, JP; Misawa, Reiji, Ohta-ku,
Tokyo, JP**

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

(54) Bezeichnung: **Bildverarbeitungsvorrichtung, Bildverarbeitungsverfahren und Speichermedium**

(57) Zusammenfassung: Selbst wenn Beschriftungen einer Vielzahl von Objekten einen identischen Anker Ausdruck verwenden, kann die vorliegende Erfindung einen geeigneten Erläuterungstext in einem Textkörper als Metadaten mit den Objekten verknüpfen.



Beschreibung**HINTERGRUND DER ERFINDUNG****Gebiet der Erfindung**

[0001] Die vorliegende Erfindung bezieht sich auf eine Bildverarbeitungsvorrichtung, ein Bildverarbeitungsverfahren und ein Speichermedium, durch die elektronische Dokumentdaten zum Suchen eines Objekts in einem Dokument erzeugt werden.

Beschreibung des Standes der Technik

[0002] Herkömmlich wurde ein Ansatz erwogen, um eine Suche eines Objekts zur Verfügung zu stellen, das in einem Dokument umfasst ist, wie etwa eine Fotografie, eine Zeichnung (Strich- bzw. Linienzeichnung) oder eine Tabelle. (Der Begriff „Objekt“ wird hierin so verwendet, dass er bedeutet, dass ein anderes Objekt als Zeichen, wie etwa eine Fotografie, eine Zeichnung (Strich- bzw. Linienzeichnung) oder eine Tabelle, umfasst ist.)

[0003] Zum Beispiel gibt es ein Verfahren, durch das in der Nähe eines aus einem Dokument extrahierten Objekts eine Zeichenkette (Beschriftung) eingefügt wird, die das Objekt beschreibt, und diese als Metadaten verknüpft wird, so dass das Objekt gesucht werden kann.

[0004] Wenn eine Beschriftung in einem allgemeinen bzw. normalen Dokument einen (hierin nachstehend als „Ankerausdruck“ bezeichneten) Ausdruck zum Identifizieren des Objekts umfasst, wie etwa eine Zeichnungsnummer (z. B. „Fotografie 1“, „erste Zeichnung“ oder „Tabelle 1“), wird eine genauere Beschreibung des Objekts auch in einem Textkörper unter Verwendung des Ankerausdrucks beschrieben. Der Ankerausdruck, wie er vorstehend beschrieben ist, wurde auch als ein Mittel zum Identifizieren eines Objekts in einem Dokument verwendet. Gemäß der Erfindung, die in der japanischen Patentschrift Nr. H11-025113 (1999) offenbart ist, wird ein (hierin nachstehend als „Erläuterungstext in Textkörper“ bezeichneten) erläuternder Teil in einem Textkörper, der einen Ankerausdruck umfasst, extrahiert und als Metadaten für das Objekt verknüpft. Wenn zum Beispiel eine an ein Objekt einer Zeichnung angrenzende Beschriftung einen Ankerausdruck „Fig. 1“ umfasst und ein Textkörper die Erläuterung „die Fig. 1 ist AAA“ umfasst, ist ein Ankerausdruck „Fig. 1“ als Identifikationsinformation des Objekts der Zeichnung verknüpft. Gleichzeitig ist auch der Erläuterungstext in dem Textkörper, nämlich „die Fig. 1 ist AAA“ als Metadaten verknüpft, wodurch die Suche durch die Metadaten des Objekts der Zeichnung zur Verfügung gestellt wird.

[0005] In den letzten Jahren haben einige Textverarbeitungssysteme zum Beispiel eine Editierfunktion, wie etwa eine Funktion zum automatischen Erzeugen eines Ankerausdrucks und eine Funktion zum Verknüpfen eines in einem Dokument vorhandenen Objekts mit einem Erläuterungstext in einem Textkörper. Die durch diese Funktionen gegebenen Informationen (Metadaten) können in einem elektronischen Dokument gespeichert werden, um dadurch ein effizientes Editieren des Dokuments zu realisieren.

[0006] Scanner bzw. Abtastvorrichtungen haben in den letzten Jahren eine Funktion, wie etwa einen automatischen Dokumenteneinzug, und können somit viele Seiten von Papieren bzw. Unterlagen auf einfache Weise lesen. Somit können solche Scanner auch gleichzeitig eine Vielzahl von Dokumenttypen lesen. Wenn ein solcher Scanner andererseits eine Mischung unterschiedlicher Dokumente lesen muss, besteht die Möglichkeit, dass eine Vielzahl von Objekten mit Beschriftungen hervorgebracht werden können, die einen identischen Ankerausdruck umfassen. Zum Beispiel kann es einen Fall geben, bei dem eines der gleichzeitig gelesenen Dokumente ein Tabellenobjekt mit einer Beschriftung „die Tabelle 1 ist YYY“ aufweist und ein anderes Dokument unter diesen ein Tabellenobjekt mit einer Beschriftung „die Tabelle 1 zeigt ZZZ“ aufweist. Falls die vorstehend beschriebene Verknüpfungsverarbeitung unter solchen Umständen einfach durchgeführt wird, wird der Ankerausdruck für die gleiche „Tabelle 1“ mit den beiden Tabellenobjekten verknüpft, wodurch es misslingt, einen Erläuterungstext in einem Textkörper bereitzustellen, der dem Ankerausdruck für „die Tabelle 1“ auf geeignete Weise entspricht.

[0007] Infolge der vorgenannten Situation wurde ein derartiges Verfahren verlangt, durch das, selbst wenn eine Vielzahl von Dokumenttypen gelesen werden müssen und eine Vielzahl von Beschriftungen einen identischen Ankerausdruck verwenden, eine Beschriftung oder ein Erläuterungstext in einem Textkörper auf geeignete Weise als Metadaten mit dem Objekt verknüpft werden kann.

KURZFASSUNG DER ERFINDUNG

[0008] Eine Bildverarbeitungsvorrichtung gemäß der vorliegenden Erfindung umfasst: eine Bereichsaufteilungseinheit, die zum jeweiligen Aufteilen eines Bildes einer Vielzahl von Seiten in eine Vielzahl von Bereichen konfiguriert ist; eine Attributinformationen-Hinzufügungseinheit, die zum Hinzufügen von den Bereichen entsprechenden Attributen zu der jeweiligen Vielzahl von aufgeteilten Bereichen konfiguriert ist; eine Zeichenerkennungseinheit, die zum Durchführen einer Zeichenerkennungsverarbeitung auf einem Beschriftungsbereich und einem Textkörperbereich konfiguriert ist, zu denen ein Beschriftungsattribut beziehungsweise ein Textkörperattribut durch

die Attributinformationen-Hinzufügungseinheit hinzugefügt sind; und eine Metadaten-Verarbeitungseinheit, die zum Verknüpfen von Metadaten mit einem Objektbereich konfiguriert ist, der durch den Beschriftungsbereich begleitet ist, wobei die Metadaten-Verarbeitungseinheit aufweist: eine erste Extraktionseinheit, die zum Extrahieren eines Ankerausdrucks, der aus einer vorbestimmten Zeichenkette besteht, und eines Beschriftungsausdrucks, der aus einer von dem Ankerausdruck abweichenden Zeichenkette besteht, aus dem Ergebnis der Zeichenerkennungsverarbeitung auf dem Beschriftungsbereich konfiguriert ist; eine Bestimmungseinheit, die zum Bestimmen konfiguriert ist, ob es eine Vielzahl von Objektbereichen gibt, die durch Beschriftungsbereiche begleitet sind, die einen identischen Ankerausdruck umfassen; eine zweite Extraktionseinheit, die zum Extrahieren eines Erläuterungstexts, der den Ankerausdruck umfasst, aus dem Ergebnis der Zeichenerkennungsverarbeitung auf dem Textkörperbereich konfiguriert ist; eine erste Verknüpfungseinheit, die zum Verknüpfen des Objektbereichs mit Metadaten, die aus dem durch die zweite Extraktionseinheit extrahierten Erläuterungstext erhalten werden, konfiguriert ist, falls die Bestimmungseinheit bestimmt, dass es einen Objektbereich gibt, der durch einen Beschriftungsbereich begleitet ist, der den identischen Ankerausdruck umfasst; eine Ähnlichkeitsgrad-Berechnungseinheit, die zum jeweiligen Berechnen von Ähnlichkeitsgraden zwischen Beschriftungsausdrücken der jeweiligen Beschriftungsbereiche, die den identischen Ankerausdruck umfassen, und dem durch die zweite Extraktionseinheit extrahierten Erläuterungstext, der den identischen Ankerausdruck umfasst, konfiguriert ist, falls die Bestimmungseinheit bestimmt, dass es eine Vielzahl von Objektbereichen gibt, die durch Beschriftungsbereiche begleitet sind, die den identischen Ankerausdruck umfassen; und eine zweite Verknüpfungseinheit, die zum Bestimmen, basierend auf den durch die Ähnlichkeitsgrad-Berechnungseinheit berechneten Ähnlichkeitsgraden, derartiger Erläuterungstexte konfiguriert ist, die für die jeweilige Vielzahl von Objektbereichen optimal sind, um aus den bestimmten optimalen Erläuterungstexten erhaltene Metadaten mit den jeweiligen Objektbereichen zu verknüpfen.

[0009] Gemäß der vorliegenden Erfindung kann das Objekt selbst dann mit geeigneten Metadaten verknüpft werden, wenn es eine Vielzahl von Objekten gibt, die einen identischen Ankerausdruck umfassen. Somit kann selbst aus Bilddaten heraus, die eine Mischung einer Vielzahl von Dokumenten umfassen, ein derartiges elektronisches Dokument erstellt werden, das geeignete Metadaten zu den einzelnen Objekten hinzugefügt hat.

[0010] Weitere Merkmale der vorliegenden Erfindung werden aus der folgenden Beschreibung bei-

spielhafter Ausführungsbeispiele (unter Bezugnahme auf die begleitenden Zeichnungen) ersichtlich.

KURZE BESCHREIBUNG DER ZEICHNUNGEN

[0011] [Fig. 1](#) ist ein Blockschaltbild, das die Konfiguration eines Bildverarbeitungssystems gemäß der vorliegenden Erfindung veranschaulicht;

[0012] [Fig. 2](#) veranschaulicht die Konfiguration von einem MFP **100**;

[0013] [Fig. 3](#) ist ein Blockschaltbild, das die interne Konfiguration einer Datenverarbeitungseinheit **218** veranschaulicht;

[0014] [Fig. 4](#) ist ein Blockschaltbild, das die interne Konfiguration einer Metadaten-Verarbeitungseinheit **304** veranschaulicht;

[0015] [Fig. 5A](#) bis [Fig. 5C](#) veranschaulichen die Verarbeitungsdetails durch eine Datenverarbeitungseinheit **218**. [Fig. 5A](#) veranschaulicht, wie Bilddaten in Bereiche aufgeteilt werden. [Fig. 5B](#) ist eine Tabelle, die ein Beispiel des Verarbeitungsergebnisses durch eine Bereichsaufteilungseinheit **301**, eine Attributinformationen-Hinzufügungseinheit **302** und eine Zeichenerkennungseinheit **303** veranschaulicht. [Fig. 5C](#) veranschaulicht ein Beispiel der durch eine Formatumwandlungseinheit **305** verwendeten Entsprechungstabelle;

[0016] [Fig. 6A](#) und [Fig. 6B](#) veranschaulichen ein gemäß der vorliegenden Erfindung erzeugtes elektronisches Dokument. [Fig. 6A](#) veranschaulicht ein Beispiel eines elektronischen Dokuments, das durch ein SVG-Format beschrieben ist. [Fig. 6B](#) veranschaulicht Metadaten in dem elektronischen Dokument durch eine Tabelle;

[0017] [Fig. 7](#) ist ein Ablaufdiagramm, das den Überblick der Verarbeitung durch die Bildverarbeitungsvorrichtung gemäß der vorliegenden Erfindung veranschaulicht;

[0018] [Fig. 8A](#) und [Fig. 8B](#) sind Ablaufdiagramme, die den Ablauf von Metadatenextraktions- und Hinzufügungsverarbeitungen veranschaulichen;

[0019] [Fig. 9A](#) und [Fig. 9B](#) veranschaulichen die Verarbeitung durch die Bildverarbeitungsvorrichtung gemäß der vorliegenden Erfindung. [Fig. 9A](#) veranschaulicht ein Beispiel von an die Datenverarbeitungseinheit **218** eingegebenen Bilddaten und einer Bereichsaufteilung von diesen. [Fig. 9B](#) ist eine Tabelle, die ein Beispiel des Verarbeitungsergebnisses in dem Fall der Bilddaten durch die Bereichsaufteilungseinheit **301**, die Attributinformationen-Hinzufügungseinheit **302** und die Zeichenerkennungseinheit **303** veranschaulicht;

[0020] **Fig. 10** ist ein Ablaufdiagramm, das den Ablauf der Verarbeitung gemäß Ausführungsbeispiel 1 zum Verknüpfen eines Erläuterungstexts in einem Textkörper mit einem Objektbereich veranschaulicht;

[0021] **Fig. 11** ist ein Ablaufdiagramm, das den Ablauf einer Verarbeitung zum Berechnen des Ähnlichkeitsgrads gemäß Ausführungsbeispiel 1 veranschaulicht;

[0022] **Fig. 12A** und **Fig. 12B** veranschaulichen das Zwischenergebnis, das durch die Verarbeitung zum Berechnen des Ähnlichkeitsgrads erhalten wird. **Fig. 12A** veranschaulicht ein Beispiel des Ergebnisses, das durch die Verarbeitungen bis zu Schritt 1103 erhalten wird. **Fig. 12B** veranschaulicht ein Beispiel der Verarbeitungen in Schritten 1105 und 1106;

[0023] **Fig. 13** veranschaulicht durch eine Tabelle die Metadateninformationen, die erhalten werden, indem Bilddaten **901** bis **907** gemäß **Fig. 9A** der Verarbeitung gemäß Ausführungsbeispiel 1 unterzogen werden;

[0024] **Fig. 14** ist eine Darstellung, die die Beziehung zwischen **Fig. 14A** und **Fig. 14B** zeigt;

[0025] **Fig. 14A** und **Fig. 14B** sind Ablaufdiagramme, die den Ablauf der Verarbeitung gemäß Ausführungsbeispiel 2 zum Verknüpfen eines Erläuterungstexts in einem Textkörper mit einem Objektbereich veranschaulichen;

[0026] **Fig. 15** ist ein Ablaufdiagramm, das den Ablauf der Verarbeitung gemäß Ausführungsbeispiel 2 zum Berechnen des Ähnlichkeitsgrads veranschaulicht;

[0027] **Fig. 16** ist ein Ablaufdiagramm, das den Ablauf der Verarbeitung gemäß Ausführungsbeispiel 3 zum Verknüpfen eines Erläuterungstexts in einem Textkörper mit einem Objektbereich veranschaulicht;

[0028] **Fig. 17** veranschaulicht ein Beispiel des Benutzerschnittstelle-(UI-)Bildschirms, der auf einer Bedienungseinheit **203** angezeigt wird;

[0029] **Fig. 18** veranschaulicht ein Beispiel des Warnungsbildschirms, der auf der Bedienungseinheit **203** angezeigt wird; und

[0030] **Fig. 19** ist ein Ablaufdiagramm, das den Ablauf der Verarbeitung gemäß Ausführungsbeispiel 4 zum Verknüpfen eines Erläuterungstexts in einem Textkörper mit einem Objekt veranschaulicht.

BESCHREIBUNG DER AUSFÜHRUNGSBEISPIELE

[Ausführungsbeispiel 1]

[0031] Der folgende Abschnitt wird ein Ausführungsbeispiel zum Ausüben der vorliegenden Erfindung unter Bezugnahme auf die Zeichnungen beschreiben.

[0032] **Fig. 1** ist ein Blockschaltbild, das die Konfiguration eines Bildverarbeitungssystems gemäß diesem Ausführungsbeispiel veranschaulicht.

[0033] Gemäß **Fig. 1** ist das in dem Büro A aufgebaute LAN **102** mit einem Multifunktionsperipheriegerät (MFP) **100** verbunden, das eine Bildverarbeitungsvorrichtung zum Realisieren einer Vielzahl von Funktionstypen (z. B. einer Kopierfunktion, einer Druckfunktion, einer Übertragungsfunktion) ist. Das LAN **102** ist über einen Proxy-Server **103** auch mit einem externen Netzwerk **104** verbunden. Ein Client-PC **101** empfängt über das LAN **102** von dem MFP **100** übertragene Daten und verwendet eine dem MFP **100** eigene Funktion. Zum Beispiel kann der Client-PC **101** auch Druckdaten an das MFP **100** übertragen, um dadurch zu veranlassen, dass durch das MFP **100** Drucksachen basierend auf den Druckdaten gedruckt werden. Die Konfiguration gemäß **Fig. 1** ist ein Beispiel. Daher kann auch eine andere Konfiguration verwendet werden, bei der eine Vielzahl von Büros mit den gleichen Bestandselementen wie denjenigen des Büros A über ein Netzwerk **104** verbunden sind. Das Netzwerk **104** ist typischerweise ein Kommunikationsnetzwerk, das zum Beispiel durch das Internet, LAN, WAN, eine Telefonleitung, eine digitale Einzelleitung, ATM, eine Frame-Relay-Leitung, eine Kommunikationssatellitenleitung, eine Kabelfernsehleitung oder eine drahtlose Datenrundsendungsleitung realisiert ist. Das Netzwerk **104** kann jedes beliebige Netzwerk sein, solange Daten durch dieses gesendet und empfangen werden können. Die jeweiligen Endgeräte, nämlich der Client-PC **101** und der Proxy-Server **103**, weisen standardmäßige Bestandselemente auf, die in einen Universalrechner bereitgestellt sind, zum Beispiel einschließlich einer CPU, einem RAM, einem ROM, einer Festplatte, einer externen Speichervorrichtung, einer Netzwerkschnittstelle, einer Anzeige, einer Tastatur und einer Maus.

[0034] **Fig. 2** veranschaulicht die Konfiguration von dem MFP **100**.

[0035] Zunächst ist die Konfiguration von dem MFP **100** hauptsächlich in eine Scannereinheit **201**, die eine Bildeingabevorrichtung ist, eine Druckereinheit **202**, die eine Bildausgabevorrichtung ist, eine Steuereinheit **204**, die durch eine CPU **205** konfiguriert ist,

und eine Bedienungseinheit **203**, die eine Benutzerschnittstelle ist, gliedert.

[0036] Die Steuereinheit **204** ist mit der Scannereinheit **201**, der Druckereinheit **202** und der Bedienungseinheit **203** verbunden. Die Steuereinheit **204** ist auch eine Steuerungseinrichtung, die mit dem LAN **219** oder einer öffentlichen Leitung (WAN) **220** verbunden ist, die ein gewöhnliches Telefonleitungsnetz darstellt, um dadurch eine Eingabe und eine Ausgabe von Bildinformationen und Geräte- bzw. Vorrichtungsinformationen bereitzustellen.

[0037] Die CPU **205** steuert die jeweiligen Einheiten, die in der Steuereinheit **204** umfasst sind.

[0038] Der RAM **206** ist ein Systemarbeitspeicher für den Betrieb der CPU **205**, und ist auch ein Bildspeicher zum vorübergehenden Speichern von Bilddaten.

[0039] Der ROM **210** ist ein Lade-ROM, in dem ein Programm, wie etwa ein Systemladeprogramm, gespeichert wird.

[0040] Eine Speichereinheit **211** ist ein Festplattenlaufwerk, in dem Systemsteuersoftware und Bilddaten gespeichert werden.

[0041] Eine Bedienungseinheit-I/F **207** ist eine Schnittstelleneinheit mit einer Bedienungseinheit (UI) **203**, die auf der Bedienungseinheit **203** anzuzeigende Bilddaten an die Bedienungseinheit **203** ausgibt. Die Bedienungseinheit-I/F **207** weist auch eine Funktion zum Übertragen von Informationen, die durch einen Benutzer dieser Bildverarbeitungsvorrichtung durch die Bedienungseinheit **203** eingegeben werden, an die CPU **205** auf.

[0042] Eine Netzwerk-I/F **208** verbindet diese Bildverarbeitungsvorrichtung mit dem LAN **219**, um eine Eingabe und eine Ausgabe von Informationen eines Pakettyps bereitzustellen.

[0043] Ein Modem **209** verbindet diese Bildverarbeitungsvorrichtung mit einem WAN **220**, um eine Datendemodulation und -modulation bereitzustellen, um dadurch eine Eingabe und eine Ausgabe von Informationen bereitzustellen. Die Vorrichtungen, wie sie vorstehend beschrieben sind, sind an einem Systembus **221** eingerichtet.

[0044] Eine Bildbus-I/F **212** ist eine Busbrücke, die den Systembus **221** mit einem Bildbus **222** zum Übermitteln von Bilddaten mit einer hohen Geschwindigkeit verbindet, um die Datenstruktur umzuwandeln.

[0045] Der Bildbus **222** besteht zum Beispiel aus einem PCI-Bus und IEEE 1394. Der Bildbus **222** hat die folgenden Vorrichtungen darauf eingerichtet.

[0046] Ein Rasterbildverarbeiter (RIP) **213** realisiert eine sogenannte Rendering-Verarbeitung, um einen Code einer Seitenbeschreibungssprache (PDL) zu analysieren, um den Code in ein Bitmapbild mit einer festgelegten Auflösung zu entwickeln. Diese Entwicklung fügt Attributinformationen auf Grundlage einer Pixeleinheit oder einer Bereichseinheit hinzu. Dies wird als eine Bildbereichsbestimmungsverarbeitung bezeichnet. Durch die Bildbereichsbestimmungsverarbeitung werden Attributinformationen, die den Objekttyp zeigen, wie etwa Zeichen (Text), eine Linie, Grafik, oder ein Bild, mit Bezug auf jedes Pixel bzw. Bildelement oder jeden Bereich hinzugefügt. Zum Beispiel gibt ein RIP **213** abhängig von dem Typ des Objekts der PDL-Beschreibung in einem PDL-Code ein Bildbereichssignal aus. Dann werden die Attributinformationen abhängig von dem durch den Signalwert gezeigten Attribut gespeichert, während sie mit einem Pixel bzw. Bildelement oder einem Bereich verknüpft bzw. verbunden werden, der dem Objekt entspricht. Somit werden die Bilddaten durch die mit diesen verknüpften bzw. verbundenen Attributinformationen begleitet.

[0047] Eine Geräte-I/F **214** verbindet über eine Signalleitung **223** die Scannereinheit **201** als eine Eingabevorrichtung mit der Steuereinheit **204**. Die Geräte-I/F **214** verbindet über eine Signalleitung **224** auch die Druckereinheit **202** als eine Bildausgabevorrichtung mit der Steuereinheit **204**. Die Geräte-I/F **214** stellt eine synchrone/asynchrone Umwandlung von Bilddaten bereit.

[0048] Eine Scannerbild-Verarbeitungseinheit **215** korrigiert, verarbeitet und editiert eingegebene Bilddaten.

[0049] Eine Druckerbild-Verarbeitungseinheit **216** unterzieht Druckausgabebilddaten, die an die Druckereinheit **202** auszugeben sind, zum Beispiel einer Korrektur und einer Auflösungsumwandlung abhängig von der Druckereinheit **202**.

[0050] Eine Bilddrehungseinheit **217** dreht die eingegebenen Bilddaten, um so die Daten aufzurichten und die Daten auszugeben.

[0051] Die Datenverarbeitungseinheit **218** wird unter Bezugnahme auf [Fig. 3](#) ausführlich beschrieben.

<Datenverarbeitungseinheit>

[0052] Wie es gemäß [Fig. 3](#) gezeigt ist, besteht die Datenverarbeitungseinheit **218** aus: einer Bereichsaufteilungseinheit (Bereichsextraktionseinheit) **301**, der Attributinformationen-Hinzufügungseinheit **302**, der Zeichenerkennungseinheit **303**, der Metadaten-Verarbeitungseinheit **304** und einer Formatumwandlungseinheit **305**. Die Datenverarbeitungseinheit **218** unterzieht eingegebene Bilddaten **300** den vorbe-

stimmten Verarbeitungen durch die jeweiligen Verarbeitungseinheiten **301** bis **305**. Schließlich wird ein elektronisches Dokument **310** erzeugt und ausgegeben.

[0053] Die Bereichsaufteilungseinheit **301** empfängt Bilddaten, die durch die Scannereinheit **201** gelesen werden, oder Bilddaten, die von einer externen Vorrichtung, wie etwa einem Client-PC, empfangen werden und in der Speichereinheit gespeichert werden. Um auf Grundlage einer Seiteneinheit aus den eingegebenen Bilddaten Bereiche zu extrahieren, die basierend auf jedem Attribut aufgeteilt sind, wie etwa einem Zeichen, einer Fotografie, einer Zeichnung oder einer Tabelle, werden die Pixel bzw. Bildelemente in den Bilddaten einer Verarbeitung, wie etwa einer Extraktion oder Gruppierung, unterzogen. Dies kann durch ein bekanntes Bereichsaufteilungsverfahren (Bereichsextraktionsverfahren) durchgeführt werden. Ein Beispiel eines solchen Verfahrens wird nachstehend beschrieben. Zunächst wird ein eingegebenes Bild binarisiert bzw. digitalisiert, um ein binäres Bild zu erzeugen. Dann wird bewirkt, dass das binäre Bild eine niedrigere Auflösung aufweist, um dadurch ein Culling- bzw. Auslese-/Aussonderungsbild (reduziertes Bild) zu erstellen. Zum Beispiel wird, um ein $1/(M \times N)$ -Ausdünnungsbild zu erstellen, das binäre Bild aufgeteilt, so dass es den jeweiligen $M \times N$ Pixeln entspricht. Falls die $M \times N$ Pixel schwarze Pixel umfassen, werden entsprechende Pixel in dem reduzierten Bild als schwarze Pixel festgelegt. Falls die $M \times N$ Pixel keine schwarzen Pixel umfassen, werden entsprechende Pixel in dem reduzierten Bild als weiße Pixel festgelegt. Auf diese Art und Weise wird ein Ausdünnungsbild erstellt. Als Nächstes werden Teile in dem Ausdünnungsbild extrahiert, in denen schwarze Pixel verbunden sind bzw. zusammenhängen (schwarze Verbindungspixel), um dadurch ein Rechteck zu erstellen, das durch die schwarzen Verbindungspixel begrenzt ist. Wenn Rechtecke mit einer ähnlichen Größe wie diejenige eines Zeichenbildes (Rechtecke für ein Zeichen) angeordnet sind, oder wenn ähnliche Rechtecke, deren Höhe oder Breite ähnlich zu derjenigen einer Zeichenbildgröße (Rechtecken von schwarzen Verbindungspixeln, die eine Vielzahl von Zeichen verbunden haben) ist, in der Nähe der kurzen Seiten angeordnet sind, können diese Rechtecke mit einer hohen Wahrscheinlichkeit Zeichenbilder bilden, die eine Zeichenreihe bzw. -folge darstellen. In diesem Fall sind die Rechtecke miteinander verbunden, um dadurch ein Rechteck zu erhalten, das eine Zeichenreihe bzw. -folge darstellt. In dem Fall einer Ansammlung von Rechtecken, die eine Zeichenreihe bzw. -folge darstellen, wobei deren kurze Seiten im Wesentlichen die gleiche Länge aufweisen, und die in einer Spaltenrichtung in im Wesentlichen gleichen Abständen angeordnet sind, kann diese Ansammlung mit einer gewissen Wahrscheinlichkeit einen Textkörper bzw. Körpertext darstellen. Somit werden diese Rechtecke verbunden und wird

der resultierende Teil als ein Textkörper- bzw. Körpertextbereich extrahiert. Ein Fotografiebereich, ein Zeichnungsbereich und ein Tabellenbereich werden durch verbundene schwarze Pixel extrahiert, deren Größe größer ist als diejenige eines Zeichenbildes. Als Ergebnis zum Beispiel die entsprechenden Bereiche extrahiert, wie sie durch die Bezugszeichen **501** bis **506** gemäß [Fig. 5A](#) gezeigt sind. Die Attribute der jeweiligen Bereiche werden, wie es nachstehend beschrieben ist, zum Beispiel basierend auf den Größen, dem Seitenverhältnis oder einer Dichte schwarzer Pixel oder dem Konturenverfolgungsergebnis von weißen Pixeln, die in den schwarzen Verbindungspixeln umfasst sind, bestimmt.

[0054] Die Attributinformationen-Hinzufügungseinheit **302** fügt die Attributinformationen zu den jeweiligen Bereichen hinzu, die durch die Bereichsaufteilungseinheit **301** aufgeteilt sind. Die folgende Beschreibung wird basierend auf der Annahme vorgenommen, dass Bilddaten **500** durch die Bereichsaufteilungseinheit **301** aufgeteilt sind, wie es gemäß [Fig. 5A](#) gezeigt ist.

[0055] Ein Bereich **506** auf einer Seite umfasst darin Zeichen und Linien gleich oder größer einer vorbestimmten Anzahl und weist zum Beispiel eine Form eines Absatzes auf. Daher wird der Bereich **506** mit einem Attribut „Textkörper“ versehen. Um die Zeichnung einfach zu veranschaulichen, sind die in dem Bereich **506** gemäß [Fig. 5A](#) umfassten Zeichen durch schwarze Punkte dargestellt. Der Bereich **506** umfasst jedoch eigentlich eine Vielzahl von Zeichenbildern.

[0056] Mit Bezug auf die verbleibenden Bereiche **501** bis **505** wird bestimmt, ob die Bereiche **501** bis **505** ein Bereich sind oder nicht, der Zeichen umfasst. Im Speziellen erscheinen in dem Fall eines Zeichen umfassenden Bereichs in dem Bereich Rechtecke für Zeichenbilder in zyklischer bzw. periodischer Weise. Somit wird bestimmt, ob der Bereich ein Bereich ist oder nicht, der Rechtecke umfasst, deren Größe ähnlich einer Zeichenbildgröße ist. Als Ergebnis werden der Bereich **501**, der Bereich **504** und der Bereich **505** als Zeichen umfassende Bereiche bestimmt und werden sie mit einem Attribut „Zeichen umfassender Bereich“ versehen.

[0057] Mit Bezug auf einen anderen Bereich als die vorstehend beschriebenen wird andererseits zunächst die Größe des Bereichs bestimmt. Wenn die Bereichsgröße sehr klein ist, wird deren Attribut als „Rauschen“ bestimmt. Wenn die Bereichsgröße nicht so klein und gleich oder größer einer festgelegten Größe ist, ist der Bereich ein Bereich für ein bestimmtes Objekt. Daher wird dieser Bereich weiter der folgenden Bestimmung unterzogen, um ein Attribut zu bestimmen, das zu dem Bereich hinzugefügt ist (d. h., welchem Objekt dieser Bereich ent-

spricht). Zunächst werden weiße Pixel in schwarzen Verbindungspixeln mit einer niedrigen Pixeldichte einer Konturenverfolgung unterzogen. Dann wird bestimmt, ob Rechtecke, die die Konturen weißer Pixel begrenzen, geordnet angeordnet sind oder nicht. Wenn die Rechtecke, die die Konturen weißer Pixel begrenzen, geordnet angeordnet sind, wird der Bereich dahingehend bestimmt, dass er ein Attribut „Tabelle“ aufweist. Wenn die Rechtecke, die die Konturen weißer Pixel begrenzen, nicht geordnet angeordnet sind, wird der Bereich dahingehend bestimmt, dass er ein Attribut „Zeichnung (Strich- bzw. Linienzeichnung)“ aufweist. Andere Bereiche als dieser Bereich, die eine hohe Pixeldichte aufweisen, werden als einem Bild oder einer Fotografie entsprechend bestimmt und werden dahingehend bestimmt, dass sie ein Attribut „Fotografie“ aufweisen. In dem vorstehenden Abschnitt wurde ein Bereich basierend auf den drei Typen von Objektattributen klassifiziert, nämlich „Fotografie“, „Zeichnung (Strich- bzw. Linienzeichnung)“ und „Tabelle“. Die vorliegende Erfindung ist jedoch nicht darauf beschränkt. Es können auch andere Bestimmungskriterien verwendet werden, so dass ein Bereich basierend auf beliebigen Typen von Objektattributen klassifiziert werden kann.

[0058] Falls ein Zeichen umfassender Bereich angrenzend an (oder genau über oder genau unter) einem Objektbereich vorhanden ist, der mit einem Attribut, wie etwa „Fotografie“, „Zeichnung“ oder „Tabelle“ versehen ist, wird bestimmt, dass dieser Bereich ein Bereich von Zeichen zum Beschreiben des Objektbereichs ist und wird er daher mit einem Attribut „Beschriftung“ versehen. Ein mit einem Attribut „Beschriftung“ versehener Bereich wird gespeichert, während er mit einem Bereich, wie etwa „Fotografie“, „Zeichnung“ oder „Tabelle“ verknüpft ist, der durch die Beschriftung begleitet wird, so dass der letztere Bereich identifiziert werden kann.

[0059] Wenn ein Bereich größer ist als das Zeichenbild eines Textkörperteils und sich an einer anderen Position befindet als der Position der Spalteneinstellung des Textkörperteils, wird dieser Bereich mit einem Attribut „Überschrift“ versehen. Wenn ein Bereich größer ist als das Zeichenbild eines Textkörperteils und auf der oberen Seite der Spalteneinstellung des Textkörperteils liegt, wird dieser Bereich mit einem Attribut „Zwischenüberschrift“ versehen. Wenn ein Bereich für ein Zeichenbild mit einer kleineren Größe als der Größe des Zeichenbildes des Textkörperteils vorgesehen ist und an einem unteren Ende oder einem oberen Ende eines Manuskripts liegt, wird dieser Bereich mit einem Attribut „Seite“ (oder „Seitenkopfzeile“ oder „Seitenfußteile“) versehen. In dem Fall des Bereichs, der als ein Zeichen umfassender Bereich bestimmt wurde und der nicht als „Textkörper“, „Überschrift“, „Zwischenüberschrift“, „Beschriftung“ oder „Seite“ bestimmt wurde, wird der Bereich mit einem Attribut „Zeichen“ versehen.

[0060] Durch die Verarbeitung zum Hinzufügen von Attributinformationen, wie sie vorstehend beschrieben ist, werden die Bilddaten **500** so eingestellt, dass ein Bereich **501** mit einem Überschrift-Attribut versehen wird, ein Bereich **502** mit einem Tabelle-Attribut versehen wird, ein Bereich **503** mit einem Fotografie-Attribut versehen wird, ein Bereich **504** mit einem Zeichen-Attribut versehen wird, ein Bereich **505** mit einem Beschriftung-Attribut versehen wird (das den Bereich **503** begleitet), und der Bereich **506** mit einem Textkörper-Attribut versehen wird. Die Unterstreichungen, die den jeweiligen Attributen für die Bereiche **501**, **504** und **505** gegeben sind, zeigen an, dass ein Attribut „Zeichen umfassender Bereich“ zu diesen hinzugefügt ist.

[0061] Die Zeichenerkennungseinheit **303** führt eine bekannte Zeichenerkennungsverarbeitung (OCR-Verarbeitung) auf Zeichenbild umfassenden Bereichen (Bereichen, die mit Attributen wie „Zeichen“, „Textkörper“, „Überschrift“, „Zwischenüberschrift“ und „Beschriftung“ versehen sind) durch. Ein Bereich für ein Zeichenbild in „Tabelle“ kann auch mit einem Attribut „tabelleninternes Zeichen“ versehen werden und anschließend einer Zeichenerkennungsverarbeitung unterzogen werden. Dann wird eine durch die Zeichenerkennungsverarbeitung erhaltene Zeichencodespalte als Zeichencodeinformation gespeichert und mit einem Zielbereich verknüpft.

[0062] Wie es vorstehend beschrieben ist, speichert die Speichereinheit **211** darin zum Beispiel die Position und Größe des Bereichs, die Bereichsattributinformation, die Seiteninformation, die Information für das Ergebnis der Zeichenerkennungsverarbeitung (Zeichencodespalte), die durch die jeweiligen Einheiten der Bereichsaufteilungseinheit **301**, der Attributinformationen-Hinzufügungseinheit **302** und der Zeichenerkennungseinheit **303** extrahiert werden. [Fig. 5B](#) veranschaulicht durch eine Tabelle das Ergebnis der vorstehend beschriebenen Verarbeitung der Bilddaten **500**. Somit wird das Ergebnis in der Speichereinheit **211** in einer Tabellenform gespeichert. Der Bereich **504** ist ein Bereich für ein Zeichenbild, das in einer Fotografie **503** liegt. Somit wird der Bereich **504** mit einem Attribut „in Fotografie **503**“ versehen. Obwohl die Plätze/Stellen in der Tabelle hinsichtlich der Position und Größe des Bereichs (z. B. Koordinaten X/Y, Breite W, Höhe H) Markierungen wie etwa X1 umfassen, umfassen diese Plätze/Stellen in einem realen Fall numerische Werte.

[0063] Der Objektbereich, wie etwa „Fotografie“, „Zeichnung“ oder „Tabelle“, der durch die Attributinformationen-Hinzufügungseinheit **302** als ein durch eine Beschriftung begleiteter Bereich bestimmt wird, wird durch die Metadaten-Verarbeitungseinheit **304** mit Metadaten verknüpft bzw. verbunden, die als Informationen zum Suchen des Objekts fungieren. Dann speichert die Attributinformationen-Hinzufü-

gungseinheit **302** die resultierenden Daten in der Speichereinheit **211**. Im Speziellen verknüpft die Attributinformationen-Hinzufügungseinheit **302**, als Metadaten zum Suchen des Objekts, eine Zeichenkette, die als eine Beschriftung für einen speziellen Objektbereich beschrieben ist, mit einem Erläuterungsteil in einem Textkörper (einem Erläuterungstext in einem Textkörper), der ein identisches Wort wie als ein in der Zeichenkette umfasstes Wort verwendet, wie etwa eine Figurennummer (Ankerausdruck).

[0064] Die Verknüpfung von Metadaten wird durch eine (hierin nachstehend als „Beschriftungskennung“ bezeichnete) Kennung bzw. ein Identifizierungszeichen durchgeführt, das jedem Objekt gegeben ist. Diese Beschriftungskennung kann verwendet werden, um eine Beschriftung oder einen Erläuterungstext in einem Textkörper mit einem durch die Beschriftung begleiteten Objektbereich auf geeignete Weise zu verknüpfen. Durch Hinzufügen von unterschiedlichen Beschriftungskennungen zu den einzelnen Objekten können selbst dann, wenn es eine Vielzahl von Objekten gibt, die einen identischen Ankerausdruck umfassende Beschriftungen aufweisen, Metadaten auf geeignete Weise mit den Objekten verknüpft werden. Bei diesem Ausführungsbeispiel ist die Beschriftungskennung eine ID zum eindeutigen Identifizieren eines durch eine Beschriftung begleiteten Objektbereichs (d. h. eine Seriennummer (positive Ganzzahl), die mit einem Wert „1“ beginnt). Die Beschriftungskennung kann auch eine Positionsinformation sein, wie etwa eine Adresse oder ein Zeiger, die/der die Position anzeigt, an der die Zeichenerkennungsinformation für eine Beschriftung oder einen Textkörper gespeichert ist, sowie eine Bezugsinformation, wie etwa XPath oder URL.

[0065] Die Formatumwandlungseinheit **305** verwendet die jeweiligen Typen von Informationen (z. B. Seiteninformation, die Position oder Größe des Bereichs, Attribut, Zeichenerkennungsinformation, Metadaten), die durch die jeweiligen vorgenannten Verarbeitungseinheiten erhalten werden, um die Bilddaten in ein elektronisches Dokument mit einem vorbestimmten Format umzuwandeln. Das vorbestimmte Format kann zum Beispiel PDF, SVG, XPS oder OfficeOpenXML sein. Das durch die Formatumwandlung erzeugte elektronische Dokument umfasst zum Beispiel Seitenanzeigeeinformationen durch graphische Darstellungen (z. B. ein anzuzeigendes Bild) und Inhaltsinformationen durch semantische Beschreibung, wie etwa Zeichen (z. B. Metadaten).

[0066] Die Formatumwandlungseinheit **305** führt hauptsächlich die folgenden beiden Verarbeitungen durch. Eine Verarbeitung besteht darin, einen Bildbereich Filterverarbeitungen (z. B. Histogrammausgleichung, Glättung, Kanten hervorhebung, Farbquantisierung, Binarisierung) zu unterziehen, so dass die Bilddaten (z. B. ein Bild des Teils, der einem Bereich

entspricht, der mit einem Attribut „Zeichnung (Strich- bzw. Linienzeichnung)“ versehen ist) in einem elektronischen Dokument mit einem vorbestimmten Format gespeichert werden können. Die Bilddaten können in einem elektronischen Dokument mit einem vorbestimmten Format gespeichert werden, indem die Bilddaten in Vektorpfadbeschreibung-Grafikdaten (Vektordaten) oder Bitmapbeschreibung-Grafikdaten (z. B. JPEG-Daten) umgewandelt werden. Die Umwandlung in Vektordaten kann durch eine bekannte Vektorisierungstechnik durchgeführt werden. Diese Umwandlung erzeugt auch eine Grafikbeschreibung (Vektorpfadbeschreibung), wie etwa einen Rahmen, der angezeigt wird, wenn das Suchergebnis für eine Objektsuche identifiziert oder hervorgehoben wird. Die andere Verarbeitung besteht darin, zu den erzeugten Vektordaten oder Bitmapdaten Bereichsinformationen (Position, Größe, Attribut), Informationen für die Zeichenerkennung in dem Bereich und in der Speichereinheit **211** gespeicherte Metadaten hinzuzufügen, um dadurch ein elektronisches Dokument mit einem vorbestimmten Format zu erstellen.

[0067] Ein Umwandlungsverarbeitungsverfahren, das durch die Formatumwandlungseinheit **305** auf jedem Bereich durchgeführt werden sollte, hängt häufig von dem Attribut des Bereichs ab. Zum Beispiel ist die Vektorumwandlungsverarbeitung für eine Grafik geeignet, die aus schwarz und weiß oder einer geringen Anzahl von Farben zusammengesetzt ist, wie etwa ein Zeichen oder eine Strich- bzw. Linienzeichnung, aber sie ist nicht für einen Bildbereich mit Gradierung, wie etwa eine Fotografie, geeignet. Um abhängig von dem Attribut von jedem Bereich eine geeignete Umwandlung bereitzustellen, kann auch eine Entsprechungstabelle, wie sie gemäß [Fig. 5C](#) gezeigt ist, im Vorhinein bereitgestellt werden.

[0068] Zum Beispiel ist die gemäß [Fig. 5C](#) gezeigte Entsprechungstabelle so eingestellt, dass als eine Umwandlungsverarbeitung Bereiche mit Attributen „Zeichen“, „Zeichnung (Strich- bzw. Linienzeichnung)“ und „Tabelle“ einer Vektorumwandlungsverarbeitung unterzogen werden und ein Bereich mit einem Attribut „Fotografie“ einer Bildausschnittverarbeitung unterzogen wird. Die gemäß [Fig. 5C](#) gezeigte Entsprechungstabelle ist auch so eingestellt, dass für jedes Attribut eingestellt ist, ob die Pixelinformationen des Bereichs aus den Bilddaten gelöscht werden oder nicht. Zum Beispiel ist, wenn die Daten für einen Bereich mit einem Attribut „Zeichen“ in Vektorpfadbeschreibungsdaten umgewandelt werden, der Bereich so eingestellt, dass er der Lösungsverarbeitung unterzogen wird. In diesem Fall wird eine derartige Verarbeitung durchgeführt, dass Pixel, die einem Teil entsprechen, der durch den umgewandelten Vektorpfad abgedeckt wird, mit der umgebenden Farbe gefüllt werden. Auf ähnliche Weise wird, wenn der Bereich mit einem Attribut „Fotografie“ als ein rechteckiger Bildteil ausgeschnitten wird, ein Gebiet, das

dem ausgeschnittenen Gebiet entspricht, zum Beispiel mit der umgebenden Farbe gefüllt. Der Zweck der Lösungsverarbeitung wie vorstehend beschrieben besteht darin, Bilddaten, für die die jeweiligen Bereiche bereits der Füllungsverarbeitung unterzogen wurden, als Teile zu verwenden, die die „Hintergrund“-Bilddaten bilden. Die Hintergrundbilddaten (das Hintergrundbild) umfassen verbleibende Teile, die von den durch die Bereichsaufteilungsverarbeitung extrahierten Bereichen abweichen (z. B. Pixel, die einer Basis bzw. einem Unterbau in Bilddaten entsprechen). Die Datenbeschreibung eines elektronischen Dokuments wird durch Überlagerung der Grafikdaten (Vordergrundbild), die durch eine nicht gezeigte Vektorumwandlungsverarbeitungseinheit oder Bildausschnittverarbeitungseinheit erhalten werden, auf das Hintergrundbild erhalten, um dadurch das resultierende Bild anzuzeigen. Dies kann folglich ein Fehlen von Informationen für Hintergrundpixel (Hintergrundfarbe) verhindern und nicht redundante Grafikdaten konfigurieren.

[0069] Alternativ kann auch eine Vielzahl von Entsprechungstabellen im Vorhinein erstellt werden, so dass eine geeignete Entsprechungstabelle abhängig von einer Anwendung des elektronischen Dokuments ausgewählt werden kann. Zum Beispiel werden, wenn die Entsprechungstabelle gemäß [Fig. 5C](#) verwendet wird, die Objekte größtenteils so umgewandelt, dass sie eine Vektorpfadbeschreibung aufweisen, wodurch eine hohe Bildqualität für eine vergrößerte oder verkleinerte Anzeige bereitgestellt wird. Daher ist zum Beispiel eine andere Entsprechungstabelle bereitgestellt, gemäß der unterschiedliche Binärbilder für die jeweiligen Zeichenfarben von Zeichenbildern erzeugt werden und einer verlustfreien Komprimierung unterzogen werden, und andere Bilder als die vorgenannten Bilder als ein Hintergrundbild einer JPEG-Komprimierung unterzogen werden. Das Erstgenannte ist zum Beispiel für eine Anwendung zum Wiederverwenden eines Grafikeditors geeignet. Das Letztgenannte ist für einen Fall geeignet, bei dem es gewünscht ist, ein derartiges elektronisches Dokument zu erstellen, für das Zeichenbilder auf einfache Weise gelesen werden können, während eine hohe Komprimierungsrate erreicht wird. Durch Verwendung unterschiedlicher Entsprechungstabellen abhängig von einer Anwendung, kann ein geeignetes elektronisches Dokument erstellt werden, das für eine Anwendung für einen Benutzer geeignet ist.

[0070] [Fig. 6A](#) veranschaulicht ein Beispiel eines durch die Datenverarbeitungseinheit **218** erzeugten elektronischen Dokuments. [Fig. 6A](#) zeigt ein elektronisches Dokument **600**, das basierend auf den Bilddaten **500** gemäß [Fig. 5A](#) erzeugt ist und das basierend auf dem SVG-Format (SVG: Scalable Vector Graphics) beschrieben ist. [Fig. 6B](#) veranschaulicht Informationen für zu dem elektronischen Dokument

600 hinzugefügte Metadaten der Einfachheit halber in der Form einer Tabelle.

[0071] In diesem Fall bezeichnen die Bezugszeichen **601** bis **606** gemäß [Fig. 6A](#) Grafikbeschreibungen, die den Bereichen **501** bis **506** in den Bilddaten **500** entsprechen. Die Bezugszeichen **601** und **604** bis **606** bezeichnen Zeichenzeichnungsbeschreibungen durch Zeichencodes. Das Bezugszeichen **602** bezeichnet eine Vektorpfadbeschreibung für den Rahmen der Tabelle, der einer Vektorumwandlung unterzogen ist. Das Bezugszeichen **603** bezeichnet eine Beschreibung zum Anhängen eines Fotografiebildes. Die Beschreibung **603** umfasst einen Beschriftungsbezeichner (caption_id) **608** mit Wert „1“.

[0072] Das Bezugszeichen **607** bezeichnet eine Beschreibung für Metadaten. Die Beschreibung **607** beschreibt die Zeichen in dem Fotografieobjekt, nämlich „neues Produkt“, die Beschriftung „[Fig. 1](#) AAA“, einen Anker Ausdruck in der Beschriftung, d. h. „[Fig. 1](#)“, und einen Erläuterungstext in einem Textkörper, d. h. „[Fig. 1](#) ist XX betreffend AAA“, der basierend auf dem Anker Ausdruck und der Zeichenkette des Textkörpers extrahiert ist. Außerdem ist auch ein Bezeichner **609** beschrieben, der identisch zu einem Beschriftungsbezeichner **608** ist. Anstelle des Erläuterungstexts in dem Textkörper (Satz), d. h. „[Fig. 1](#) ist XX betreffend AAA“, der den Anker Ausdruck „[Fig. 1](#)“ umfasst, kann auch ein Wort „AAA“ extrahiert und als Metadaten hinzugefügt werden. Alternativ können auch der Erläuterungstext in dem Textkörper und das Wort beide als Metadaten verwendet werden. Das Gleiche gilt für eine Zeichenkette einer Beschriftung. Somit kann ein aus einer Beschriftung extrahiertes Wort weiter als Metadaten hinzugefügt werden.

<Metadaten-Verarbeitungseinheit>

[0073] [Fig. 4](#) ist ein Blockschaltbild, das die interne Konfiguration der Metadaten-Verarbeitungseinheit **304** veranschaulicht.

[0074] Das Bezugszeichen **401** bezeichnet eine Metadaten-Hinzufügungsziel-Auswahleinheit, die eine Verarbeitung zum Auswählen eines Bereichs, der der Extraktion/Hinzufügung von Metadaten zu unterziehen ist, aus eingegebenen Bilddaten durchführt. Bei diesem Ausführungsbeispiel werden Bilddaten mit einem Beschriftungsbereich als ein Ziel ausgewählt, um der Extraktion/Hinzufügung von Metadaten, wie etwa einer Beschriftung, einem Anker Ausdruck oder einem Erläuterungstext in einem Textkörper, unterzogen zu werden.

[0075] Das Bezugszeichen **402** bezeichnet eine Anker Ausdruck-Extraktionseinheit, die einen Anker Ausdruck aus dem Beschriftungsbereich extrahiert, der durch die Metadaten-Hinzufügungsziel-Auswahleinheit **401** ausgewählt wird. Im Speziellen analysiert die

Anker Ausdruck-Extraktionseinheit **402** die mit dem ausgewählten Beschriftungsbereich verknüpften Zeichenerkennungsinformationen, um nach dem Vorhandensein irgendeines Anker Ausdrucks, wie etwa „**Fig. 1**“, in den Zeichenerkennungsinformationen zu suchen. Wenn ein Anker Ausdruck gefunden wird, wird der entsprechende Teil als ein Anker Ausdruck extrahiert und werden andere Teile als der vorgenannte Teil als ein Beschriftungsausdruck extrahiert. Diese Extraktion verwendet zum Beispiel eine Zeichencodecharakteristik oder ein Wörterbuch, um eine unbedeutende Zeichenkette (z. B. eine unsinnige Symbolfolge) auszuschließen. Dies verhindert folglich, dass eine Aufteilungslinie oder ein Bild, das an einer Grenze eines Textteils eines Dokuments erscheint, bei der Zeichenerkennung fälschlich als ein Zeichen erkannt wird. Außerdem kann ein Anker Ausdruck durch Verwendung eines mehrsprachigen Zeichenkettenmusters, wie etwa einer Zeichnungsnummer, oder eines Falschen Zeichenerkennungsmusters für das Zeichenkettenmuster, extrahiert werden, um dadurch eine Anker Ausdruck-Extraktionsgenauigkeit zu erhalten und die Zeichen eines Anker Ausdrucks zu korrigieren. Ein Beschriftungsausdruck kann auch einer Analyse durch eine Verarbeitung natürlicher Sprache oder der Korrektur einer falschen Zeichenerkennung unterzogen werden. Somit kann auch eine Grenze mit einem Anker Ausdruck oder ein Symbol oder eine Zeichenverzerrung, was oben oder unten erscheint, korrigiert und ausgeschlossen werden.

[0076] Das Bezugszeichen **403** bezeichnet eine textkörperinterne Sucheinheit, die einen Ausdruck (Satz) in einem Textkörper, der einen Anker Ausdruck umfasst, der durch die Anker Ausdruck-Extraktionseinheit **402** erhalten wird, in den Zeichenerkennungsinformationen sucht, die mit dem Textkörperteil verknüpft sind, um diesen als einen Erläuterungstext in dem Textkörper zu extrahieren. Um eine Hochgeschwindigkeitssuche zu realisieren, kann auch ein Suchindex erstellt und verwendet werden. Die Erstellung des Index und eine Hochgeschwindigkeitssuche unter Verwendung des Index können durch eine bekannte Technik erreicht werden, und sie stellen keinen wesentlichen Punkt gemäß der vorliegenden Erfindung dar und werden daher nicht weiter beschrieben. Alternativ können viele Anker Ausdrücke gleichzeitig gesucht werden, um dadurch eine hohe Geschwindigkeit zu realisieren. Ein Erläuterungstext in einem Textkörper, der bei der Suche gefunden wird, kann auch mit einem Anker Ausdruck kombiniert werden, der mit einem mehrsprachigen Zeichenkettenmuster beschrieben ist.

[0077] Das Bezugszeichen **404** bezeichnet eine Ausdrucksähnlichkeitsgrad-Berechnungseinheit, die basierend auf dem Anker Ausdruck einen Beschriftungsausdruck mit einem Erläuterungstext in einem Textkörper vergleicht, um den Grad einer Ähnlichkeit

zwischen diesen zu berechnen. Im Speziellen vergleicht die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** den Beschriftungsausdruck mit dem Erläuterungstext in dem Textkörper auf der Zeichenebene oder durch eine Analyse natürlicher Sprache, um dadurch einen Vergleich basierend auf einer Wort- oder Bedeutungsebene zu vollziehen. Dann wird der Ähnlichkeitsgrad berechnet, wobei die Positionen von Anker Ausdrücken in den beiden Ausdrücken berücksichtigt werden. Der Ähnlichkeitsgrad ist ein Kriterium, das einen erhöhten Wert hat, wenn die beiden Ausdrücke ein Zeichen oder ein Wort aufweisen, das näher an dem Anker Ausdruck ist. Die Details des Ähnlichkeitsgrads werden nachstehend beschrieben.

[0078] Das Bezugszeichen **405** bezeichnet eine Metadaten-Sammel-/Ausgabeeinheit, die durch die jeweiligen vorgenannten Einheiten extrahierte Metadaten sammelt, um die Metadaten mit Bilddaten zu verknüpfen, die mit den Metadaten zu versehen sind, um dadurch das Ergebnis an die Formatumwandlungseinheit **305** auszugeben.

[0079] Das Bezugszeichen **406** stellt eine Metadaten-Verarbeitungssteuereinheit dar, die Bilddaten **300** basierend auf den Bereichsinformationen (Position, Größe, Attribut) **411**, den bereichsinternen Zeichenerkennungsinformationen **412** und den Metadaten **413**, die in der Speichereinheit **211** gespeichert sind, an geeignete Verarbeitungseinheiten **401** bis **405** zuweist. Dann steuert die Metadaten-Verarbeitungssteuereinheit **406** die Gesamtheit der Metadaten-Verarbeitungseinheit **304**, so dass die von den jeweiligen Verarbeitungseinheiten **401** bis **404** ausgegebenen Daten durch die Metadaten-Sammel-/Ausgabeeinheit **405** integriert werden können.

[0080] Als Nächstes wird der folgende Abschnitt den Überblick der Verarbeitung durch eine Bildverarbeitungsvorrichtung (das MFP **100**) gemäß diesem Ausführungsbeispiel unter Bezugnahme auf das Ablaufdiagramm gemäß **Fig. 7** beschreiben. Die folgende Beschreibung wird basierend auf einem Fall vorgenommen, bei dem elektronische Dokumente, für die die jeweiligen Objekte mit geeigneten Metadaten versehen sind, mit Bezug auf gemäß **Fig. 9** gezeigte Bilddaten (**901** bis **903** und **904** bis **907**), die der Gesamtheit von sieben Seiten entsprechen, die aus zwei unterschiedlichen Dokumententypen zusammengesetzt sind, erzeugt werden.

[0081] Die Verarbeitungen in den jeweiligen Schritten, die hierin als diejenigen beschrieben werden, die durch die jeweiligen Verarbeitungseinheiten durchzuführen sind, werden realisiert, indem der CPU **205** ermöglicht wird, ein in der Speichereinheit **211** gespeichertes Computerprogramm zu lesen. Die vorliegende Erfindung ist jedoch nicht darauf beschränkt. Die jeweiligen Verarbeitungseinheiten können zum Bei-

spiel auch vollständig oder teilweise durch eine Hardware realisiert werden, so dass jede der jeweiligen Verarbeitungseinheiten, die die Datenverarbeitungseinheit **218** bilden, durch eine Hardware, wie etwa eine elektronische Schaltung, realisiert werden kann.

[0082] Zunächst liest die CPU **205** in Schritt 701 die Dokumente durch die Scannereinheit **201** oder die Dokumente, die zum Beispiel von dem Client-PC **101** gesendet wurden und in der Speichereinheit **211** gespeichert sind, um dadurch die Bilddaten für die Vielzahl von Seiten zu erfassen. Die erfassten Bilddaten **901** bis **907** werden an die Datenverarbeitungseinheit **218** eingegeben und von der Metadaten-Verarbeitungssteuereinheit **406** an die Bereichsaufteilungseinheit **301** gesendet.

[0083] In Schritt 702 teilt die Bereichsaufteilungseinheit **301** die eingegebenen Bilddaten **901** bis **907** auf Basis einer Seiteneinheit und basierend auf den Attributen von diesen auf. Die Bilddaten **901** bis **907** werden in die jeweiligen Bereiche **908**, **910**, **911**, **912**, **913**, **915**, **917**, **918**, **919** und **920** aufgeteilt.

[0084] In Schritt 703 fügt die Attributinformationen-Hinzufügungseinheit **302** Attributinformationen zu den jeweiligen Bereichen hinzu, die in Schritt 702 aufgeteilt sind. Zum Beispiel wird in dem Fall der Bilddaten **903** für die dritte Seite der Bereich **911** mit einem Attribut „Fotografie“ versehen und wird der Bereich **912** mit einem Attribut „Beschriftung“ versehen. Die Beschriftung **912** wird auch mit Informationen versehen, die anzeigen, dass der begleitende Bereich (das begleitende Objekt) der Bereich **911** ist.

[0085] In Schritt 704 führt die Zeichenerkennungseinheit **303** eine Zeichenerkennungsverarbeitung auf den Bereichen durch, die in Schritt 703 mit einem zeichenbezogenen Attribut (z. B. Textkörper, Beschriftung, Überschrift, Zwischenüberschrift) versehen sind, und verknüpft sie das Ergebnis als Zeichenerkennungsinformationen mit den Bereichen, um dadurch das Ergebnis in der Speichereinheit **211** zu speichern. Die Zeichenerkennungsverarbeitungsergebnisse der jeweiligen Bereiche sind so, wie sie in den Plätzen/Stellen für die Zeichenerkennungsinformationen in der gemäß [Fig. 9B](#) gezeigten Tabelle beschrieben sind.

[0086] In Schritt 705 bestimmt die Datenverarbeitungseinheit **218**, ob alle Seiten den jeweiligen Verarbeitungen in Schritt 702 bis 704 unterzogen sind oder nicht. Da die Bilddaten **901** bis **907**, die den sieben Seiten entsprechen, eingegeben sind, kann die Vollendung der Verarbeitungen bezüglich aller Seiten bestimmt werden, wenn alle Informationen, wie sie in der Tabelle gemäß [Fig. 9B](#) gezeigt sind, allesamt in der Speichereinheit **211** gespeichert sind. Wenn die Vollendung der Verarbeitungen bezüglich aller Seiten bestimmt wird, schreitet die Verarbeitung zu Schritt

706 voran. Falls irgendeine Seite gefunden wird, die noch nicht den Verarbeitungen unterzogen ist, kehrt die Verarbeitung zu Schritt 702 zurück und werden die vorgenannten Verarbeitungen wiederholt.

[0087] In Schritt 706 führt die Metadaten-Verarbeitungseinheit **304** eine Verarbeitung des Extrahierens und Hinzufügens von Metadaten durch. Die Details dieser Verarbeitung werden nachstehend beschrieben.

[0088] In Schritt 707 verwendet die Formatumwandlungseinheit **305** die Entsprechungstabelle, wie sie gemäß [Fig. 5C](#) gezeigt ist, um basierend auf den jeweiligen Informationen, die in der Speichereinheit **211** gespeichert sind, die Bilddaten **901** bis **907** in elektronische Dokumente mit vorbestimmten Formaten umzuwandeln, die im Vorhinein festgelegt sind.

(Extraktion und Hinzufügung von Metadaten)

[0089] Als Nächstes wird der folgende Abschnitt die Einzelheiten der Verarbeitung des Extrahierens und Hinzufügens von Metadaten in dem vorstehend beschriebenen Schritt 706 mit Bezug auf die Ablaufdiagramme gemäß [Fig. 8A](#) und [Fig. 8B](#) beschreiben. Diese Verarbeitung besteht hauptsächlich aus den folgenden beiden Verarbeitungen. Die erste Verarbeitung ist die erste Extraktion für die Extraktion eines Ankerausdrucks und eines Beschriftungsausdrucks ([Fig. 8A](#)). Die zweite Verarbeitung ist die zweite Extraktion für die Extraktion eines Erläuterungstexts in einem Textkörper ([Fig. 8B](#)). Die Verarbeitung für die Extraktion eines Ankerausdrucks und eines Beschriftungsausdrucks wird als erstes durchgeführt. Dann wird die Verarbeitung für die Extraktion eines Erläuterungstexts in einem Textkörper durchgeführt.

[0090] Zunächst wird der folgende Abschnitt die Verarbeitung für die Extraktion eines Ankerausdrucks und eines Beschriftungsausdrucks unter Bezugnahme auf das Ablaufdiagramm gemäß [Fig. 8A](#) beschreiben.

[0091] In Schritt 801 nimmt die Metadaten-Hinzufügsziel-Auswahleinheit **401** auf Bereichsinformationen in der Speichereinheit **211** Bezug, um aus mit einem Attribut „Beschriftung“ versehenen Bereichen einen Bereich auszuwählen, der der Verarbeitung zum Extrahieren eines Ankerausdrucks und eines Beschriftungsausdrucks noch nicht unterzogen ist. Im Speziellen wird bestimmt, ob es einen noch nicht verarbeiteten Beschriftungsbereich gibt oder nicht. Falls es einen noch nicht verarbeiteten Beschriftungsbereich gibt, wird der Bereich als ein Verarbeitungsziel ausgewählt. Dann schreitet die Verarbeitung zu Schritt 802 voran. Falls es keinen Bereich gibt, der ein Attribut „Beschriftung“ aufweist, oder falls alle Bereiche bereits der Verarbeitung unterzogen sind, ist die Verarbeitung zum Extrahieren

eines Ankersausdrucks und eines Beschriftungsausdrucks vollendet. In dem Fall, in dem Bilddaten **901** bis **907** eingegeben sind, wählt die erste Routine einen Beschriftungsbereich **912** aus und wählen die anschließenden Routinen sequentiell Beschriftungsbereiche **918** und **920** aus.

[0092] In Schritt 802 fügt die Metadaten-Verarbeitungssteuereinheit **406** einen Beschriftungsbezeichner zu dem ausgewählten Objektbereich hinzu, der durch einen Beschriftungsbereich begleitet wird, und sichert sie einen Speicherbereich für Metadaten für den hinzugefügten Beschriftungsbezeichner in der Speichereinheit **211**. In dem Fall, in dem die Beschriftung **912** ausgewählt wird, wird ein Beschriftungsbezeichner „1“ zu dem durch die Beschriftung **912** begleiteten Fotografieobjekt **911** hinzugefügt und wird ein Speicherbereich in der Speichereinheit **211** gesichert, um die Metadaten für den Beschriftungsbezeichner „1“ zu speichern.

[0093] In Schritt 803 extrahiert die Anker Ausdruck-Extraktionseinheit **402** einen Anker Ausdruck und einen Beschriftungsausdruck aus den Zeichenerkennungsinformationen des ausgewählten Beschriftungsbereichs. Eine ein Objekt begleitende Beschriftung kann nur einen Anker Ausdruck, nur einen Beschriftungsausdruck oder sowohl einen Anker Ausdruck als auch einen Beschriftungsausdruck umfassen. Zum Beispiel wird ein Anker Ausdruck einer Zeichnung häufig durch eine Kombination einer speziellen Zeichenkette (Ankerzeichenkette), wie etwa „Zeichnung“, „Figur“ oder „Fig.“, und einer Nummer oder eines Symbols ausgedrückt. Daher wird im Vorhinein ein Wörterbuch erstellt, das registrierte Ankerzeichenketten als Kandidaten umfasst. Diese Kandidaten in dem Wörterbuch können mit den Zeichenerkennungsinformationen verglichen werden, um einen Anker Ausdruck (Ankerzeichenkette + Nummer/Symbol) zu identifizieren. Dann wird unter den Zeichenerkennungsinformationen des Beschriftungsbereichs eine Zeichenkette als ein Beschriftungsausdruck bestimmt, die keinen Anker Ausdruck bildet. Wenn zum Beispiel die Zeichenerkennungsinformation des Beschriftungsbereichs der Bereich **912** ist, der aus einer Zeichenkette „Fig. 1 AAA“ besteht, ist der Teil „Fig. 1“ ein Anker Ausdruck und ist der Teil „AAA“ ein Beschriftungsausdruck. Es kann einen Fall geben, in dem der Beschriftungsausdruck eine sehr kleine Anzahl von Zeichen aufweist oder keine signifikante Zeichenkette (z. B. eine Symbolkette „----“) aufweist. In einem solchen Fall besteht die Möglichkeit, dass eine Markierung, die kein Zeichen darstellt, wie etwa eine Grenze bzw. ein Rand eines Dokuments, fälschlich als eine Zeichenkette erkannt wird. Daher wird diese Markierung nicht als ein Beschriftungsausdruck extrahiert.

[0094] In Schritt 804 bestimmt die Metadaten-Verarbeitungssteuereinheit **406**, ob in Schritt 803 aus

dem Beschriftungsbereich ein Anker Ausdruck und/oder ein Beschriftungsausdruck extrahiert ist oder nicht. Im Speziellen wird bestimmt, ob ein Anker Ausdruck und ein Beschriftungsausdruck extrahiert sind oder nicht, die als Metadaten eines mit einem Beschriftungsbezeichner versehenen Objekts fungieren. Wenn ein solcher Ausdruck extrahiert ist, schreitet die Verarbeitung zu Schritt 805 voran. Wenn ein solcher Ausdruck nicht extrahiert ist, kehrt die Verarbeitung zu Schritt 801 zurück.

[0095] In Schritt 805 speichert die Metadaten-Verarbeitungssteuereinheit **406** die aus dem Beschriftungsbereich extrahierten Metadaten (einen Anker Ausdruck oder einen Beschriftungsausdruck oder sowohl einen Anker Ausdruck als auch einen Beschriftungsausdruck) in dem Speicherbereich für die Metadaten, der in Schritt 802 gesichert ist.

[0096] Durch die vorstehend beschriebenen Verarbeitungen wird zum Beispiel ein extrahierter Anker Ausdruck auf geeignete Weise über einen Beschriftungsbezeichner als Metadaten Für ein spezielles Objekt verknüpft bzw. zugeordnet.

[0097] Wenn die Verarbeitung zum Extrahieren von Metadaten durch die Extraktion eines Anker Ausdrucks und eines Beschriftungsausdrucks vollendet ist, wird eine Verarbeitung für die Extraktion eines Erläuterungstexts in einem Textkörper gestartet.

[0098] Der folgende Abschnitt wird die Verarbeitung für die Extraktion eines Erläuterungstexts in einem Textkörper unter Bezugnahme auf das Ablaufdiagramm gemäß [Fig. 8B](#) beschreiben.

[0099] In Schritt 806 wählt die Metadaten-Verarbeitungssteuereinheit **406** einen Anker Ausdruck aus, der der Verarbeitung für die Extraktion eines Erläuterungstexts in einem Textkörper noch nicht unterzogen ist. Im Speziellen bestimmt die Metadaten-Verarbeitungssteuereinheit **406**, ob es einen noch nicht verarbeiteten Anker Ausdruck gibt oder nicht. Falls es einen noch nicht verarbeiteten Anker Ausdruck gibt, wird der Anker Ausdruck ausgewählt. Dann schreitet die Verarbeitung zu Schritt 807 voran. Fall es keinen Anker Ausdruck gibt, oder falls alle Anker Ausdrücke bereits der Verarbeitung unterzogen sind, ist die Verarbeitung für die Extraktion eines Erläuterungstexts in einem Textkörper vollendet. In dem vorgenannten Beispiel wird „Fig. 1“ als ein Anker Ausdruck aus dem Beschriftungsbereich **912** extrahiert. Daher wird dieser Anker Ausdruck zunächst ausgewählt. Dann schreitet die Verarbeitung zu Schritt 807 voran.

[0100] In Schritt 807 extrahiert die Metadaten-Verarbeitungssteuereinheit **406** einen Teil, der den extrahierten Anker Ausdruck (Satz) umfasst, aus den Zeichenerkennungsinformationen für den Textkörperbereich, die in der Speichereinheit **211** gespeichert sind,

um den extrahierten Erläuterungstext in dem Textkörper als Metadaten mit dem Objekt zu verknüpfen. Bei dem vorgenannten Beispiel wird „Fig. 1“ als ein Anker Ausdruck extrahiert. Daher umfasst der Textkörperbereich **908** der Bilddaten **901** ein Wort „Fig. 1“, das durch das Bezugszeichen **916** gezeigt ist. Da dieses Wort identisch zu dem Anker Ausdruck „Fig. 1“ ist, wird der Textkörperbereich **908** als ein Erläuterungstext in einem Textkörper extrahiert und als die Metadaten des Fotografieobjekts **911** verknüpft bzw. zugeordnet.

[0101] Der extrahierte Textkörperbereich als ein Erläuterungstext in einem Textkörper kann weiter auch einer Analyse unterzogen werden, um dadurch Metadaten zu extrahieren. Zum Beispiel kann eine morphologische Analyse für eine Verarbeitung natürlicher Sprache zum Beispiel verwendet werden, um Wörter auszusondern, oder können die resultierenden Wortklassen oder wichtigen Wörter zum Beispiel bestimmt werden, um dadurch das Wort „Kamera“ in dem Textkörperbereich **908** als Metadaten zu extrahieren.

[0102] Es besteht die Möglichkeit, dass eine Vielzahl von Erläuterungstexten in einem Textkörper extrahiert werden. Es kann auch die Möglichkeit bestehen, dass Metadaten ferner aus einem extrahierten Erläuterungstext in einem Textkörper erhalten werden. Somit kann dieser Schritt eine Vielzahl von Metadaten(-stücken) extrahieren. Wenn die Verarbeitung für die Extraktion eines Erläuterungstexts in einem Textkörper vollendet ist, schreitet die Verarbeitung zu Schritt 808 voran.

[0103] In Schritt 808 bestimmt die Metadaten-Verarbeitungssteuereinheit **406**, ob ein Erläuterungstext in einem Textkörper, der als Metadaten eines mit einem Beschriftungsbezeichner versehenen Objektbereichs fungiert, extrahiert ist oder nicht. Falls ein derartiger Erläuterungstext in einem Körper extrahiert ist, schreitet die Verarbeitung zu Schritt 809 voran. Falls ein derartiger Erläuterungstext in einem Körper nicht extrahiert ist, kehrt die Verarbeitung zu Schritt 806 zurück und wird die Verarbeitung bezüglich eines weiteren noch nicht verarbeiteten Anker Ausdrucks wiederholt.

[0104] In Schritt 809 speichert die Metadaten-Verarbeitungssteuereinheit **406** den extrahierten Erläuterungstext in einem Textkörper in dem in Schritt 802 gesicherten Speicherbereich für Metadaten (oder fügt sie diesen dort hinzu). Nach der Speicherung kehrt die Verarbeitung zu Schritt 806 zurück, um die Verarbeitung bezüglich eines weiteren noch nicht verarbeiteten Anker Ausdrucks fortzusetzen.

[0105] Durch die Verarbeitungen, wie sie vorstehend beschriebenen sind, wird der extrahierte Erläuterungstext in einem Textkörper über einen Beschriftungsbezeichner auf geeignete Weise als Metadaten für ein spezielles Objekt verknüpft bzw. zugeordnet.

tungsbezeichner auf geeignete Weise als Metadaten für ein spezielles Objekt verknüpft bzw. zugeordnet.

[0106] Wenn die Verarbeitung bezüglich aller Anker Ausdrücke vollendet ist, sammelt die Metadaten-Sammel-/Ausgabereinheit **405** die resultierenden Metadaten, und fügt sie Metadaten in der Form von Bilddaten, die durch die Formatumwandlungseinheit **305** empfangen werden können, zu Bilddaten hinzu. Dann werden die mit den Metadaten versehenen Bilddaten an die Formatumwandlungseinheit **305** gesendet.

(Verknüpfung zwischen einem Erläuterungstext in einem Textkörper und einem Objekt)

[0107] Als Nächstes wird der folgende Abschnitt die Einzelheiten der Verarbeitung in Schritt 807 gemäß **Fig. 8B** unter Bezugnahme auf das Ablaufdiagramm gemäß **Fig. 10** beschreiben.

[0108] In Schritt 1001 erfasst die Metadaten-Verarbeitungssteuereinheit **406** die Anzahl von Objekten mit der Beschriftung, die den in Schritt 806 gemäß **Fig. 8B** ausgewählten speziellen Anker Ausdruck umfasst. Es wird ein Fall angenommen, bei dem zum Beispiel ein Anker Ausdruck „Fig. 2“ für die Bilddaten **901** bis **907** ausgewählt ist. In diesem Fall gibt es kein anderes Objekt, das „Fig. 2“ umfasst, als das Fotografieobjekt **919**. Daher wird „1“ als die Anzahl von Objekt(en) erfasst. Wenn andererseits ein Anker Ausdruck „Fig. 1“ ausgewählt ist, existieren das Fotografieobjekt **911** und das Fotografieobjekt **917** als Objekte mit einer Beschriftung, die „Fig. 1“ umfasst. Daher wird „2“ als die Anzahl von Objekt(en) ausgewählt.

[0109] In Schritt 1002 bestimmt die Metadaten-Verarbeitungssteuereinheit **406**, ob die Anzahl von Objekten, die in Schritt 1001 erfasst ist, eine Mehrzahl ist oder nicht. Wenn die Anzahl der erfassten Objekte „1“ ist (d. h., wenn es kein anderes Objekt mehr gibt, das eine Beschriftung aufweist, das einen identischen Anker Ausdruck umfasst), schreitet die Verarbeitung zu Schritt 1003 voran. Wenn die Anzahl der erfassten Objekte eine Mehrzahl ist (d. h., wenn Beschriftungen einer Vielzahl von Objekten einen identischen Anker Ausdruck verwenden), schreitet die Verarbeitung zu Schritt 1005 voran.

[0110] In Schritt 1003 sucht die textkörperinterne Sucheinheit **403** den speziellen Anker Ausdruck in den Zeichenerkennungsinformationen des Textkörperbereichs. In dem Fall des vorstehenden Beispiels wird der Textkörperbereich **915**, der den Anker Ausdruck **916**, nämlich „Fig. 2“, umfasst, als ein Erläuterungstext in einem Textkörper extrahiert. Wenn ein den Anker Ausdruck umfassender Ausdruck in den Zeichenerkennungsinformationen des Textkörperbereichs gefunden wird, schreitet die Verarbeitung zu Schritt 1004 voran.

[0111] In Schritt 1004 verknüpft die Metadaten-Verarbeitungssteuereinheit **406** den Ausdruck in dem Textkörper, der den speziellen Anker Ausdruck umfasst, (einen Erläuterungstext in einem Textkörper) mit dem Objekt. In dem Fall des vorstehenden Beispiels wird der Textkörperbereich **915**, der als ein Erläuterungstext in einem Textkörper extrahiert ist, der den Anker Ausdruck „**Fig. 2**“ umfasst, mit dem Fotografieobjekt **919** verknüpft.

[0112] In Schritt 1005 sucht die textkörperinterne Sucheinheit **403**, wie in Schritt 1003, den speziellen Anker Ausdruck in den Zeichenerkennungsinformationen des Textkörperbereichs. In dem Fall des vorstehenden Beispiels werden die Textkörperbereiche **908** und **913** als Erläuterungstexte in einem Textkörper gefunden, der den Anker Ausdruck „**Fig. 1**“ umfasst. Wenn ein Erläuterungstext in einem Textkörper, der den speziellen Anker Ausdruck umfasst, in den Zeichenerkennungsinformationen des Textkörperbereichs gefunden wird, wie es vorstehend beschrieben ist, schreitet die Verarbeitung zu Schritt 1006 voran.

[0113] In Schritt 1006 erstellt die Metadaten-Verarbeitungssteuereinheit **406**, mit Bezug auf den gefundenen Erläuterungstext in einem Textkörper (Textkörperbereich), eine erforderliche Anzahl von Kombinationen zum Auffinden einer optimalen Entsprechungsbeziehung mit einer Beschriftung eines möglichen Objektbereichs. Dann wird der Ähnlichkeitsgrad mit Bezug auf die einzelnen Kombinationen berechnet. In dem Fall des vorstehenden Beispiels werden zum Beispiel zwei Kombinationen zwischen den Beschriftungen **912** und **918**, die den identischen Anker Ausdruck „**Fig. 1**“ umfassen, und den Textkörperbereichen **908** und **913**, die den Anker Ausdruck umfassen, erstellt. Im Speziellen wird in diesem Fall, falls eine Beschriftung bestimmt wird, die einem Erläuterungstext in einem Textkörper entspricht, die andere ebenfalls bestimmt. Somit werden zwei Kombinationen erstellt, nämlich eine Kombination der Beschriftung **912** und des Textkörperbereichs **908** und eine Kombination der Beschriftung **918** und des Textkörperbereichs **908**. Alternativ können auch zwei andere entsprechende Kombinationen erstellt werden, nämlich eine Kombination der Beschriftung **912** und des Textkörperbereichs **913** und eine Kombination der Beschriftung **918** und des Textkörperbereichs **913**. Falls es drei Beschriftungen (Objekte) gibt, die einen identischen Anker Ausdruck umfassen, und drei Erläuterungstexte in einem Textkörper gefunden werden, die den Beschriftungen entsprechen, werden zumindest fünf Kombinationen berechnet.

[0114] Die Einzelheiten der Berechnung des Ähnlichkeitsgrads werden nachstehend beschrieben. Nach der Vollendung der Berechnung des Ähnlichkeitsgrads schreitet die Verarbeitung zu Schritt 1007 voran.

[0115] In Schritt 1007 nimmt die Metadaten-Verarbeitungssteuereinheit **406** basierend auf dem Berechnungsergebnis von Schritt 1006 an, dass eine Kombination mit dem höchsten Wert des Ähnlichkeitsgrads optimal ist. Dann verknüpft die Metadaten-Verarbeitungssteuereinheit **406** einen Erläuterungstext in einem Textkörper in der Kombination mit dem Objektbereich.

(Verarbeitung zum Berechnen
des Ähnlichkeitsgrads)

[0116] Die Berechnung des Ähnlichkeitsgrads in Schritt 1006 gemäß **Fig. 10** wird unter Bezugnahme auf das Ablaufdiagramm gemäß **Fig. 11** mit Bezug auf das Beispiel der vorstehend beschriebenen beiden Kombinationen (der ersten Kombination einer Beschriftung **912** und eines Textkörperbereichs **908** und der zweiten Kombination der Beschriftung **918** und des Textkörperbereichs **908**) beschrieben.

[0117] In Schritt 1101 erfasst die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** einen Beschriftungsausdruck aus der Beschriftung der Kombination von Berechnungszielen. Wenn die erste Kombination ein Berechnungsziel ist, wird „AAA“, was durch Entfernen des Anker Ausdrucks „**Fig. 1**“ aus der Beschriftung **912** erhalten wird, als ein Beschriftungsausdruck erfasst. Nach der Erfassung schreitet die Verarbeitung zu Schritt 1102 voran.

[0118] In Schritt 1102 unterzieht die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** den erfassten Beschriftungsausdruck einer Wortabsonderung durch eine morphologische Analyse, um dadurch die Wortklasseninformationen bezüglich der jeweiligen Wörter zu erhalten. Dann wird basierend auf den erhaltenen Wortklasseninformationen ein (hierin nachstehend als „Beschriftungswort“ bezeichnetes) Wort mit einer Wortklasse, wie etwa ein Nomen, als ein Vergleichsziel ausgewählt. Im Speziellen wird ein Wort oder ein Anker Ausdruck, das/der nicht wichtig ist, von Vergleichszielen ausgeschlossen, so dass ein derartiges Wort oder ein derartiger Anker Ausdruck nicht ausgewählt wird. Zum Beispiel stellt der Beschriftungsausdruck „AAA“ ein Nomen „AAA“ bereit. Es werden ein Beschriftungswort „AAA“ und eine Wortklasseninformation „Nomen“ erhalten. Als Ergebnis wird „AAA“ als ein Vergleichszielkandidat ausgewählt. In diesem Schritt können eine Vielzahl von Beschriftungswörtern ausgewählt werden.

[0119] In Schritt 1103 stellt die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** die Gewichtung für das Beschriftungswort auf 1 ein. Im Speziellen wird bei diesem Ausführungsbeispiel der Gewichtungswert zum Beispiel nicht basierend auf dem Abstand vor dem Anker Ausdruck oder der Wortklasse oder einer Charakteristik des Worts berechnet. Der durch die vorstehend beschriebene Berechnung er-

haltene Wert kann jedoch auch als eine Gewichtung verwendet werden, wenn der Beschriftungsausdruck ein Satz mit einer bestimmten Länge ist, oder wenn die Wichtigkeit des Worts zum Beispiel basierend auf einer eindeutigen Ausdrucksextraktion oder einem Wörterbuch berechnet wird.

[0120] **Fig. 12A** veranschaulicht durch eine Tabelle das Ergebnis der ersten Kombination, das durch die vorstehend beschriebenen Verarbeitungen erhalten wird. In dieser Tabelle sind der Beschriftungsausdruck und das Beschriftungswort ist das Beschriftungswortattribut „Nomen“ und ist ein Vergleichszielkandidat ausgewählt. Die Abstandsinformation von dem Anker Ausdruck ist auf eine Information „-“ eingestellt, die anzeigt, dass diese Information nicht verwendet wird. Die Gewichtung ist auf „1“ eingestellt.

[0121] Als Nächstes erfasst die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** in Schritt 1104 einen Erläuterungstext in einem Textkörper (Textkörperbereich) in der Kombination von Berechnungszielen. In dem Fall der ersten Kombination wird der Textkörperbereich **908** als der Erläuterungstext in einem Textkörper erfasst, nämlich „Kamera AAA (**Fig. 1**)“ dient z. B. zum Fotografieren einer Landschaft“, der den Anker Ausdruck **909** umfasst, nämlich „**Fig. 1**“.

[0122] In Schritt 1105 unterzieht die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** den erfassten Erläuterungstext in dem Textkörper einer Wortabsonderung durch eine morphologische Analyse, um dadurch die Wortklasseninformationen für die jeweiligen Wörter zu erhalten. Dann wird basierend auf den erhaltenen Wortklasseninformationen ein (hierin nachstehend als „Erläuterungstextwort“ bezeichnetes) Wort mit einer Wortklasse, wie etwa ein Nomen, als ein Vergleichsziel ausgewählt. Diese Auswahl wird auch so durchgeführt, dass ein nicht wichtiges Wort oder ein nicht wichtiger Anker Ausdruck von Vergleichszielen ausgeschlossen und nicht ausgewählt wird. In diesem Schritt werden eine Vielzahl von Erläuterungstextwörtern ausgewählt.

[0123] In Schritt 1106 ordnet die Ausdrucksähnlichkeitsgrad-Vergleichseinheit **404** die ausgewählten Erläuterungstextwörter in einer Reihenfolge von denjenigen mit einer näheren Position in dem Erläuterungstext in dem Textkörper zu dem Anker Ausdruck an (d. h. in einer Reihenfolge derjenigen mit einem kürzeren Abstand (Wortabstand) von dem Anker Ausdruck). Der Abstand von dem Anker Ausdruck wird einfach als die Anzahl von Wörtern von dem Anker Ausdruck zu dem Erläuterungstextwort eingestellt. Nachstehend wird hierin eine Spalte von Erläuterungstextwörtern, die in einer Reihenfolge des Wortabstands angeordnet sind, als eine „Erläuterungstextwortkette“ bezeichnet. Dann wird die Anzahl von in dieser Erläuterungstextwortkette umfassten Erläuterungstextwörtern als der Wert der Variablen N eingestellt.

terungstextwörtern als der Wert der Variablen N eingestellt.

[0124] **Fig. 12B** veranschaulicht durch eine Tabelle das Ergebnis der Verarbeitungen von Schritten 1105 und 1106 auf dem erfassten Erläuterungstext in dem Textkörper, nämlich „Kamera AAA (**Fig. 1**)“ dient z. B. zum Fotografieren einer Landschaft“. Die drei Erläuterungstextwörter werden als Vergleichszielkandidaten ausgewählt und mit Werten „3“, „2“ und „1“ als eine Reihenfolge in der Erläuterungstextwortkette versehen, so dass diese Wörter in einer Reihenfolge derjenigen mit einem kürzeren Abstand zu dem Anker Ausdruck „**Fig. 1**“ extrahiert werden können (d. h. in einer Reihenfolge von „AAA“, „Kamera“ und „Landschaft“). In diesem Fall beträgt die Anzahl der in der Erläuterungstextwortkette umfassten Erläuterungstextwörter drei. Somit wird der Wert der Variablen N auf „3“ eingestellt.

[0125] In Schritt 1107 initialisiert die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** den Wert der Variablen S, die den Ähnlichkeitsgrad zeigt. Im Speziellen stellt die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** den vorbestimmten Wert als den Ähnlichkeitsgrad S auf „0“ ein.

[0126] In Schritt 1108 bestimmt die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404**, ob der Wert der Variablen N größer ist als „0“ oder nicht. Wenn die Variable N größer ist als „0“, bedeutet dies, dass es ein Erläuterungstextwort gibt, das ein noch nicht verarbeitetes Vergleichsziel darstellt. Somit schreitet die Verarbeitung zu Schritt 1109 voran. Wenn der Wert der Variablen N gleich „0“ ist, bedeutet dies, dass alle Erläuterungstextwörter als Vergleichsziel bereits der Verarbeitung unterzogen sind. Somit wird eine Routine durchgeführt, die aus Schritt 1108 bis 1113 besteht, und wird die Verarbeitung vollendet.

[0127] In Schritt 1109 extrahiert die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** aus der Erläuterungstextwortkette ein derartiges Erläuterungstextwort, das eine Ordnungszahl bzw. Reihenfolgenummer in der Erläuterungstextwortkette aufweist, die den gleichen Wert wie denjenigen der Variablen N aufweist. In dem Fall des vorstehenden Beispiels wird als erstes das Erläuterungstextwort „AAA“ mit „3“ als der Ordnungszahl bzw. Reihenfolgenummer der Erläuterungstextwortkette extrahiert.

[0128] In Schritt 1110 berechnet die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** die Gewichtung des Erläuterungstextworts basierend auf dem Wortabstand, der den Abstand von dem Anker Ausdruck zeigt. Die Gewichtung weist einen Wert auf, der abhängig von dem Wortabstand kleiner wird und „1“ als den höchsten Wert und „0“ als den niedrigsten Wert aufweist. Die Gewichtung ist bei diesem Beispiel ein Kehrwert des Wortabstands. In dem Fall des

Erläuterungstextworts „AAA“ ist der Wortabstand von diesem „2“. Somit wird „0,5“ als eine Gewichtung berechnet. Die Gewichtung kann auch durch verschiedene andere Verfahren als das vorgenannte Verfahren erhalten werden, die zum Beispiel ein Verfahren des Abschwächens der Gewichtung in einer linearen Art und Weise, ein Verfahren der Verwendung eines Logarithmus oder ein Verfahren der Verwendung einer Verteilungsfunktion umfassen. Es kann jedoch auch jedes beliebige Verfahren verwendet werden, solange das Verfahren die Dämpfung bzw. Verkleinerung der Gewichtung abhängig von dem Abstand bereitstellt.

[0129] In Schritt 1111 vergleicht die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** das ausgewählte Erläuterungstextwort mit dem Beschriftungswort, um den Ähnlichkeitsgrad S zu berechnen. Im Speziellen, wie es in der folgenden Formel 1 gezeigt ist, wird der Ähnlichkeitsgrad S, wenn das erstere Wort und das letztere Wort die gleiche Zeichenkette aufweisen, durch Multiplikation von „1“ mit den Gewichtungen des ersteren und des letzteren Worts berechnet. Wenn das erstere Wort und das letztere Wort unterschiedliche Zeichenketten aufweisen, wird der Ähnlichkeitsgrad S durch Multiplikation von „0“ mit den Gewichtungen des ersteren und des letzteren Worts berechnet.

Ähnlichkeitsgrad S = (1
oder 0) × (Gewichtung von
Beschriftungswort) × (Gewichtung von
Erläuterungstextwort) (Formel 1)

[0130] Auf diese Art und Weise wird der Ähnlichkeitsgrad S mit Bezug auf das Erläuterungstextwort als ein Vergleichsziel berechnet. Wenn es eine Vielzahl von Beschriftungswörtern gibt, werden die Beschriftungswörter einzeln mit einem ausgewählten Erläuterungstextwort verglichen. Dann wird der Ähnlichkeitsgrad S mit dem höchsten Wert unter den berechneten Ähnlichkeitsgraden S als S_{max} bestimmt.

[0131] Bei diesem Ausführungsbeispiel wird der Ähnlichkeitsgrad S durch eine einfache Berechnung erhalten. Der Ähnlichkeitsgrad S kann jedoch auch zum Beispiel durch einen solchen Vergleich erhalten werden, der einen Thesaurus bzw. einen Wortschatz, eine inkonsistente Notation oder eine Notation, wie etwa Hiragana, Katakana, oder Kanji, berücksichtigt. Der Ähnlichkeitsgrad S kann auch durch einen Vergleich unter Verwendung eines Thesaurus bzw. Wortschatzes oder einem Vergleich unter Verwendung einer Vergleichssprache zum Bestimmen der Vieldeutigkeit basierend auf dem Kontext erhalten werden. Der Ähnlichkeitsgrad S kann zum Beispiel auch durch Berücksichtigung der Einbeziehungsrate einer Zeichenkette oder einer Rate erhalten werden, mit der das gleiche Zeichen gemischt wird.

[0132] Als Nächstes addiert die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** in Schritt 1112 den Wert des Ähnlichkeitsgrads S (oder S_{max}), der in Schritt 1111 berechnet wird, zu den Werten des vorhergehenden Ähnlichkeitsgrads S, um dadurch den Wert des Ähnlichkeitsgrads S zu aktualisieren. In dem Fall der ersten Routine wird zum Beispiel der Anfangswert auf „0“ eingestellt. Somit wird der in Schritt 1111 berechnete Wert des Ähnlichkeitsgrads S direkt als ein aktualisierter Ähnlichkeitsgrad S verwendet. In dem Fall der zweiten Routine und der anschließenden Routinen wird der durch die vorhergehende Routine bzw. die vorhergehenden Routinen erhaltene Wert des Ähnlichkeitsgrads S mit dem Wert des zuletzt bestimmten Ähnlichkeitsgrads S (oder S_{max}) addiert.

[0133] In Schritt 1113 zieht die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** 1 (eins) von dem Wert der Variablen N ab. Dann kehrt die Verarbeitung zu Schritt 1108 zurück.

[0134] In dem Beispiel gemäß [Fig. 12B](#) gibt es drei Erläuterungstextwörter als Vergleichszielkandidat. Somit wird die vorgenannte Routine drei Mal wiederholt.

[0135] In der ersten Routine wird der Wert der Variablen N auf „3“ eingestellt. Somit wird die Bestimmung in Schritt 1108 von Schritt 1109 gefolgt. In Schritten 1109 und 1110 wird die Extraktion von „AAA“ von der Berechnung von „0,5“ als eine Gewichtung für das Wort gefolgt, wie es vorstehend beschrieben ist. Dann wird in Schritt 1111 eine Verarbeitung durchgeführt, um den Ähnlichkeitsgrad S zwischen dem Erläuterungstextwort „AAA“ und dem Beschriftungswort „AAA“ zu berechnen. In diesem Fall weisen das erstere und das letztere Wort die gleiche Zeichenkette auf. Daher wird der Ähnlichkeitsgrad S als $1 \times 1 \times 0,5$ berechnet.

[0136] Da es nur ein Beschriftungswort „AAA“ gibt, wird der resultierende Ähnlichkeitsgrad S = 0,5 in Schritt 1112 zu dem Anfangswert „0“ addiert. Dann schreitet die Verarbeitung zu Schritt 1113 voran. In Schritt 1113 wird 1 (eins) von dem Wert der Variablen N abgezogen, was in N = „2“ resultiert. Dann kehrt die Verarbeitung zu Schritt 1108 zurück. Daraufhin wird die gleiche Verarbeitung zwei Mal wiederholt. Der Ähnlichkeitsgrad S wird auch zwischen den Erläuterungstextwörtern „Kamera“ und „Landschaft“ berechnet, und dann wird der Ähnlichkeitsgrad S stets aktualisiert. Da die Erläuterungstextwörter „Kamera“ und „Landschaft“ sich jedoch von dem einzigen Beschriftungswort „AAA“ unterscheiden, beträgt der in Schritt 1111 berechnete Wert des Ähnlichkeitsgrads S für die beiden Erläuterungstextwörter „Kamera“ und „Landschaft“ gleich „0“. Daher beträgt der endgültig erhaltene Wert des Ähnlichkeitsgrads S gleich $0,5 + 0 + 0 = 0,5$.

[0137] Wie es vorstehend beschrieben ist, wurde die Verarbeitung für die erste Kombination beschrieben. Die ähnliche Verarbeitung wird auch bezüglich der zweiten Kombination durchgeführt, die aus der Beschriftung **918** und dem Textkörperbereich **908** besteht. In dem Fall der zweiten Kombination ist „BBB“ der aus der Beschriftung **918** erhaltene Beschriftungsausdruck. Somit unterscheidet sich der Beschriftungsausdruck von dem Erläuterungstextwort. Somit hat der endgültig erhaltene Ähnlichkeitsgrad *S* einen Wert von 0. Als Ergebnis nimmt in Schritt 1007 gemäß [Fig. 10](#) die Metadaten-Verarbeitungssteuereinheit **406** an, dass die erste Kombination mit dem Ähnlichkeitsgrad *S* von 0,5 eine optimale Kombination ist. Im Speziellen wird das Fotografieobjekt **911** mit der Beschriftung **912** gemäß der ersten Kombination mit dem Erläuterungstext in dem Textkörper, nämlich „Kamera AAA ([Fig. 1](#))“ dient z. B. zum Fotografieren einer Landschaft“ (dem Textkörperbereich **908**), verknüpft.

[0138] Anstelle des in dem Ablaufdiagramm gemäß [Fig. 11](#) gezeigten Verfahrens können auch andere Verfahren verwendet werden, wie etwa zum Beispiel ein Verfahren des Verwendens einer Rate, mit der das gleiche Zeichen verwendet wird, oder eines Grads, mit dem das gleiche Zeichen kontinuierlich verwendet wird, oder ein Verfahren des Extrahierens ausschließlich eines wichtigen Ausdrucks, um dadurch einen speziellen Ausdruck (eine spezielle Zeichenkette) zum Vergleich zu erhalten.

[0139] [Fig. 13](#) veranschaulicht durch eine Tabelle die Metadateninformationen, die durch die vorstehend beschriebene Verarbeitung erhalten werden, die auf den Bilddaten **901** bis **907** durchgeführt wird.

[0140] Wie es durch das Bezugszeichen **1301** gezeigt ist, sind zum Beispiel der Erläuterungstext in dem Textkörper des Textkörperbereichs **908**, die Beschriftung **912** und der Ankerausdruck hiervon, nämlich „[Fig. 1](#)“, durch den Beschriftungsbezeichner „1“ auf geeignete Weise mit dem Fotografieobjekt **911** verknüpft. Wie es durch das Bezugszeichen **1302** gezeigt ist, sind der Erläuterungstext in dem Textkörper des Textkörperbereichs **913**, die Beschriftung **918** und der Ankerausdruck hiervon, nämlich „[Fig. 1](#)“, durch den Beschriftungsbezeichner „2“ auf geeignete Weise mit dem Fotografieobjekt **917** verknüpft. Wie es durch das Bezugszeichen **1303** gezeigt ist, sind der Erläuterungstext des Textkörpers des Textkörperbereichs **915**, die Beschriftung **920** und der Ankerausdruck hiervon, nämlich „[Fig. 2](#)“, durch den Beschriftungsbezeichner „3“ auf geeignete Weise mit dem Fotografieobjekt **919** verknüpft.

[0141] Die in der Tabelle gemäß [Fig. 13](#) gezeigten Metadateninformationen sind in einem realen Fall zum Beispiel durch das SVG-Format beschrieben. Die Metadateninformationen können durch das glei-

che Verfahren wie das für [Fig. 6A](#) beschriebene beschrieben werden. Im Speziellen ist gemäß [Fig. 6A](#) der Beschriftungsbezeichner **608** zu den Objektdaten **603** hinzugefügt und ist der gleiche Bezeichner als der Beschriftungsbezeichner **609** zu den Metadaten **607** des Objekts hinzugefügt. In ähnlicher Weise ist gemäß [Fig. 13](#) der gleiche Bezeichner als der Beschriftungsbezeichner (1 bis 3), der zu den jeweiligen Objekten hinzugefügt ist, zu den den jeweiligen Objekten entsprechenden Metadaten hinzugefügt, so dass jedem Objekt entsprechende Metadaten identifiziert werden können.

[0142] Wie es vorstehend beschrieben ist, kann bei der Bildverarbeitungsvorrichtung gemäß diesem Ausführungsbeispiel ein Objekt in einem Dokument, wie etwa eine Fotografie, eine Zeichnung oder eine Tabelle, auf geeignete Weise mit einem Erläuterungstext in einem Textkörper, der die Inhalte von diesem beschreibt, verknüpft werden. Als Ergebnis können selbst dann, wenn ein spezieller Ankerausdruck in den Beschriftungen in unterschiedlichen Objekten unterschiedlicher Dokumente gemeinsam verwendet wird, elektronische Dokumentdaten, die mit geeigneten Metadaten versehen sind, mit Bezug auf die jeweiligen Objekte erstellt werden. Außerdem können Metadaten extrahiert werden, während der Einfluss durch die Seitengestaltung oder den Abstand zwischen Seiten erheblich verringert wird. Somit können geeignete Metadaten selbst aus Bilddaten heraus hinzugefügt werden, die zufällig angeordnete Seiten aufweisen.

[Ausführungsbeispiel 2]

[0143] Als Nächstes wird der folgende Abschnitt Ausführungsbeispiel 2 der Verarbeitung in Schritt 807 gemäß [Fig. 8B](#) unter Bezugnahme auf [Fig. 14A](#), [Fig. 14B](#) und [Fig. 15](#) beschreiben. Ausführungsbeispiel 2 ist ein Ausführungsbeispiel, das die Effizienz der Verarbeitung zum Extrahieren eines Erläuterungstexts in einem Textkörper, der einen ausgewählten Ankerausdruck umfasst, um den extrahierten Text mit einem Objektbereich zu verknüpfen, verbessern kann.

[0144] [Fig. 14A](#) und [Fig. 14B](#) sind Ablaufdiagramme, die die Verarbeitung zum Verknüpfen eines Erläuterungstexts in einem Textkörper mit einem Objektbereich bei diesem Ausführungsbeispiel veranschaulichen. Die gleichen Teile wie diejenigen des Ablaufdiagramms von [Fig. 10](#) gemäß Ausführungsbeispiel 1 werden einfach/vereinfacht oder nicht weiter beschrieben. Daher wird hauptsächlich ein Unterscheidungspunkt beschrieben.

[0145] In Schritt 1401 wird die Anzahl von Objekten erfasst, die eine Beschriftung aufweisen, die einen speziellen Ankerausdruck umfasst. Wenn die Anzahl der erfassten Objekte in Schritt 1402 eine Mehr-

zahl ist, schreitet die Verarbeitung zu Schritt 1405 voran. Wenn die Anzahl des Objekts „1“ ist, sind die anschließenden Verarbeitungen (S1403 und S1404) die gleichen wie Schritte S1003 und S1004 gemäß [Fig. 10](#).

[0146] In Schritt 1405 sucht die Metadaten-Verarbeitungssteuereinheit **406** den Ankerausdruck in dem Textkörperbereich. In Schritt 1406 bestimmt die Metadaten-Verarbeitungssteuereinheit **406**, ob der resultierende Erläuterungstext in dem Textkörper auf der gleichen Seite wie derjenigen des Objekts, wie etwa einer Fotografie, liegt. In dem Fall von Bilddaten für die jeweiligen Seiten, die durch Lesen dieser Seiten durch einen Scanner erhalten werden, ist es selten, dass Daten einer Seite mit Bilddaten eines anderen Dokuments gemischt sind. Wenn ein Objekt, wie etwa eine Fotografie, auf der gleichen Seite wie derjenigen eines Erläuterungstexts in einem Textkörper liegt, der durch die Suche gefunden wird, wird daher erachtet, dass das Objekt und der Erläuterungstext eine Entsprechungsbeziehung zwischen sich aufweisen. Diese Bestimmungsverarbeitung wird daher so durchgeführt, dass die Verknüpfung ohne Berechnung des Ähnlichkeitsgrads durchgeführt werden kann. Wenn das Objekt, wie etwa eine Fotografie, als auf der gleichen Seite wie derjenigen des Erläuterungstexts in einem Textkörper liegend bewertet wird, schreitet die Verarbeitung zu Schritt 1407 voran. In Schritt 1407 verknüpft die Metadaten-Verarbeitungssteuereinheit **406** den Erläuterungstext in einem Textkörper, der den Ankerausdruck umfasst, mit dem Objektbereich auf der gleichen Seite. Daraufhin schreitet die Verarbeitung zu Schritt 1409 voran. Wenn das Objekt, wie etwa eine Fotografie, als nicht auf der gleichen Seite wie derjenigen des Erläuterungstexts in dem Textkörper liegend bewertet wird, schreitet die Verarbeitung zu Schritt 1408 voran.

[0147] In Schritt 1408 bestimmt die Metadaten-Verarbeitungssteuereinheit **406**, ob andere Erläuterungstexte in einem Textkörper vorhanden sind oder nicht, der durch die Suche gefunden wird. Im Speziellen bestimmt die Metadaten-Verarbeitungssteuereinheit **406**, ob es eine Kombination eines Objektbereichs und eines Erläuterungstexts in einem Textkörper gibt, für die der Ähnlichkeitsgrad berechnet werden muss. Falls bestimmt wird, dass es keinen Erläuterungstext mehr in einem Textkörper gibt, wird diese Verarbeitung vollendet. Falls es einen weiteren Erläuterungstext in einem Textkörper gibt, schreitet die Verarbeitung zu Schritt 1409 voran.

[0148] In Schritt 1409 bestimmt die Metadaten-Verarbeitungssteuereinheit **406**, ob die in Schritt 1401 erfasste Anzahl der Objekte 3 oder mehr ist oder nicht. Wenn die Anzahl der Objekte 2 ist, schreitet die Verarbeitung zu Schritt 1410 voran. Wenn die Anzahl von Objekten 3 oder mehr ist, schreitet die Verarbeitung

zu Schritt 1413 voran. Die Verarbeitungen in Schritt 1413 und Schritt 1414, wenn die Anzahl der Objekte 3 oder mehr ist, sind gleich denjenigen von Schritt 1006 und Schritt 1007 gemäß [Fig. 10](#).

[0149] In Schritt 1410 unterzieht die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** die Kombination von einem Objektbereich und dem Erläuterungstext in einem Textkörper der vorstehend beschriebenen Verarbeitung des Ablaufdiagramms gemäß [Fig. 11](#), um dadurch den Ähnlichkeitsgrad S zu berechnen.

[0150] Als Nächstes verwendet die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** in Schritt 1411 den in Schritt 1410 erhaltenen Ähnlichkeitsgrad S, um eine Kombination des anderen Objektbereichs und des Erläuterungstexts in einem Textkörper den durch das Ablaufdiagramm gemäß [Fig. 15](#) gezeigten Verarbeitungen zu unterziehen. Im Speziellen werden die folgenden Verarbeitungen durchgeführt.

[0151] Zunächst erfasst die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** in Schritt 1501 den in Schritt 1410 erhaltenen Ähnlichkeitsgrad S, um diesen Ähnlichkeitsgrad S in dem RAM **206** als den Ähnlichkeitsgrad Scomp festzuhalten, der in nachstehend beschriebenem Schritt 1514 als ein Vergleichsziel verwendet wird. Dann schreitet die Verarbeitung zu Schritt 1502 voran.

[0152] In Schritt 1502 bis Schritt 1513 werden die gleichen Verarbeitungen wie diejenigen von Schritt 1101 bis Schritt 1112 gemäß [Fig. 11](#) durchgeführt. Wenn eine Verarbeitung für die erste Aktualisierung des Ähnlichkeitsgrads S in Schritt 1513 durchgeführt wird, schreitet die Verarbeitung zu Schritt 1514 voran.

[0153] In Schritt 1514 vergleicht die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** den in Schritt 1501 erfassten und festgehaltenen Wert Scomp mit dem in Schritt 1513 aktualisierten Wert des Ähnlichkeitsgrads S, um zu bestimmen, welcher der beiden Werte höher ist. Wenn der Wert des aktualisierten Ähnlichkeitsgrads S höher ist als Scomp, wird die Verarbeitung vollendet. Dann schreitet die Verarbeitung zu Schritt 1412 voran. Der Grund dafür besteht darin, dass es festgelegt ist, dass der erfasste Ähnlichkeitsgrad S (Scomp) niedriger ist als der Ähnlichkeitsgrad S gemäß Schritt 1410. Wenn der Wert des aktualisierten Ähnlichkeitsgrads S kleiner ist als der Wert Scomp, schreitet die Verarbeitung zu Schritt 1515 voran, um die zweite Routine durchzuführen, um den Ähnlichkeitsgrad S zu berechnen.

[0154] Wenn der Wert der Variablen N in der zweiten Routine und den anschließenden Routinen gleich 0 ist, ohne zu bestimmen, dass der Wert des aktualisierten Ähnlichkeitsgrads S höher ist als Scomp, wird die Verarbeitung vollendet. Dann schreitet die Verar-

beitung zu Schritt 1412 voran. Zu diesem Zeitpunkt ist festgelegt, dass der erfasste Ähnlichkeitsgrad S (Scomp), der in Schritt 1410 erfasst wird, höher ist als der in Schritt 1411 erfasste Ähnlichkeitsgrad S .

[0155] Dann verknüpft die Metadaten-Verarbeitungssteuereinheit **406** in Schritt 1412 den Erläuterungstext in einem Textkörper in der Kombination mit einem höheren Wert des Ähnlichkeitsgrads S mit dem Objektbereich.

[0156] Wie es vorstehend beschrieben ist, kann gemäß diesem Ausführungsbeispiel abhängig davon, ob das Objekt, wie etwa eine Fotografie, und der Erläuterungstext in einem Textkörper, auf der gleichen Seite liegen oder nicht, und davon, ob die Anzahl der Objekte 3 oder mehr beträgt, ein Teil der Verarbeitungen aus-/weggelassen werden. Somit können die Verarbeitungen mit einer höheren Geschwindigkeit durchgeführt werden.

[Ausführungsbeispiel 3]

[0157] Bei Ausführungsbeispiel 1 und Ausführungsbeispiel 2 wird ein Objektbereich, wie etwa eine Fotografie, mit einem Erläuterungstext in einem Textkörper nur basierend darauf verknüpft, ob der berechnete Wert des Ähnlichkeitsgrads höher oder niedriger ist. Der folgende Abschnitt wird ein Ausführungsbeispiel beschreiben, bei dem nach der Berechnung des Ähnlichkeitsgrads bestimmt wird, ob der berechnete Wert des Ähnlichkeitsgrads einen vorbestimmten Schwellenwert erreicht oder nicht. Nur wenn der berechnete Wert des Ähnlichkeitsgrads den vorbestimmten Schwellenwert erreicht, wird der Objektbereich mit dem Erläuterungstext in einem Textkörper verknüpft.

[0158] [Fig. 16](#) ist ein Ablaufdiagramm, das den Ablauf der Verarbeitung gemäß diesem Ausführungsbeispiel veranschaulicht, um einen Erläuterungstext in einem Textkörper mit einem Objektbereich zu verknüpfen. Das Ablaufdiagramm gemäß [Fig. 16](#) entspricht dem Ablaufdiagramm gemäß [Fig. 10](#) von Ausführungsbeispiel 1 und dem Ablaufdiagramm gemäß [Fig. 14A](#) und [Fig. 14B](#) von Ausführungsbeispiel 2. Daher werden gemeinsame Teile zwischen diesen einfach/vereinfacht oder nicht weiter beschrieben. Daher wird hauptsächlich ein Unterscheidungspunkt beschrieben.

[0159] Die Verarbeitungen von Schritt 1601 bis Schritt 1608 sind die gleichen wie die Verarbeitungen von Schritt 1401 bis Schritt 1408 gemäß [Fig. 14A](#) und [Fig. 14B](#). Die Verarbeitung von Schritt 1609 ist die gleiche wie die Verarbeitung von Schritt 1006 gemäß [Fig. 10](#). Wenn Ähnlichkeitsgrade von Kombinationen in Schritt 1609 berechnet werden, schreitet die Verarbeitung zu Schritt 1610 voran.

[0160] In Schritt 1610 vergleicht die Ausdrucksähnlichkeitsgrad-Berechnungseinheit **404** den Wert des höchsten Ähnlichkeitsgrads mit einem vorbestimmten Schwellenwert, der im Vorhinein eingestellt ist, um zu bestimmen, ob der Wert des Ähnlichkeitsgrads gleich oder größer dem Schwellenwert ist oder nicht. Wenn der Wert des Ähnlichkeitsgrads gleich oder größer dem Schwellenwert ist, schreitet die Verarbeitung zu Schritt 1611 voran, um einen Erläuterungstext in einem Textkörper mit einem Objektbereich zu verknüpfen. Wenn der Wert des Ähnlichkeitsgrads den Schwellenwert nicht erreicht, wird der Erläuterungstext in dem Textkörper nicht mit dem Objektbereich verknüpft und wird die Verarbeitung vollendet.

[0161] [Fig. 17](#) veranschaulicht ein Beispiel des Benutzerschnittstelle-(UI-)Bildschirms, der auf der Bedienungseinheit **203** von dem MFP **100** angezeigt wird. Der UI-Bildschirm **1701** weist darauf Schaltflächen **1702** und **1703** zum Festlegen der Ebene bzw. des Niveaus der Funktion zum Suchen eines Objekts in einem Dokument, wie etwa einer Fotografie, einer Zeichnung oder einer Tabelle, auf. Wenn die Schaltfläche **1702** gewählt wird, wird ein elektronisches Dokument durch ein Verfahren mit einer anspruchsvollen bzw. überlegenen Objektsuchfunktion erstellt (d. h. dem Verfahren gemäß der vorliegenden Erfindung). Wenn die Schaltfläche **1703** gewählt wird, wird ein elektronisches Dokument durch ein herkömmliches Verfahren erstellt, das der Komprimierung einer Dateigröße Vorrang einräumt. Das Bezugszeichen **1704** bezeichnet eine Schaltfläche, um einem Benutzer zu ermöglichen, den vorstehend beschriebenen Schwellenwert auf einen beliebigen Wert einzustellen. Um die Ebene bzw. das Niveau der Extraktion von Metadaten in einem Dokument zu erhöhen, wird die Schaltfläche **1704** auf die rechte Seite geschoben. Dies verringert folglich den Schwellenwert, um dadurch mehr Metadaten zu extrahieren. Um die Ebene bzw. das Niveau der Extraktion von Metadaten in dem Dokument andererseits zu verringern, wird die Schaltfläche **1704** auf die linke Seite geschoben. Dies erhöht folglich den Schwellenwert, um dadurch weniger Metadaten zu extrahieren. Durch diese Benutzerschnittstelle kann der Benutzer den Schwellenwert beliebig ändern, um die Ebene bzw. das Niveau der Extraktion von Metadaten auf eine gewünschte Ebene bzw. ein gewünschtes Niveau zu ändern. Das Bezugszeichen **1705** bezeichnet eine Löschschriftfläche zum Löschen der gewählten Inhalte. Das Bezugszeichen **1706** bezeichnet eine OK-Schriftfläche zum Festlegen der eingestellten Inhalte.

[0162] Gemäß diesem Ausführungsbeispiel wird der Fall, bei dem der Wert des Ähnlichkeitsgrads kleiner ist als der Schwellenwert, als der Fall angenommen, bei dem korrekte Metadaten nicht extrahiert werden können, um dadurch zu verhindern, dass Metadaten hinzugefügt werden. Dies kann folglich einen Umstand verhindern, bei dem eine falsche Verknüpfung

eines Erläuterungstexts in einem Textkörper mit einem Objektbereich eine Hinzufügung von falschen Metadaten verursacht. Somit kann die anschließende Objektsuche korrekt durchgeführt werden.

[Ausführungsbeispiel 4]

[0163] Als Nächstes wird Ausführungsbeispiel 4 beschrieben, bei dem, wenn herausgefunden wird, dass es eine Vielzahl von Objekten gibt, die eine Beschriftung aufweisen, die einen speziellen Ankerausdruck umfasst, der Benutzer durch eine Warnungsanzeige bezüglich des Vorhandenseins gewarnt wird.

[0164] [Fig. 18](#) veranschaulicht ein Beispiel eines Warnungsbildschirms, der auf der Bedienungseinheit **203** von dem MFP **100** bei diesem Ausführungsbeispiel angezeigt wird. [Fig. 19](#) ist ein Ablaufdiagramm des Ablaufs der Verarbeitung gemäß diesem Ausführungsbeispiel zum Verknüpfen eines Erläuterungstexts in einem Textkörper mit einem Objektbereich. Die gleichen Teile wie diejenigen des Ablaufdiagramms von [Fig. 10](#) gemäß Ausführungsbeispiel 1 werden einfach/vereinfacht oder nicht weiter beschrieben. Somit wird hauptsächlich ein Unterscheidungspunkt beschrieben.

[0165] In Schritt 1901 wird die Anzahl von Objekten mit einer Beschriftung erfasst, die einen speziellen Ankerausdruck umfasst. Wenn in Schritt 1902 bestimmt wird, dass die Anzahl der erfassten Objekte eine Mehrzahl ist, schreitet die Verarbeitung zu Schritt 1905 voran.

[0166] In Schritt 1905 zeigt die Metadaten-Verarbeitungssteuereinheit **406** zum Beispiel einen Warnungsbildschirm, wie er gemäß [Fig. 18](#) gezeigt ist, auf der Bedienungseinheit **203** an. Dieser Warnungsbildschirm umfasst eine Nachricht, die anzeigt, dass eine identische Zeichnungsnummer in unterschiedlichen Zeichnungen erfasst ist, und er umfasst auch eine Schaltfläche zum Festlegen, ob die Verarbeitung fortgesetzt wird oder nicht.

[0167] Wenn der Benutzer eine Schaltfläche „Fortsetzen“ **1802** auf dem Warnungsbildschirm **1801** wählt, schreitet die Verarbeitung zu Schritt 1907 voran, um die Verarbeitung fortzusetzen. Die Verarbeitungen von Schritt 1907 bis Schritt 1909 sind die gleichen wie die Verarbeitungen von Schritt 1005 bis 1007 gemäß [Fig. 10](#). Wenn der Benutzer andererseits eine Schaltfläche „Beenden“ **1803** wählt, wird die Verarbeitung angehalten, um zu einem Status vor dem Start einer Abtastung zurückzukehren.

[0168] Gemäß diesem Ausführungsbeispiel kann der Benutzer, wenn herausgefunden wird, dass es eine Vielzahl von Objekten mit Beschriftungen gibt, die einen identischen Ankerausdruck umfassen, eine Gelegenheit dazu bekommen, zu erwägen, ob die

Verarbeitung fortzusetzen ist oder nicht. Somit kann der Benutzer eine Gelegenheit dazu bekommen, die Verarbeitung anzuhalten, um die zweite Abtastung zu versuchen, um korrekte Metadaten hinzuzufügen.

[Weitere Ausführungsbeispiele]

[0169] Aspekte der vorliegenden Erfindung können auch durch einen Computer eines Systems oder einer Vorrichtung (oder Vorrichtungen bzw. Geräte, wie etwa eine CPU oder MPU) realisiert werden, der ein auf einer Speichereinheit aufgezeichnetes Programm ausliest und ausführt, um die Funktionen des/der vorstehend beschriebenen Ausführungsbeispiels/Ausführungsbeispiele durchzuführen, sowie durch ein Verfahren, dessen Schritte durch einen Computer eines Systems oder einer Vorrichtung durchgeführt werden, indem zum Beispiel ein auf einer Speichereinheit aufgezeichnetes Programm ausgelesen und ausgeführt wird, um die Funktionen des/der vorstehend beschriebenen Ausführungsbeispiels/Ausführungsbeispiele durchzuführen. Zu diesem Zweck wird an dem Computer das Programm zum Beispiel über ein Netzwerk oder von einem Aufzeichnungsmedium verschiedener Typen, das als die Speichereinheit dient (z. B. ein computerlesbares Medium), bereitgestellt.

[0170] Während die vorliegende Erfindung unter Bezugnahme auf beispielhafte Ausführungsbeispiele beschrieben wurde, ist es selbstverständlich, dass die Erfindung nicht auf die offenbarten beispielhaften Ausführungsbeispiele beschränkt ist. Dem Umfang der folgenden Patentansprüche ist die breiteste Auslegung zuzugestehen, um alle derartigen Modifikationen und äquivalente Strukturen und Funktionen zu umfassen.

[0171] Selbst wenn Beschriftungen einer Vielzahl von Objekten einen identischen Ankerausdruck verwenden, kann die vorliegende Erfindung einen geeigneten Erläuterungstext in einem Textkörper als Metadaten mit den Objekten verknüpfen.

ZITATE ENTHALTEN IN DER BESCHREIBUNG

Diese Liste der vom Anmelder aufgeführten Dokumente wurde automatisiert erzeugt und ist ausschließlich zur besseren Information des Lesers aufgenommen. Die Liste ist nicht Bestandteil der deutschen Patent- bzw. Gebrauchsmusteranmeldung. Das DPMA übernimmt keinerlei Haftung für etwaige Fehler oder Auslassungen.

Zitierte Patentliteratur

- JP 11-025113 [[0004](#)]

Zitierte Nicht-Patentliteratur

- IEEE 1394 [[0045](#)]

Patentansprüche

1. Bildverarbeitungsvorrichtung mit:
 einer Bereichsaufteilungseinheit, die zum Aufteilen eines Bildes einer Seite in eine Vielzahl von Bereichen konfiguriert ist;
 einer Attributinformationen-Hinzufügungseinheit, die zum Hinzufügen von den Bereichen entsprechenden Attributen zu der Vielzahl von aufgeteilten Bereichen konfiguriert ist;
 einer Zeichenerkennungseinheit, die zum Durchführen einer Zeichenerkennungsverarbeitung auf einem Beschriftungsbereich und einem Textkörperbereich konfiguriert ist, zu denen ein Beschriftungsattribut beziehungsweise ein Textkörperattribut durch die Attributinformationen-Hinzufügungseinheit hinzugefügt sind; und
 einer Metadaten-Verarbeitungseinheit, die zum Verknüpfen von Metadaten mit einem Objektbereich konfiguriert ist, der durch den Beschriftungsbereich begleitet ist, wobei
 die Metadaten-Verarbeitungseinheit aufweist:
 eine erste Extraktionseinheit, die zum Extrahieren eines Ankerausdrucks, der aus einer vorbestimmten Zeichenkette besteht, und eines Beschriftungsausdrucks, der aus einer von dem Ankerausdruck abweichenden Zeichenkette besteht, aus dem Ergebnis der Zeichenerkennungsverarbeitung auf dem Beschriftungsbereich konfiguriert ist;
 eine Bestimmungseinheit, die zum Bestimmen konfiguriert ist, ob es eine Vielzahl von Objektbereichen gibt, die durch Beschriftungsbereiche begleitet sind, die einen identischen Ankerausdruck umfassen;
 eine zweite Extraktionseinheit, die zum Extrahieren eines Erläuterungstexts, der den Ankerausdruck umfasst, aus dem Ergebnis der Zeichenerkennungsverarbeitung auf dem Textkörperbereich konfiguriert ist;
 eine erste Verknüpfungseinheit, die zum Verknüpfen des Objektbereichs mit Metadaten, die aus dem durch die zweite Extraktionseinheit extrahierten Erläuterungstext erhalten werden, konfiguriert ist, falls die Bestimmungseinheit bestimmt, dass es einen Objektbereich gibt, der durch einen Beschriftungsbereich begleitet ist, der den identischen Ankerausdruck umfasst;
 eine Ähnlichkeitsgrad-Berechnungseinheit, die zum jeweiligen Berechnen von Ähnlichkeitsgraden zwischen Beschriftungsausdrücken der jeweiligen Beschriftungsbereiche, die den identischen Ankerausdruck umfassen, und dem durch die zweite Extraktionseinheit extrahierten Erläuterungstext, der den identischen Ankerausdruck umfasst, konfiguriert ist, falls die Bestimmungseinheit bestimmt, dass es eine Vielzahl von Objektbereichen gibt, die durch Beschriftungsbereiche begleitet sind, die den identischen Ankerausdruck umfassen; und
 eine zweite Verknüpfungseinheit, die zum Bestimmen, basierend auf den durch die Ähnlichkeitsgrad-Berechnungseinheit berechneten Ähnlichkeits-

graden, derartiger Erläuterungstexte, die für die jeweilige Vielzahl von Objektbereichen optimal sind, und zum Verknüpfen von aus den bestimmten optimalen Erläuterungstexten erhaltenen Metadaten mit den jeweiligen Objektbereichen konfiguriert ist.

2. Bildverarbeitungsvorrichtung gemäß Anspruch 1, wobei die Objektbereiche Bereiche sind, zu denen durch die Attributinformationen-Hinzufügungseinheit irgendein Attribut einer Fotografie, einer Zeichnung oder einer Tabelle hinzugefügt ist.

3. Bildverarbeitungsvorrichtung gemäß Anspruch 1, wobei die zweite Verknüpfungseinheit zum Bestimmen eines Erläuterungstexts mit dem höchsten Ähnlichkeitsgrad, der durch die Ähnlichkeitsgrad-Berechnungseinheit berechnet ist, als einen Erläuterungstext, der für den Objektbereich optimal ist, und zum Verknüpfen von aus dem optimalen Erläuterungstext erhaltenen Metadaten mit dem Objektbereich konfiguriert ist.

4. Bildverarbeitungsvorrichtung gemäß Anspruch 1, wobei
 die Bereichsaufteilungseinheit zum Aufteilen jedes Bildes einer Vielzahl von Seiten in eine Vielzahl von Bereichen konfiguriert ist,
 die Metadaten-Verarbeitungseinheit ferner eine zweite Bestimmungseinheit aufweist, die zum Bestimmen konfiguriert ist, ob sowohl der durch die zweite Extraktionseinheit extrahierte Erläuterungstext als auch der Objektbereich innerhalb des Bildes der gleichen Seite vorhanden sind, und
 die Metadaten-Verarbeitungseinheit zum Überspringen der Berechnung des Ähnlichkeitsgrads durch die Ähnlichkeitsgrad-Berechnungseinheit und zum Verknüpfen von Metadaten, die aus dem Erläuterungstext erhalten werden, der aus dem Bild der gleichen Seite extrahiert ist, mit dem Objektbereich, der innerhalb des Bildes der gleichen Seite vorhanden ist, konfiguriert ist, falls die zweite Bestimmungseinheit bestimmt, dass sowohl der durch die zweite Extraktionseinheit extrahierte Erläuterungstext als auch der Objektbereich innerhalb des Bildes der gleichen Seite vorhanden sind.

5. Bildverarbeitungsvorrichtung gemäß Anspruch 1, wobei die zweite Verknüpfungseinheit zum Verknüpfen von aus dem optimalen Erläuterungstext erhaltenen Metadaten mit dem Objektbereich konfiguriert ist, falls der Ähnlichkeitsgrad bezüglich des optimalen Erläuterungstexts einen vorbestimmten Schwellenwert erreicht.

6. Bildverarbeitungsvorrichtung gemäß Anspruch 1, wobei die Metadaten-Verarbeitungseinheit zum Bereitstellen eines Warnungsbildschirms konfiguriert ist, der das Vorhandensein der Vielzahl von Objektbereichen anzeigt, falls die Bestimmungseinheit bestimmt, dass es eine Vielzahl von Objektbereichen

gibt, die von Beschriftungsbereichen begleitet sind, die den identischen Ankerausdruck umfassen.

7. Bildverarbeitungsvorrichtung gemäß Anspruch 1, wobei die Bildverarbeitungsvorrichtung ferner eine dritte Verknüpfungseinheit aufweist, die zum Verknüpfen von dem durch die erste Extraktionseinheit extrahierten Beschriftungsausdruck als Metadaten mit einem Objektbereich, der von einem Beschriftungsbereich begleitet ist, der den Beschriftungsausdruck umfasst, konfiguriert ist.

8. Bildverarbeitungsvorrichtung gemäß Anspruch 1, wobei die Bildverarbeitungsvorrichtung ferner eine Erzeugungseinheit aufweist, die zum Verwenden von Informationen bezüglich der Vielzahl von Bereichen, die aus dem Bild einer Seite durch die Bereichsaufteilungseinheit aufgeteilt sind, und mit dem Objektbereich verknüpften Metadaten konfiguriert ist, um ein elektronisches Dokument mit einem vorbestimmten Format zu erzeugen.

9. Bildverarbeitungsverfahren, mit:
 einem Bereichsaufteilungsschritt zum Aufteilen eines Bildes einer Vielzahl von Seiten in eine Vielzahl von Bereichen;
 einem Attributinformationen-Hinzufügungsschritt zum Hinzufügen von von den Bereichen abhängigen Attributen zu der jeweiligen Vielzahl von aufgeteilten Bereichen;
 einem Zeichenerkennungsschritt zum Durchführen einer Zeichenerkennungsverarbeitung auf einem Beschriftungsbereich und einem Textkörperbereich, zu denen ein Beschriftungsattribut beziehungsweise ein Textkörperattribut durch den Attributinformationen-Hinzufügungsschritt hinzugefügt sind; und
 einem Metadaten-Verarbeitungsschritt zum Verknüpfen von Metadaten mit einem Objektbereich, der durch den Beschriftungsbereich begleitet ist, wobei
 der Metadaten-Verarbeitungsschritt aufweist:
 einen ersten Extraktionsschritt zum Extrahieren eines Ankerausdrucks, der aus einer vorbestimmten Zeichenkette besteht, und eines Beschriftungsausdrucks, der aus einer von dem Ankerausdruck abweichenden Zeichenkette besteht, aus dem Ergebnis der Zeichenerkennungsverarbeitung auf dem Beschriftungsbereich;
 einen Bestimmungsschritt zum Bestimmen, ob es eine Vielzahl von Objektbereichen gibt, die durch Beschriftungsbereiche begleitet sind, die einen identischen Ankerausdruck umfassen;
 einen zweiten Extraktionsschritt zum Extrahieren eines Erläuterungstexts, der den Ankerausdruck umfasst, aus dem Ergebnis der Zeichenerkennungsverarbeitung auf dem Textkörperbereich;
 einen ersten Verknüpfungsschritt zum Verknüpfen von Metadaten, die aus dem durch den zweiten Extraktionsschritt extrahierten Erläuterungstext erhalten werden, mit dem Objektbereich, falls der Bestim-

mungsschritt bestimmt, dass es einen Objektbereich gibt, der durch einen Beschriftungsbereich begleitet ist, der den identischen Ankerausdruck umfasst;
 einen Ähnlichkeitsgrad-Berechnungsschritt zum jeweiligen Berechnen von Ähnlichkeitsgraden zwischen Beschriftungsausdrücken der jeweiligen Beschriftungsbereiche, die den identischen Ankerausdruck umfassen, und dem durch den zweiten Extraktionsschritt extrahierten Erläuterungstext, der den identischen Ankerausdruck umfasst, falls der Bestimmungsschritt bestimmt, dass es eine Vielzahl von Objektbereichen gibt, die durch Beschriftungsbereiche begleitet sind, die den identischen Ankerausdruck umfassen; und
 einen zweiten Verknüpfungsschritt zum Bestimmen, basierend auf den durch den Ähnlichkeitsgrad-Berechnungsschritt berechneten Ähnlichkeitsgraden, derartiger Erläuterungstexte, die für die jeweilige Vielzahl von Objektbereichen optimal sind, und zum Verknüpfen von aus den bestimmten optimalen Erläuterungstexten erhaltenen Metadaten mit den jeweiligen Objektbereichen.

10. Nicht vorübergehendes computerlesbares Speichermedium mit computerausführbaren Instruktionen zum Durchführen von:
 einem Bereichsaufteilungsschritt zum jeweiligen Aufteilen eines Bildes einer Vielzahl von Seiten in eine Vielzahl von Bereichen;
 einem Attributinformationen-Hinzufügungsschritt zum Hinzufügen von von den Bereichen abhängigen Attributen zu der jeweiligen Vielzahl von aufgeteilten Bereichen;
 einem Zeichenerkennungsschritt zum Durchführen einer Zeichenerkennungsverarbeitung auf einem Beschriftungsbereich und einem Textkörperbereich, zu denen ein Beschriftungsattribut beziehungsweise ein Textkörperattribut durch den Attributinformationen-Hinzufügungsschritt hinzugefügt sind; und
 einem Metadaten-Verarbeitungsschritt zum Verknüpfen von Metadaten mit einem Objektbereich, der durch den Beschriftungsbereich begleitet ist, wobei
 der Metadaten-Verarbeitungsschritt aufweist:
 einen ersten Extraktionsschritt zum Extrahieren eines Ankerausdrucks, der aus einer vorbestimmten Zeichenkette besteht, und eines Beschriftungsausdrucks, der aus einer von dem Ankerausdruck abweichenden Zeichenkette besteht, aus dem Ergebnis der Zeichenerkennungsverarbeitung auf dem Beschriftungsbereich;
 einen Bestimmungsschritt zum Bestimmen, ob es eine Vielzahl von Objektbereichen gibt, die durch Beschriftungsbereiche begleitet sind, die einen identischen Ankerausdruck umfassen;
 einen zweiten Extraktionsschritt zum Extrahieren eines Erläuterungstexts, der den Ankerausdruck umfasst, aus dem Ergebnis der Zeichenerkennungsverarbeitung auf dem Textkörperbereich;

einen ersten Verknüpfungsschritt zum Verknüpfen von Metadaten, die aus dem durch den zweiten Extraktionsschritt extrahierten Erläuterungstext erhalten werden, mit dem Objektbereich, wenn der Bestimmungsschritt bestimmt, dass es einen Objektbereich gibt, der durch einen Beschriftungsbereich begleitet ist, der den identischen Anker Ausdruck umfasst;

einen Ähnlichkeitsgrad-Berechnungsschritt zum jeweiligen Berechnen von Ähnlichkeitsgraden zwischen Beschriftungsausdrücken der jeweiligen Beschriftungsbereiche, die den identischen Anker Ausdruck umfassen, und dem durch den zweiten Extraktionsschritt extrahierten Erläuterungstext, der den identischen Anker Ausdruck umfasst, wenn der Bestimmungsschritt bestimmt, dass es eine Vielzahl von Objektbereichen gibt, die durch Beschriftungsbereiche begleitet sind, die den identischen Anker Ausdruck umfassen; und

einen zweiten Verknüpfungsschritt zum Bestimmen, basierend auf den durch den Ähnlichkeitsgrad-Berechnungsschritt berechneten Ähnlichkeitsgraden, derartiger Erläuterungstexte, die für die jeweilige Vielzahl von Objektbereichen optimal sind, um aus den bestimmten optimalen Erläuterungstexten erhaltene Metadaten mit den jeweiligen Objektbereichen zu verknüpfen.

Es folgen 24 Blatt Zeichnungen

Anhängende Zeichnungen

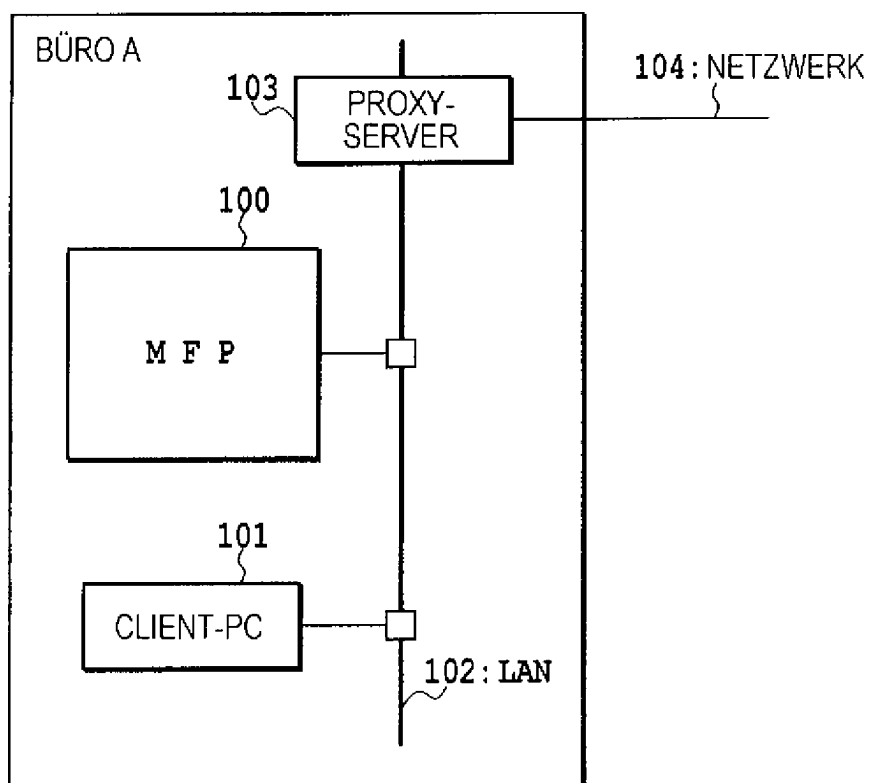


FIG.1

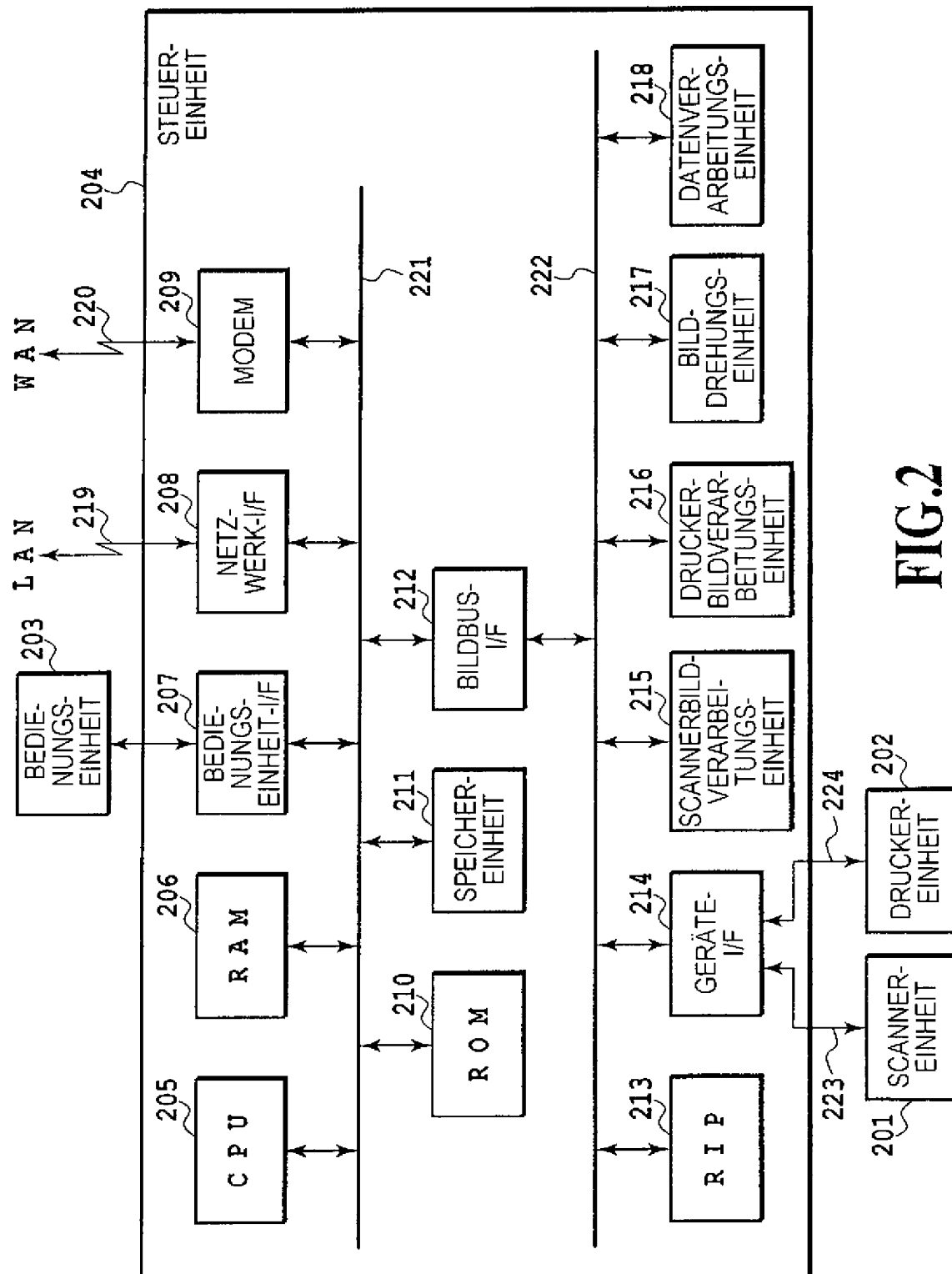


FIG. 2

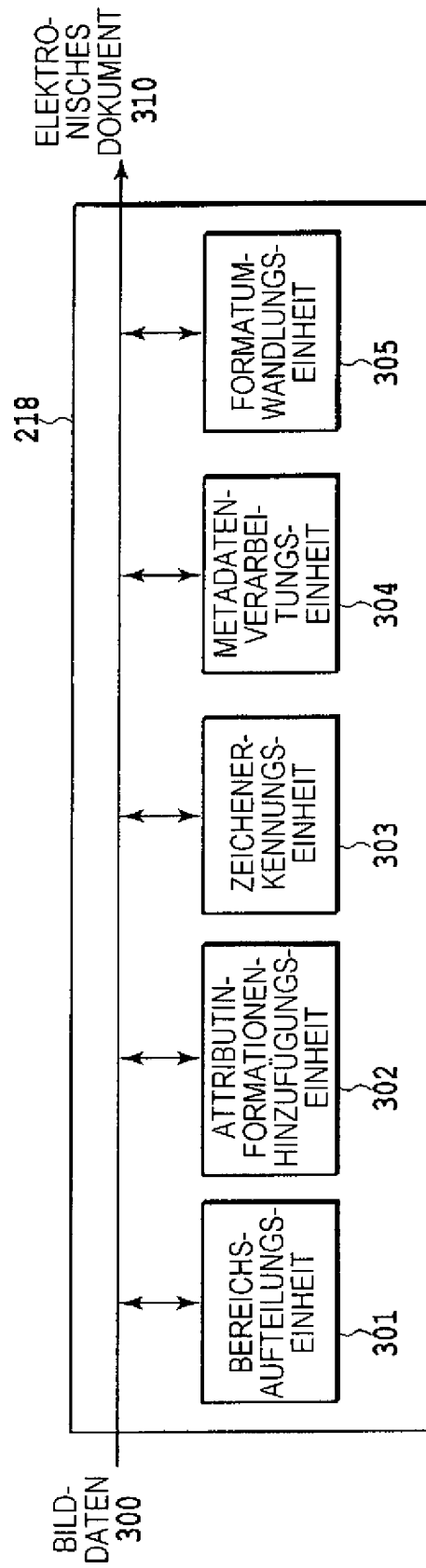


FIG.3

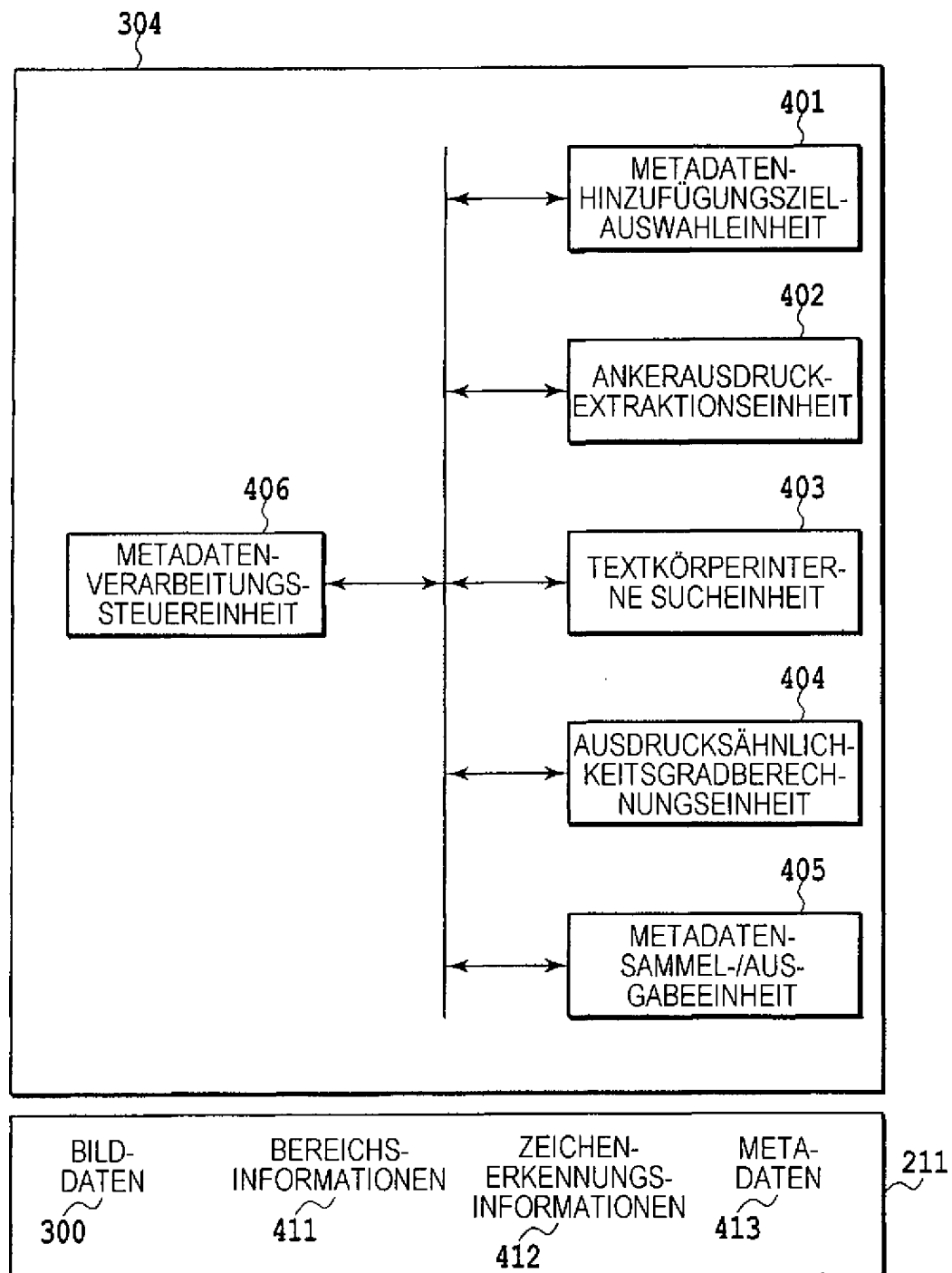


FIG.4

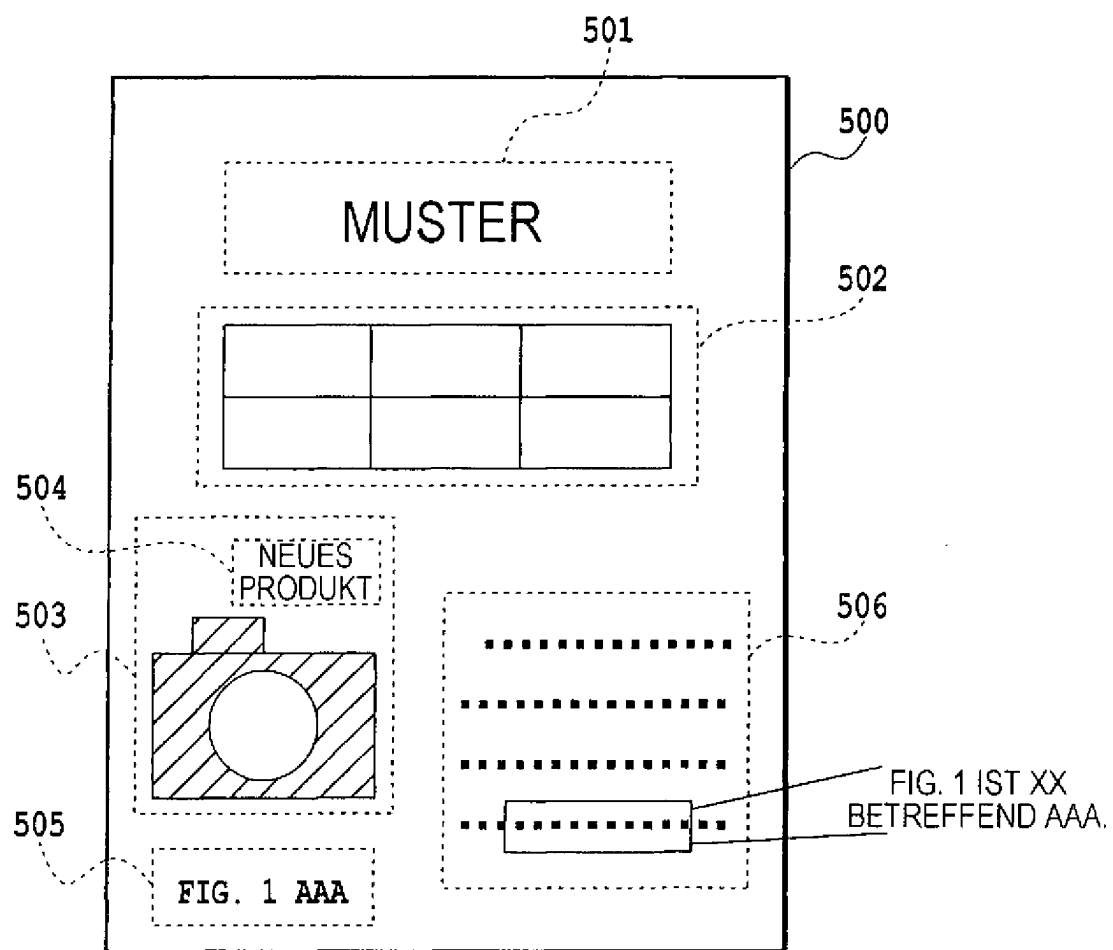


FIG.5A

| BE- REICH | ATTRIBUT | DURCH BESCHRIF- TUNG BEGLEITE- TER BEREICH | KOORDI- NATE X | KOORDI- NATE Y | BREITE W | HÖHE H | SEITE | ZEICHEN- ERKENNUNGS- INFORMATIONEN |
|--------------|---------------------------------------|--|----------------------|----------------------|-------------|-----------|-------|--|
| 501 | <u>ÜBERSCHRIFT</u> | - | X1 | Y1 | W1 | H1 | 1 | MUSTER |
| 502 | TABELLE | - | X2 | Y2 | W2 | H2 | 1 | - |
| 503 | FOTOGRAFIE | - | X3 | Y3 | W3 | H3 | 1 | - |
| 504 | <u>ZEICHEN</u> (IN FOTOGRAFIE 503) | - | X4 | Y4 | W4 | H4 | 1 | NEUES PRODUKT |
| 505 | <u>BESCHRIFTUNG</u> | 503 | X5 | Y5 | W5 | H5 | 1 | FIG. 1 AAA |
| 506 | TEXTKÖRPER | - | X6 | Y6 | W6 | H6 | 1 | ... FIG. 1 IST XX BETREFFEND AAA ... |

FIG.5B

| ATTRIBUT VERARBEI- TUNGSDETAILS | ZEICHEN | FOTOGRAFIE | ZEICHNUNG (STRICH-/LINIEN- ZEICHNUNG) | TABELLE |
|---------------------------------------|-----------------------|---------------------|---|--|
| | VEKTORUM- WANDLUNG | BILD- AUSSCHNITT | VEKTORUM- WANDLUNG | VEKTORUMWANDLUNG (TABELLENRAHMENTEIL) |
| | LÖSCHUNGSVERARBEITUNG | JA | JA | JA |

FIG.5C

```

<?xml version="1.0" ?>
<svg xmlns="http://www.w3.org/2000/svg" width="2480" height="3520"
  xmlns:x="http://@@@.jp">
  600
  601 { <text x="X1" y="Y1" color="#000000">MUSTER</text>
  602 { <g>
    <path stroke="#000000" d="M X4,Y4 L···"/>
    .....
  </g>
  603 { <g x:caption_id = "1"> 608
    <image x="X3" y="Y3" width="W3" height="H3"
      xlink:href="data:image/png;base64,Dez3632fsod6Xhd0dsc4d .....
      "/>
    .....
  </g>
  604 { <text x="X4" y="Y4" color="#000000">NEUES PRODUKT</text>
  605 { <text x="X5" y="Y5" color="#000000">FIG. 1 AAA</text>
  606 { <text x="X6" y="Y6" color="#000000">··· FIG. 1 IST XX BETREFFEND
    AAA. ···</text>
  607 { <desc>
    <metadata>
      609
      <x:meta id="1" x:meta_type="textInObject" str="NEUES PRODUKT"/>
      <x:meta id="1" x:meta_type="caption" str="FIG. 1 AAA"/>
      <x:meta id="1" x:meta_type="anchor" str="FIG. 1"/>
      <x:meta id="1" x:meta_type="textInBody" str="FIG. 1 IST XX
        BETREFFEND AAA."/>
    </metadata>
  </desc>
</svg>

```

FIG.6A

| BESCHRIF- TUNGSBE- ZEICHNER | HINZU- FÜGUNGS- ZIEL | EXTRA- HIERT AUS | SEITE | EXTRAKTIONS- INFORMA- TIONSTYP | EXTRAHIERTE INFORMATION |
|-----------------------------------|----------------------------|---------------------|-------|--|------------------------------------|
| 1 | 503 | 504 | 1 | ZEICHEN IN OBJEKT | NEUES PRODUKT |
| 1 | 503 | 505 | 1 | BESCHRIF- TUNG | FIG. 1 AAA |
| 1 | 503 | 505 | 1 | ANKER- AUSDRUCK | FIG. 1 |
| 1 | 503 | 506 | 1 | ERLÄUTE- RUNGSTEXT IN KÖRPERTEXT | FIG. 1 IST XX BE- TREFFEND AAA. |

FIG.6B

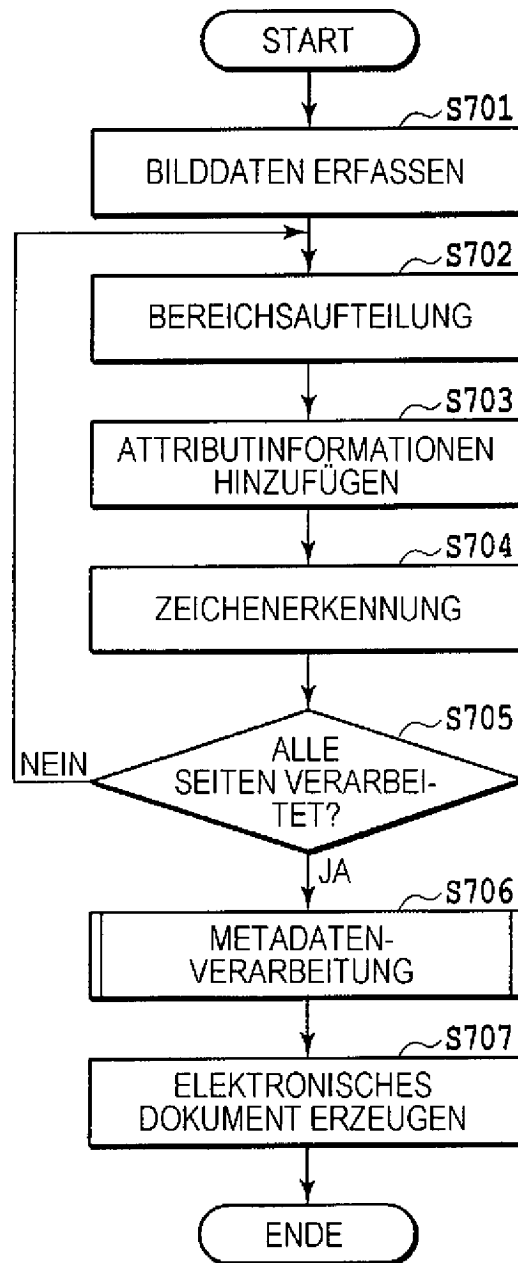


FIG.7

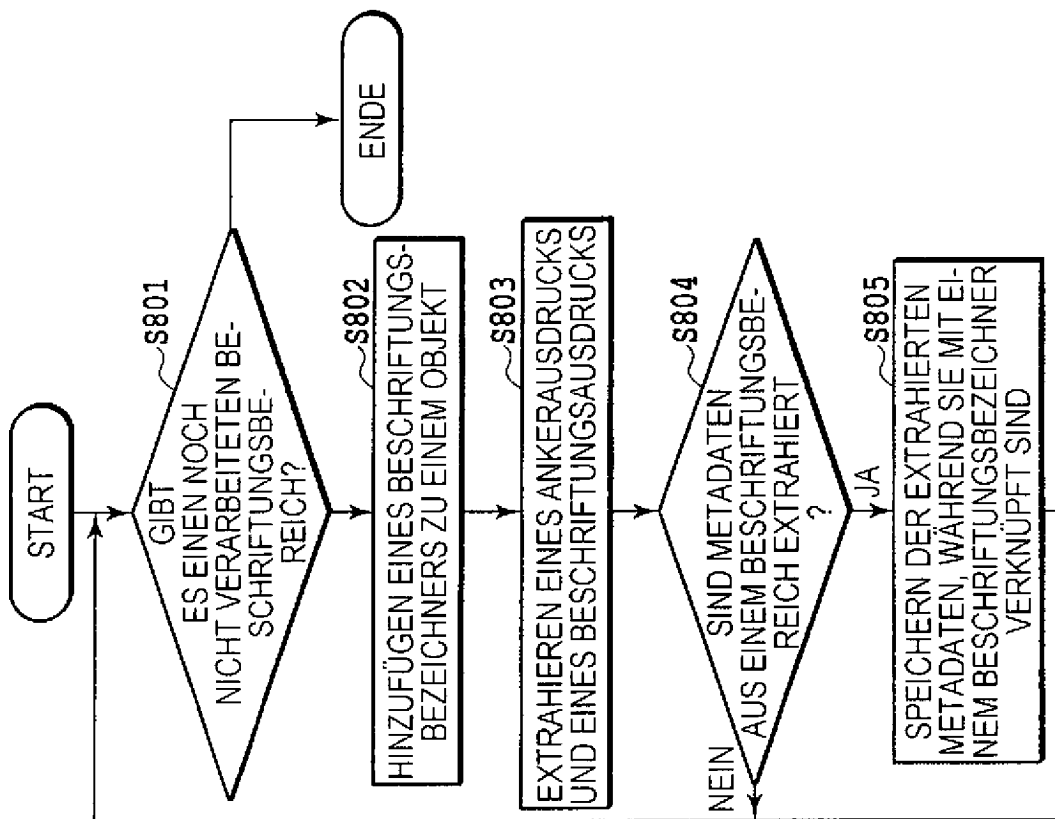


FIG. 8A

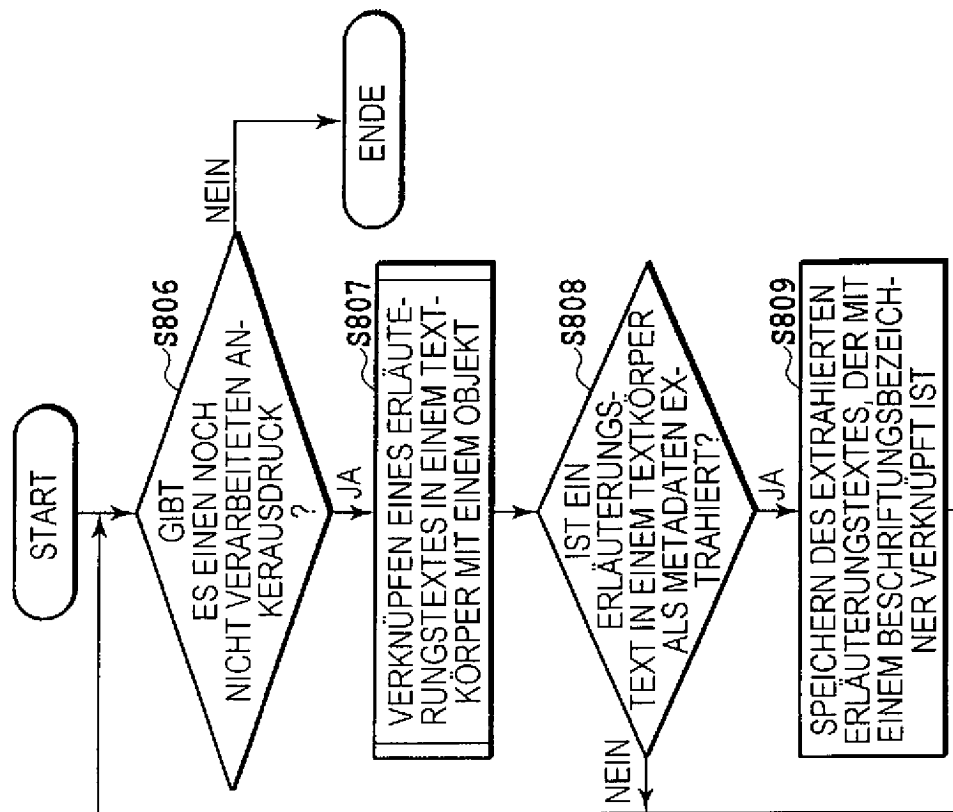


FIG. 8B

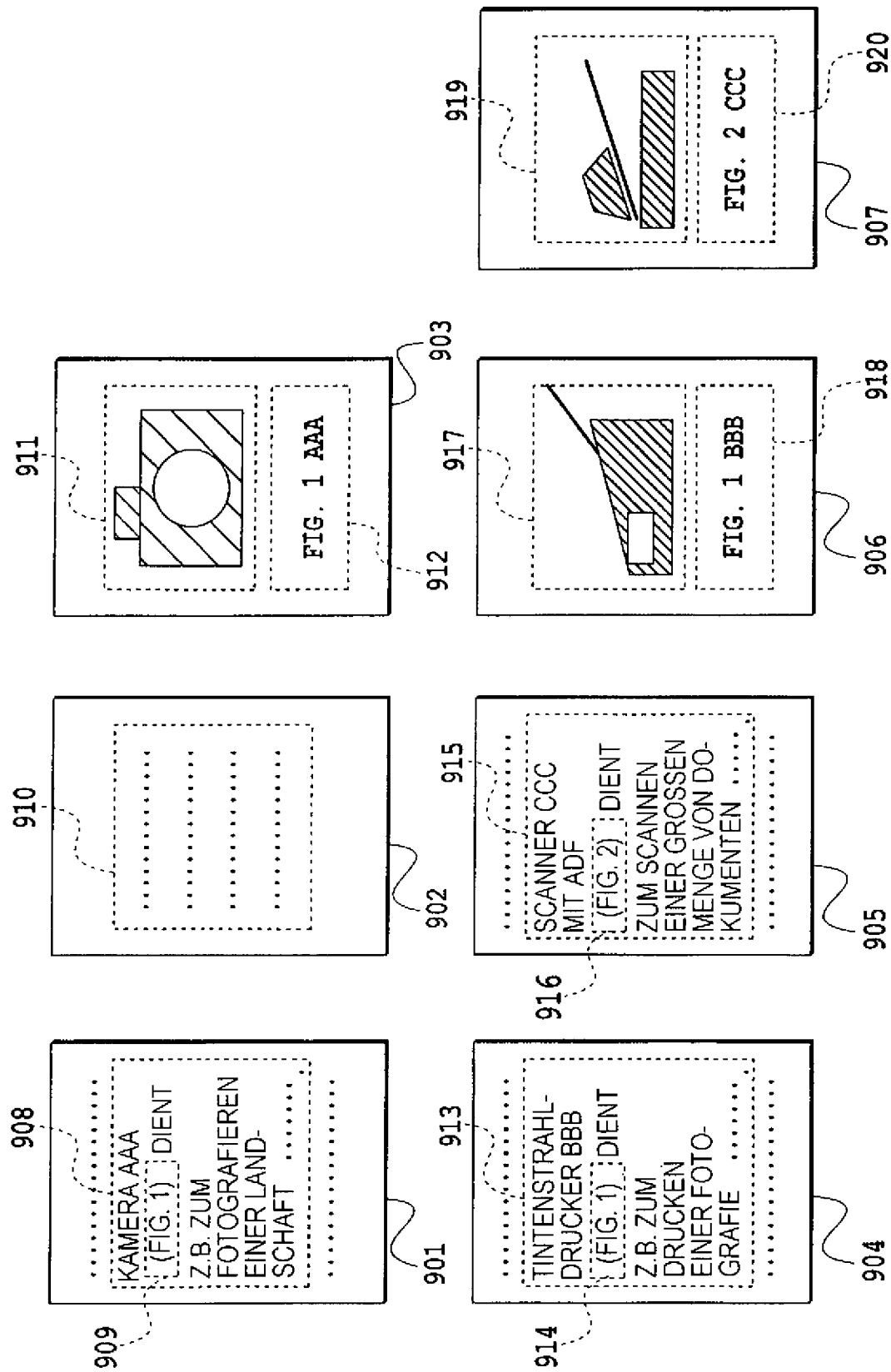


FIG. 9A

| BE- REICH | ATTRIBUT | DURCH BESCHRIF- TUNG BEGLEITE- TER BEREICH | KOORDI- NATE X | KOORDI- NATE Y | BREI- TE W | HÖHE H | SEI- TE | ZEICHEN- ERKENNUNGS- INFORMATIONEN |
|--------------|-------------------|--|----------------------|----------------------|------------------|-----------|------------|--|
| 908 | TEXTKÖRPER | - | X08 | Y08 | W08 | H08 | 1 | ... KAMERA AAA (FIG. 1) IST ... |
| 910 | TEXTKÖRPER | - | X10 | Y10 | W10 | H10 | 2 | |
| 911 | FOTOGRAFIE | - | X11 | Y11 | W11 | H11 | 3 | - |
| 912 | BESCHRIF- TUNG | 911 | X12 | Y12 | W12 | H12 | 3 | FIG. 1 AAA |
| 913 | TEXTKÖRPER | - | X13 | Y13 | W13 | H13 | 4 | ... TINTENSTRAHLDRUCKER BBB (FIG. 1) IST ... |
| 915 | TEXTKÖRPER | - | X15 | Y15 | W15 | H15 | 5 | ... SCANNER CCC MIT ADF (FIG. 2) DIENT ... EINER GROSSEN MENGE VON ... |
| 917 | FOTOGRAFIE | - | X17 | Y17 | W17 | H17 | 6 | - |
| 918 | BESCHRIF- TUNG | 917 | X18 | Y18 | W18 | H18 | 6 | FIG. 1 BBB |
| 919 | FOTOGRAFIE | - | X19 | Y19 | W19 | H19 | 7 | - |
| 920 | BESCHRIF- TUNG | 919 | X20 | Y20 | W20 | H20 | 7 | FIG. 2 CCC |

FIG.9B

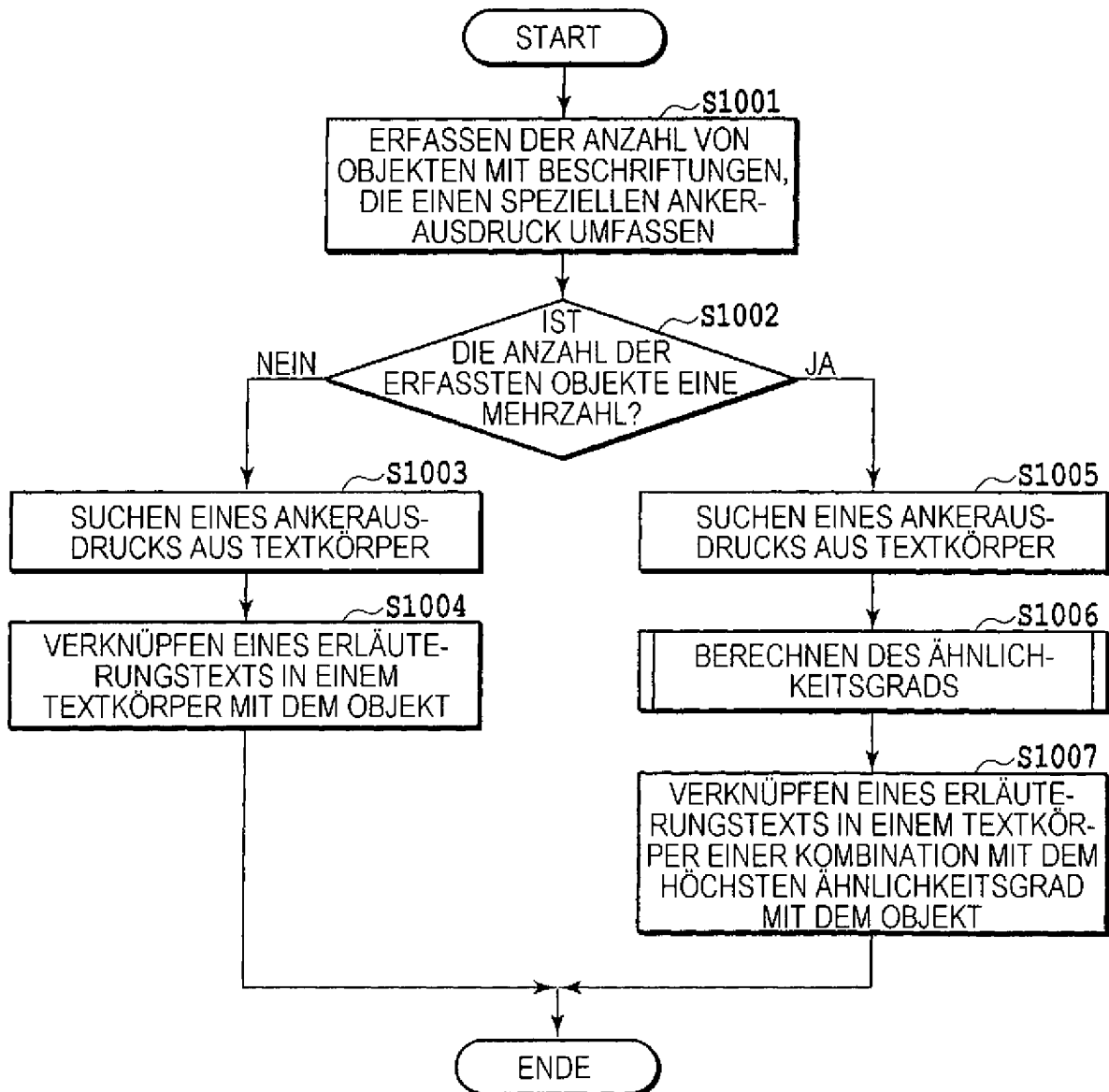


FIG.10

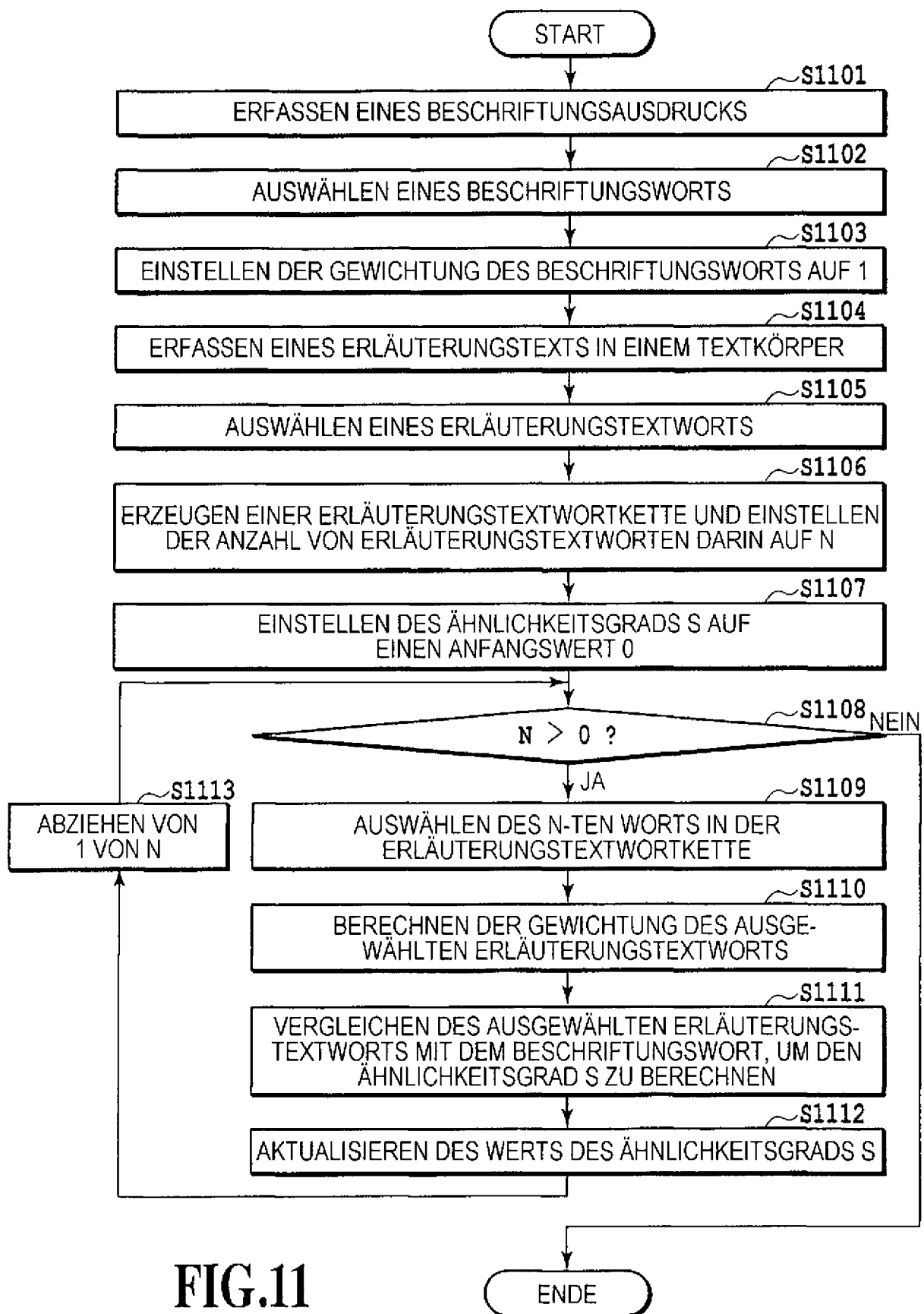


FIG.11

| | |
|------------------------------------|------|
| BESCHRIFTUNGS-AUSDRUCK | AAA |
| BESCHRIFTUNGS-WORT | AAA |
| ATTRIBUT VON BESCHRIFTUNGS-WORT | NOUN |
| VERGLEICH-SZIELKANDIDAT? | ○ |
| WORT-ABSTAND | - |
| GEWICHTUNG | 1 |

FIG.12A

| ERLÄUTE- RUNGSTEXT IN TEXTKÖRPER | KAMERA AAA (FIG. 1) DIENT Z.B. ZUM FOTOGRAFIEREN EINER LANDSCHAFT , | | | | | | | | | |
|--|---|-------|--------|----------------------|--------|----------|-------|-----------------|----------|-----|
| | KAMERA | AAA | (| FIG. 1 |) | DIENT | , | LAND- SCHAFT | Z.B. | ... |
| ERLÄUTE- RUNGS- TEXTWORT | | | | | | | | | | |
| ATTRIBUT VON ERLÄUTE- RUNGS- TEXTWORT | NOMEN | NOMEN | SYMBOL | (ANKER- AUSDRUCK) | SYMBOL | PARTIKEL | KOMMA | NOMEN | PARTIKEL | ... |
| VERGLEICH- ZIEL- KANDIDAT | ○ | ○ | x | (x) | x | x | x | ○ | x | ... |
| ORDNUNGS- ZAHL VON ER- LÄUTERUNGS- TEXTWORT- KETTE | 2 | 3 | - | - | - | - | - | 1 | - | ... |
| WORT- ABSTAND | 3 | 2 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | ... |
| GEWICHTUNG | 0.33 | 0.5 | - | - | - | - | - | 0.25 | - | ... |

FIG.12B

| BESCHRIF- TUNGSBE- ZEICHNER | HINZU- FÜGUNGS- ZIEL | EXTRA- HIERT AUS | SEITE | EXTRAK- TIONSFOR- MATIONSTYP | EXTRAKTIONS- INFORMATIONEN |
|-----------------------------------|----------------------------|---------------------|-------|--|--|
| 1 | 911 | 908 | 1 | ERLÄUTE- RUNGSTEXT IN TEXTKÖRPER | KAMERA AAA (FIG. 1) DIENT Z.B. ZUM ... |
| 1 | 911 | 912 | 3 | ANKER- AUSDRUCK | FIG. 1 |
| 1 | 911 | 912 | 3 | BESCHRIF- TUNG | FIG. 1 AAA |
| 2 | 917 | 913 | 4 | ERLÄUTE- RUNGSTEXT IN TEXTKÖRPER | TINTEN- STRAHLDRUCKER BBB (FIG. 1) DIENT ... |
| 3 | 919 | 915 | 5 | ERLÄUTE- RUNGSTEXT IN TEXTKÖRPER | SCANNER CCC MIT ADF (FIG. 2) DIENT ... |
| 2 | 917 | 918 | 6 | ANKER- AUSDRUCK | FIG. 1 |
| 2 | 917 | 918 | 6 | BESCHRIF- TUNG | FIG. 1 BBB |
| 3 | 919 | 920 | 3 | ANKER- AUSDRUCK | FIG. 2 |
| 3 | 919 | 920 | 3 | BESCHRIF- TUNG | FIG. 2 CCC |

FIG.13

FIG.14

FIG.14A

FIG.14B

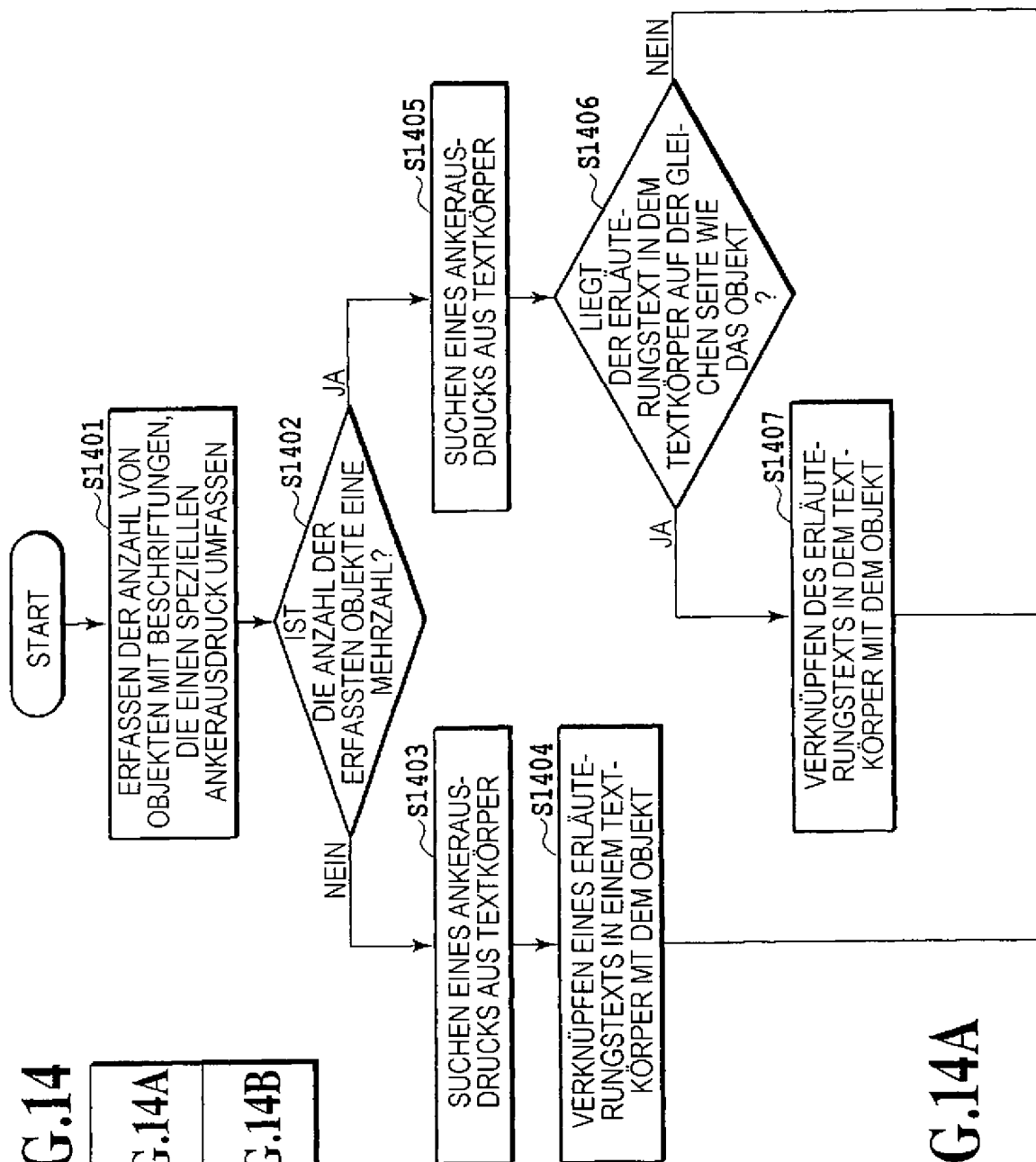


FIG.14A

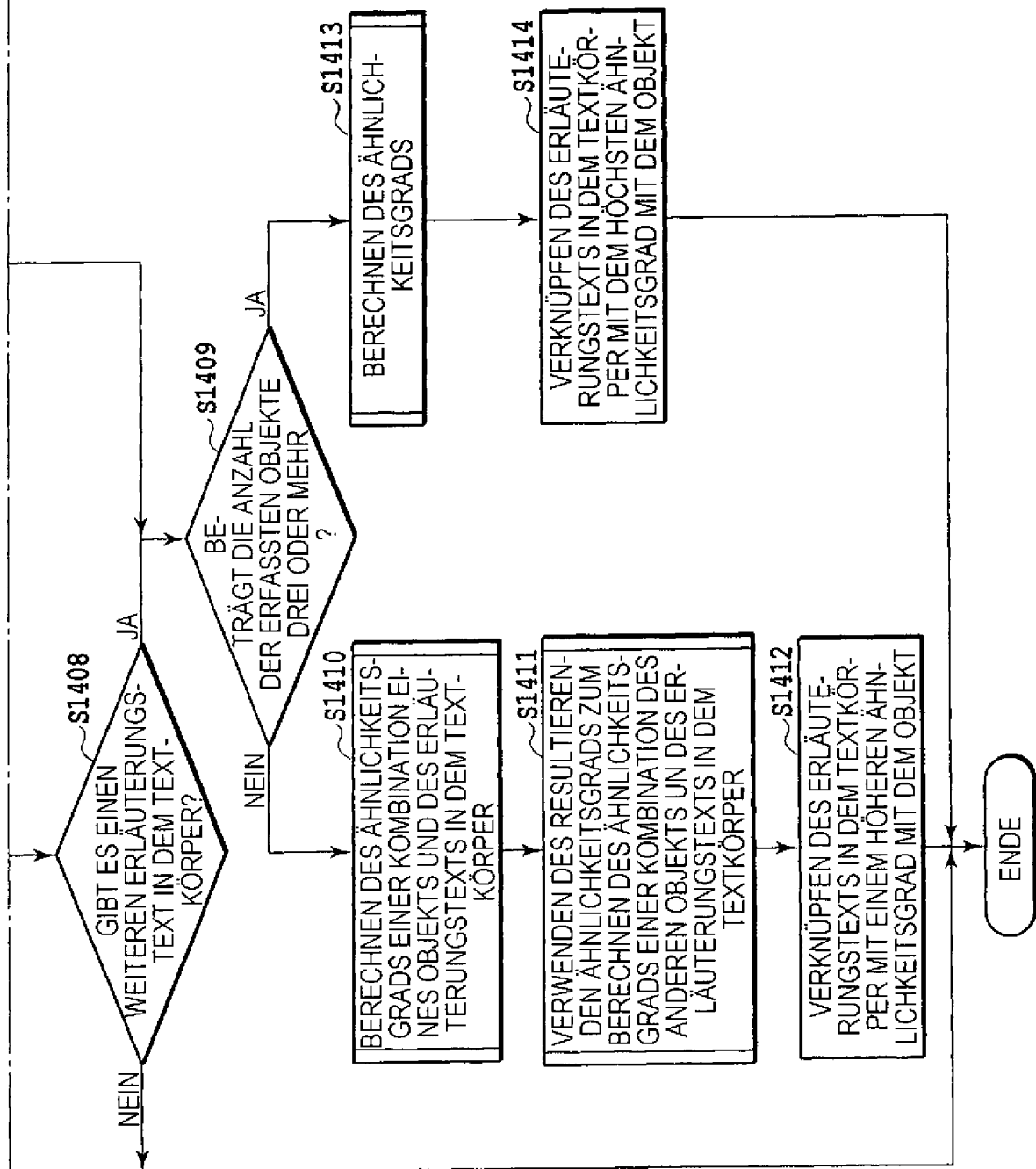
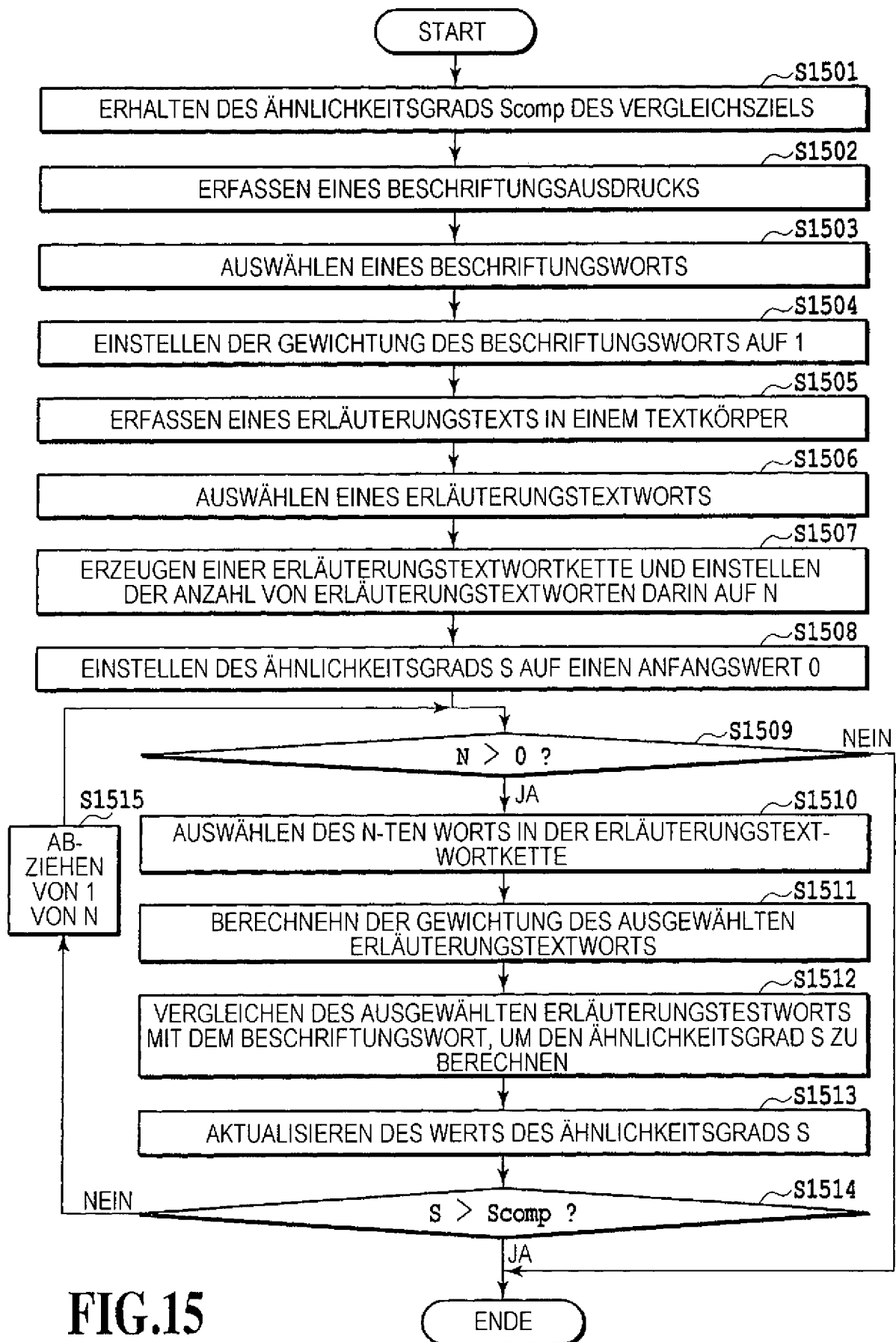


FIG. 14B



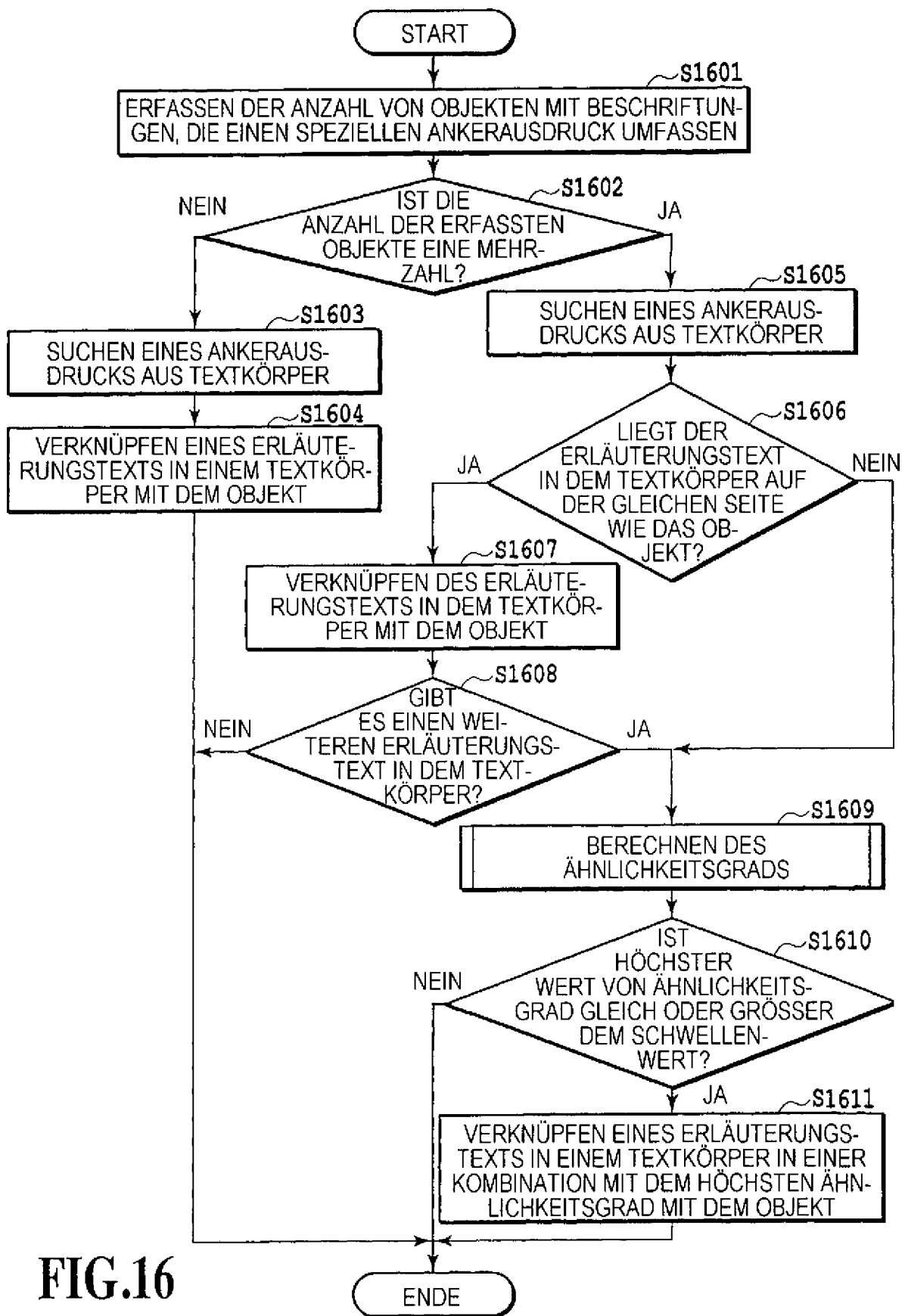


FIG.16

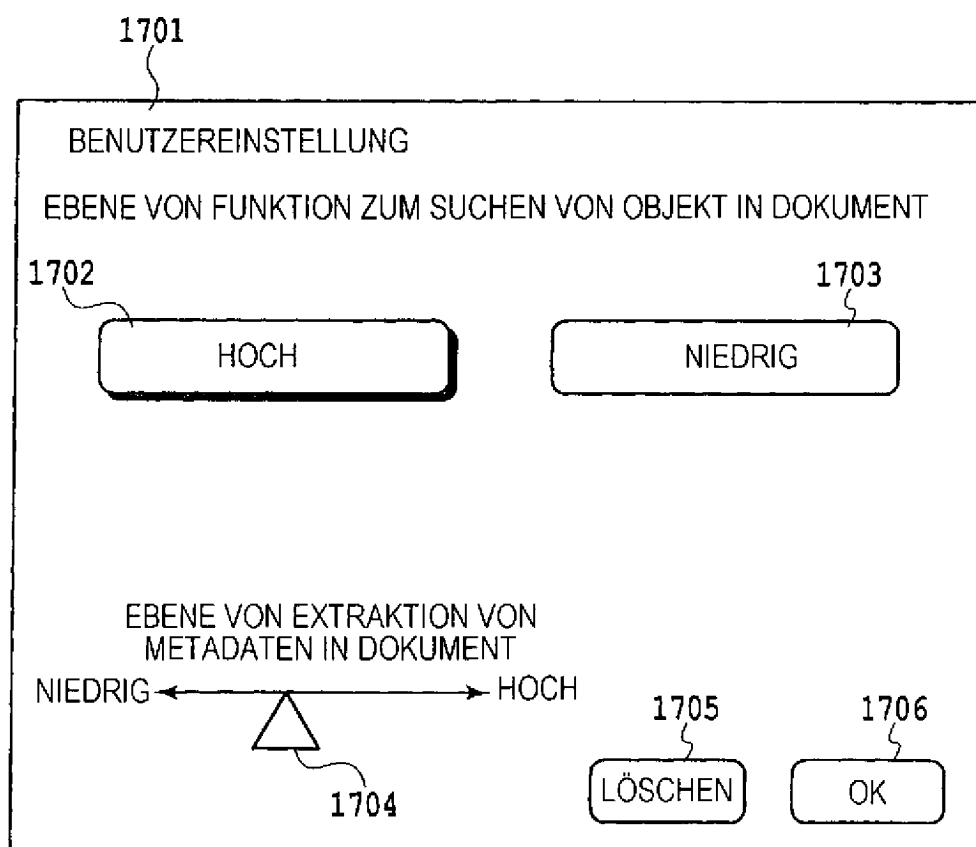


FIG.17

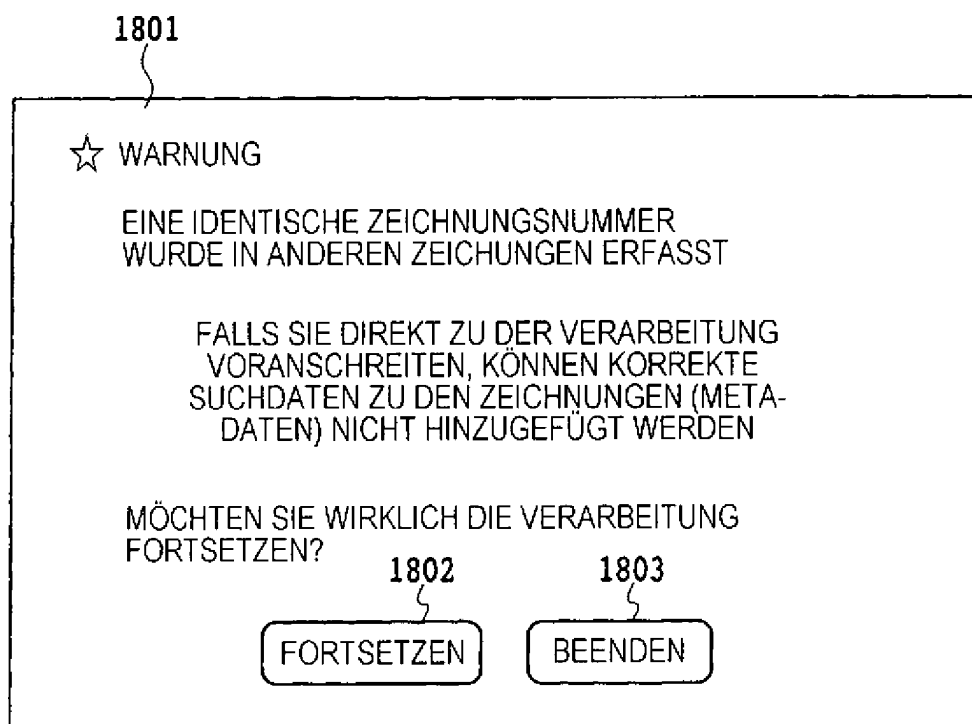
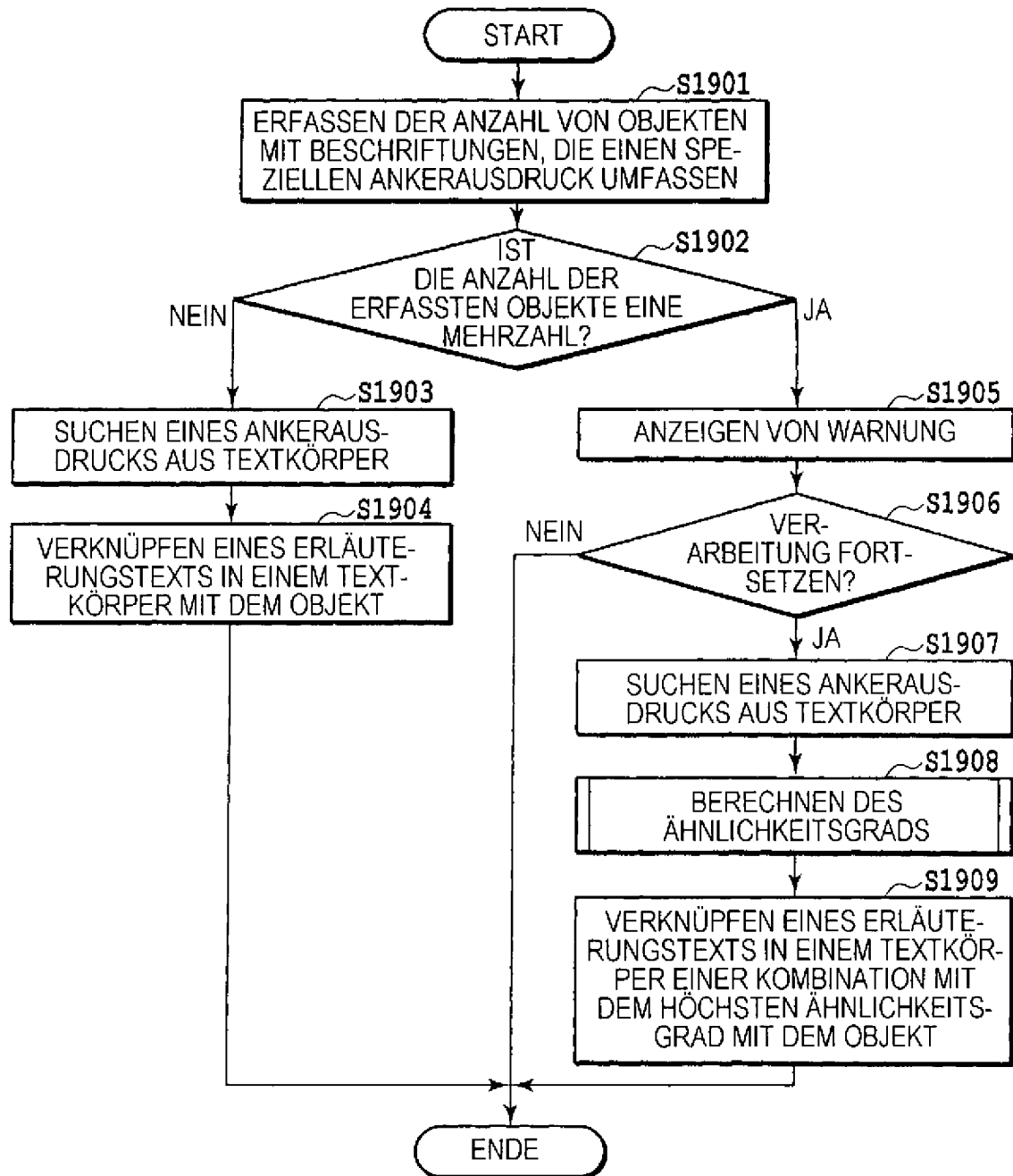


FIG.18

**FIG.19**