



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 699 23 650 T2 2005.07.07**

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 1 014 278 B1**

(21) Deutsches Aktenzeichen: **699 23 650.9**

(96) Europäisches Aktenzeichen: **99 310 220.1**

(96) Europäischer Anmeldetag: **17.12.1999**

(97) Erstveröffentlichung durch das EPA: **28.06.2000**

(97) Veröffentlichungstag

der Patenterteilung beim EPA: **09.02.2005**

(47) Veröffentlichungstag im Patentblatt: **07.07.2005**

(51) Int Cl.?: **G06F 17/28**

G06F 17/30

(30) Unionspriorität:

218431 22.12.1998 US

(73) Patentinhaber:

Xerox Corp., Rochester, N.Y., US

(74) Vertreter:

**Grünecker, Kinkeldey, Stockmair &
Schwanhäusser, 80538 München**

(84) Benannte Vertragsstaaten:

DE, FR, GB

(72) Erfinder:

**Williamowski, Jutta, 38100 Grenoble, FR;
Borghoff, Uwe M., 85635 Hoehenkirchen, DE**

(54) Bezeichnung: **System für mehrsprachige Informationswiederauffindung**

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

Beschreibung

[0001] Die vorliegende Erfindung bezieht sich auf das Durchsuchen von Informationsspeicherorten, wie etwa Datenbanken, und insbesondere auf eine Einrichtung zum Erzeugen mehrsprachiger Abfragen.

[0002] Eine beispielhafte Informations-Wiederauffindungsarchitektur ist im Aufsatz mit dem Titel "System Components For Embedded Information Retrieval From Multiple Disparate Information Sources", Ramana B. Rao, Daniel M. Russel und Jock D. Mackinlay, Proceedings of 1993 ACM Symposium on User Interface Software and Technology, Atlanta, GA, November 1993 ACM SIGGRAPH and SIGCHI beschrieben. Die Architektur enthält einen Zwischenserver, der Zugriffsanfragen zwischen einem Informationszugriffs-Client (d.h. dem Benutzer) und unterschiedlichen Informationsquellen verwaltet. Somit muß der Benutzer lediglich mit dem Informationszugriffs-Client arbeiten, um die Informationen aus unterschiedlichen Informationsquellen wiederaufzufinden.

[0003] Eine weitere Eigenschaft von Informationen im Internet besteht darin, daß sie in einer beliebigen Sprache sein können. Im allgemeinen sucht eine Suchanfrage nach Begriffen die in derselben Sprache sind wie die Suchanfrage. Bei Situationen, in denen die gefundenen Informationen in einer anderen Sprache sind, besteht der Grund hierfür normalerweise darin, daß die Informationen ein "Wort" beinhalten, das mit dem Suchbegriff übereinstimmt. Beispielsweise kann die Suche nach Informationen über eine berühmte Persönlichkeit oder Ereignis zum Erhalt von Informationen/Dokumenten in mehreren Sprachen führen.

[0004] Ein Verfahren auf der Basis eines Wörterbuchs für mehrsprachige Informationswiederauffindung ist von Lisa Ballesteros und Bruce Croft, "Dictionary Methods for Cross-Lingual Information Retrieval", Lecture Notes in Computer Science 1134 ISSN 0302-9743 (1996) beschrieben. Die Druckschrift beschreibt Versuche, die die Faktoren analysieren, die Verfahren auf der Basis eines Wörterbuchs für mehrsprachige Informationswiederauffindung und gegenwärtige Verfahren beeinflussen, die die Fehler drastisch verringern, die ein derartiger Ansatz normalerweise macht. Die Druckschrift definiert die mehrsprachige Informationswiederauffindung als die Fähigkeit, in einer Sprache zu suchen, das Wiederauffinden jedoch in mehreren Sprachen auszuführen.

[0005] Gemäß einem Aspekt der vorliegenden Erfindung umfaßt ein Verfahren zum Erzeugen einer mehrsprachigen Abfrage in einem Informations-Wiederauffindungssystem folgende Schritte:

- a) Empfangen eines Suchbegriffes;
- b) Extrahieren eines oder mehrerer elementarer

Wörter aus dem Suchbegriff;

- c) Bestimmen abgestammter Übersetzungen für jedes der einzelnen oder der mehreren Wörter;
- d) Suchen nach und Wiederauffinden von Dokumenten, die Kombinationen der abgestammten Übersetzungen enthalten;
- e) Verifizieren, daß die angestammten Übersetzungen der elementaren Wörter in den wiederaufgefundenen Dokumenten dieselbe linguistische Struktur haben wie die elementaren Wörter des Suchbegriffs; und
- f) Verwerfen der wiederaufgefundenen Dokumente, wenn die abgestammte Übersetzung der elementaren Wörter in den Dokumenten nicht dieselbe linguistische Struktur hat wie die elementaren Wörter des Suchbegriffs.

[0006] Die Erfindung bezieht sich auf das Durchsuchen von Datenbanken, die über ein Netzwerk, wie etwa das Internet, verteilt und über dieses zugänglich sind. Diese Erfindung versetzt einen Benutzer in die Lage, eine Abfrage zu erzeugen, die Suchbegriffe und Ausdrücke in ihrer Muttersprache verwendet, und festzulegen, daß die Suchergebnisse Dokumente in anderen Sprachen enthalten können. Mit der Abfrage zeigt der Benutzer die Zielsprache an, in der die Ergebnisse akzeptiert werden. Das System verarbeitet anschließend die Abfrage mit Hilfe computerlinguistischer Techniken und verifiziert die Genauigkeit der erhaltenen Ergebnisse im Bezug auf ihre Sprache und die linguistische Struktur der Ausgangssuchbegriffe. Bei einem Ausdruck, der aus mehreren Worten besteht, können sämtliche Kombinationen automatisch verifiziert werden.

[0007] Ein Beispiel eines Verfahrens und einer Vorrichtung gemäß der Erfindung wird nun unter Bezugnahme auf die beiliegenden Zeichnungen beschrieben. In diesen ist/sind:

[0008] **Fig. 1** ein Blockschaltbild, das die Bestandteile und den Vorgangsablauf eines Informations-Wiederauffindungssystems der bevorzugten Ausführungsform der vorliegenden Erfindung darstellt;

[0009] **Fig. 2-Fig. 3** ein einfaches und komplexes Abfragefenster, wie sie bei der bevorzugten Ausführungsform der Erfindung verwendet werden können;

[0010] **Fig. 4** und **Fig. 5** ein einfaches und ein komplexes Ergebnissenfenster, wie sie bei der bevorzugten Ausführungsform der vorliegenden Erfindung Verwendung finden können;

[0011] **Fig. 6-Fig. 7** ein Fenster, das eine Benutzerschnittstelle zeigt, um die mehrsprachige Informations-Wiederauffindung festzulegen, wie es bei der bevorzugten Ausführungsform der vorliegenden Erfindung ausgeführt werden kann;

[0012] **Fig. 8** eine diagrammartige Darstellung des Verfahrens für die mehrsprachige Informations-Wiederauffindung, wie sie bei der bevorzugten Ausführungsform der vorliegenden Erfindung ausgeführt werden kann;

[0013] **Fig. 9** eine diagrammartige Darstellung eines "Übersetzungs- und Stammbildungs-"Schrittes des Verfahrens von **Fig. 8**, wie es bei der bevorzugten Ausführungsform der vorliegenden Erfindung ausgeführt werden kann; und

[0014] **Fig. 10** eine Darstellung eines Systems auf Computerbasis, wie es verwendet werden kann, um die bevorzugte Ausführungsform der vorliegenden Ausführungsform auszuführen.

[0015] Die vorliegende Erfindung bezieht sich auf eine Schnittstelle für die Erzeugung von Abfragen auf und das Wiederauffinden von Informationen aus einer oder mehreren Informationsquellen. Die bevorzugte Ausführungsform wird in einem Informations-Wiederauffindungssystem ausgeführt, das Softwarevermittler bereitstellt, die mehrere heterogene Datenbanken abfragen, Informationen rekonstruieren und angepaßte Berichte erstellen können. Eine Beschreibung des Systems folgt unten, wobei eine umfangreichere Beschreibung, wie etwa zum Zeitpunkt der Einreichung dieser Anmeldung unter folgender URL (Uniform Resource Locator) im Internet zu finden ist: (<http://xrce.xerox.com/research/ct/research/cbkb.html>.) Die vorliegende Erfindung baut auf dem System auf, indem eine Einrichtung bereitgestellt wird, um dynamische komplexe Abfragen zu erzeugen. Bei einer dynamischen komplexen Abfrage sind die Abfrageattribut-Spezifikation dynamisch und nicht statisch, so daß die Ergebnisse (oder Teilergebnisse) einer Abfrage als Attribut-Spezifikation für eine weitere Abfrage verwendet werden können. Weiterhin wird die Einrichtung dem Benutzer in einer sehr intuitiven Art zur Verfügung gestellt, indem hinreichend bekannte grafische Drag-and-Drop-Schnittstellenverfahren zur Anwendung kommen.

[0016] Die bevorzugte Ausführungsform der vorliegenden Erfindung eignet sich für die Verwendung auf Computersystemen, die in der Lage sind, auf des Internet unter Verwendung eines geeigneten Java[®]-Browsers zuzugreifen. Die unterschiedlichen Softwareprogramme, die bei der Ausführung verwendet werden, sind in der hinlänglich bekannten Java-Programmiersprache beschrieben.

[0017] **Fig. 1** stellt dar, wie Informationsanfragen im Informations-Wiederauffindungssystem verarbeitet werden. Unter Bezugnahme auf **Fig. 1** fragt ein Benutzer **101** Informationen durch einen Bedarfsvermittler **102** ab. Der Bedarfsvermittler **102** bildet normalerweise eine Benutzerschnittstelle, die einem Benutzer bereitgestellt wird. Bei der bevorzugten Aus-

führungsform ist der Bedarfsvermittler ein Java[®]-Applet, das auf das Computersystem des Benutzers heruntergeladen wird und auf das unter Verwendung eines in geeigneter Weise konfigurierten Browsers (wie etwa Netscape Navigator[®] mit entsprechenden Java[®]-Softwarefähigkeiten) zugegriffen werden kann.

[0018] Die Bedarfsvermittler **102** stellen anschließend die Informationsanfrage einem Zwischenvermittler **103** zur Verfügung. Die Zwischenvermittler **103** zerlegen Informationsanfragen in voneinander abhängige Unteranfragen **104**. Die Anfragen oder Unteranfragen, die nicht weiter zerlegt werden können, werden Versorgungsvermittlern **105** zugeführt, die auf elektronische Informationsspeicherorte mit Hilfe von Suchmaschinen **106** zugreifen. Die Versorgungsvermittler **105** sind normalerweise als Schnittstelle konfiguriert, die die Protokolle einer bestimmten Suchmaschine verwenden. Es kann beispielsweise ein Versorgungsvermittler bereitgestellt sein, der mit der Suchmaschine Alta Vista[®] arbeitet und Anfragen an diese stellt. Bei der bevorzugten Ausführungsform arbeiten die Zwischenvermittler und die Versorgungsvermittler auf einem Serversystem, das die Informations-Wiederauffindung bereitstellt.

[0019] Die Ergebnisse der Informationsanfrage werden vom Zwischenvermittler **103** empfangen und anschließend wieder zusammengesetzt. Es wird darauf hingewiesen, daß das Informations-Wiederauffindungssystem dem Benutzer erweiterte Abfragemöglichkeiten zur Verfügung stellt, die durch einige Informationsquellen nicht bereitgestellt werden, weshalb das Wiederausammensetzen weiteres Filtern der Ergebnisse beinhaltet, um die erweiterten Abfragemöglichkeiten zu ermöglichen. Die Wiederausammensetzung kann weiterhin Aufgaben, wie etwa das Formatieren der empfangenen Informationen und/oder das Aufstellen einer Sortierung der Ergebnisse mit Bezug auf die Suchabfrage enthalten. Sobald die Wiederausammensetzung abgeschlossen ist, werden die Ergebnisse an die Bedarfsvermittler **102** zur Darstellung für den Benutzer **101** zurückgegeben.

[0020] Es wird darauf hingewiesen, daß Anfragen, Informationen und Ergebnisse vorzugsweise über Merkmalbeschränkungen repräsentiert werden. Anfragen müssen nicht vollständig definiert sein; sie können Teilspezifikationen der angefragten Informationen entsprechen. Weiterhin können Anfragen, die nicht vollständig erfüllt werden können, dennoch Ergebnisse in Gestalt von Teilobjekten liefern, die ursprüngliche Anfragen verfeinern, indem einige der Attribute instanziiert werden oder neue Attribute hinzugefügt werden, wodurch eine Benutzerrückmeldung erzeugt wird.

[0021] Unterschiedliche Aspekte des Informations-Wiederauffindungssystems finden sich in den folgenden PCT-Veröffentlichungen WO 98/48359

(29. Oktober 1998) und WO 98/48361 (29. Oktober 1998).

[0022] Jeder der unterschiedlichen Vermittler, die oben beschrieben wurden, ist vorzugsweise als Softwareprogramme ausgeführt, das in der Java-Programmiersprache geschrieben ist, um auf in geeigneter Weise konfigurierten Computersystemen ausgeführt zu werden.

[0023] Die Informationssystem-Benutzerschnittstelle wird dem Benutzer normalerweise über den Bedarfsvermittler dargeboten, der oben beschrieben wurde. **Fig. 2** ist eine Darstellung eines Abfragefensters zum Spezifizieren einer Abfrage. Unter Bezugnahme auf **Fig. 2** stellt ein Gebiet-Bereich **210** des Abfragefensters **200** eine Einrichtung zum Spezifizieren eines von mehreren vorbestimmten Gebieten für die Abfrage bereit. Das Abfragegebiet bezieht sich auf den Typ von Informationen, nach denen gesucht wird. Beispielsweise kann ein Abfragegebiet ein Typ von Dokumenten (z.B. Bücher oder Aufsätze) oder ein Fachbereich (z.B. Physik oder Opern) sein. Bei der bevorzugten Ausführungsform sind diese Gebiete vorbestimmt und entsprechen einer oder mehreren Informationsquellen, auf die sich die Abfrage richten wird. Die Bereiche **202–204** dienen zum Festlegen eines Suchmusters. Der Bereich **202** dient der Festlegung eines vorbestimmten Attributes (z.B. Titel oder Autor). Der Bereich **203** dient der Festlegung einer Einschränkung (z.B. exakte Inhalte oder Übereinstimmungen), die am Attribut und einem speziellen Schlüsselwort plaziert werden soll. Der Bereich **204** dient zum Festlegen eines gewünschten Schlüsselwortes (Schlüsselwörter), das einen Ausdruck definieren kann, der aus Schlüsselwörtern und Bool'schen Operatoren gebildet wird. Ist das Suchmuster korrekt, bewirkt das Drücken der Taste **205** mit der Aufschrift "Hinzufügen zur Spezifikation", daß das Muster der momentanen Spezifikation der Abfrage hinzugefügt wird, wie es im Spezifikationsbereich **206** angezeigt ist. Ist das Suchmuster nicht korrekt, kann es mit Hilfe einer Rücksetztaste **207** rückgesetzt werden. Der Spezifikationsbereich **206** enthält sämtliche gewünschte Suchmuster, die durch den Benutzer festgelegt werden.

[0024] Die Tasten **208–211** dienen dem Editieren der Festlegung. Die Taste **208** dient dem Editieren des gewählten Suchmusters. Die Taste **209** ist eine Einrichtung zum Entfernen eines Suchmusters. Die Taste **210** ist eine Einrichtung zum Löschen der gesamten Festlegung.

[0025] Schließlich dienen die Tasten **212** und **213** der Steuerung des Sendens bzw. des Abbruchs einer Abfrage.

[0026] **Fig. 3** stellt ein Fenster zum Ausführen einer Unterabfrage innerhalb einer Abfrage, d.h. einer

komplexen Abfrage, dar. Bei einer komplexen Abfrage erzeugt eine Unterabfrage einen Satz von Teilergebnissen. Die erzeugten Ergebnisse sind Teilergebnisse in dem Sinn, daß sie Werte liefern, die im Abfragemuster der Hauptabfrage (oder allgemein bei einer Abfrage höherer Ordnung) verwendet werden. Eine Unterabfrage wird erzeugt, indem eine vorbestimmte Option im Bereich **203**, nämlich die "Antwort von"-Option gewählt wird. Dadurch wird das Fenster von **Fig. 3** angezeigt. Wie es in **Fig. 3** zu erkennen ist, wurde ein neuer Reiter **301** hinzugefügt, um zu kennzeichnen, daß dies eine Unterabfrage ist. Dies versetzt den Benutzer in die Lage, zwischen der Hauptabfrage und der Unterabfrage zu queren. Es wird darauf hingewiesen, daß der übrige Teil des Unterabfragefensters der Hauptabfrage gleicht.

[0027] **Fig. 4** und **Fig. 5** zeigen eine Gesamt- bzw. eine Detailansicht des Ergebnisfensters einer Suche. Unter Bezugnahme auf **Fig. 4** zeigt eine Gesamtdarstellung eine Liste sämtlicher Treffer und einige Gesamtinformationen, wie etwa die Anzahl der Treffer. In **Fig. 5** werden alle Informationsfelder angezeigt, die sich auf einen speziellen "Treffer" beziehen. Ein Hauptbereich **501** zeigt den Textinhalt des Dokumentes. Ein Informationsbereich **502** liefert unterschiedliche Informationen, wie etwa das Datum, die URL, die Informationsquelle, Größe und Titel. Bei dieser Darstellung erfolgt das Navigieren durch die Trefferliste mit Hilfe der Tasten Zurück **503** und Vor **504**, die sich innerhalb des Fensters befinden.

[0028] Wie es oben beschrieben wurde, wäre es wünschenswert, daß es möglich ist Dokumente wiederaufzufinden, die die gewünschten Informationen enthalten, unabhängig von der Sprache der Abfrage. Die vorliegende Erfindung stellt eine derartige Einrichtung durch die Möglichkeit bereit, mehrsprachige Abfragen durchzuführen.

[0029] **Fig. 6** zeigt ein Fenster, bei dem eine mehrsprachige Informations-Wiederauffindung spezifiziert werden kann. Wie im Bezug auf den Rest der Benutzerschnittstelle werden die Fenster, die hier beschrieben werden, dem Benutzer normalerweise über den Bedarfsvermittler dargeboten. Unter Bezugnahme auf **Fig. 6** erfolgt die Steuerung durch eine Schnittstelle, mit der Abfrageoptionen festgelegt werden können, hier das Optionsfeld **601**. Auf andere Optionsfelder kann hier zugegriffen werden, wie etwa ein Ergebnisanzeige-/Ranglistenfeld **602** oder Allgemeine Optionen **603**. Bei einem beliebigen Ereignis läßt ein "Abfragesprache"-Bereich **604** den Benutzer eine Abfragesprache festlegen. Die festgelegte Abfragesprache ist diejenige, in der die Suchausdrücke erzeugt werden. Im allgemeinen ist bei der Einrichtung der vorliegenden Erfindung die gewählte Abfragesprache jene, die dem Benutzer am meisten vertraut ist. Ein zweiter "Abfrageverarbeitungs"-Bereich **605** gestattet es dem Benutzer, Abfrage-Vorverarbei-

tungsoptionen festzulegen. Abfrage-Vorverarbeitungsoptionen sind Operationen, die am Suchbegriff vor dem Durchführen einer Suche ausgeführt werden. Hier dienen die Verarbeitungsoptionen der Stammbildung **606** oder der Übersetzung **607**. Der Übersetzungsoption **607** ist ein Sprachfestlegungsbereich **608** zugeordnet. Im Sprachfestlegungsbereich **608** kann ein Benutzer die Sprachen wählen, in die die Abfrage übersetzt werden kann.

[0030] Weitere dargestellte Abfrageoptionen beinhalten einen Suchoptionsbereich **609** und einen Abfrage-Nachbearbeitungsbereich **610**. Der Suchoptionsbereich **609** enthält Optionen für eine fallbezogene Suche **611**, für das Übereinstimmen mit Wortgrenzen **612** oder zum Suchen mit Webcrawlern **613**. Es wird darauf hingewiesen, daß ein Webcrawler eine spezielle Art einer Suchmaschine ist, die eine umfassende Suche in zahlreichen Informationsquellen durchführt. Im Abfrage-Nachbearbeitungsbereich kann eine Operation zum Ermitteln der Sprache der Ergebnisse **614** festgelegt werden. Schließlich gestatten es die Tasten **615** und **616** einem Benutzer, die festgelegten Optionen entweder anzuwenden oder abzubrechen.

[0031] **Fig. 7** stellt das "Ergebnisanzeige-/Sortierfeld **602** detaillierter dar. Unter Bezugnahme auf **Fig. 7** hat ein "Ergebnisanzeige"-Bereich **700** einen ersten Unterbereich mit dem Titel "In Detailansicht" **701**, der es dem Benutzer gestattet, die Informationen festzulegen, die in der Detailansichtsoption gezeigt werden (wie es in **Fig. 5** oben dargestellt ist). Hier dienen die Optionen dazu, entweder die Hauptergebnisfelder **702** oder sämtliche Ergebnisfelder **703** zu zeigen. Im Bereich **704** "Für komplexe Abfragen" kann ein Benutzer festlegen, lediglich die Hauptabfrage **705** oder die Ergebnisse für alle Unteranfragen **706** zu zeigen.

[0032] Ein "Sortierkriterium"-Bereich **707** gestattet es dem Benutzer festzulegen, wie die Ergebnisse der Abfrage sortiert werden sollen. Eine Untergruppe des Sortierkriteriums-Bereiches **707** ist ein "Vorzugs"-Bereich **708**, der es einem Benutzer gestattet festzulegen, daß die Sortierung auf "Besser erfüllten Vorgaben" **709** basiert, und ein "Ergebnis in"-Bereich **710**, der es einem Benutzer gestattet, die Sprache festzulegen, in der sich die Ergebnisse befinden sollen. Dem Ergebnisbereich **710** ist ein Sprachfestlegungsbereich **711** zugeordnet, in dem der Benutzer die Sprache oder die Sprachen festlegen kann, in dem sich die zurückgesendeten Dokumente befinden sollen. Diese Festlegung bestimmt auch, wie die Ergebnisse sortiert werden. Schließlich gestatten es die Tasten **712** und **713** einem Benutzer, die festgelegten Optionen anzuwenden oder zu löschen.

[0033] **Fig. 8** zeigt eine diagrammartige Darstellung der mehrsprachigen Wiederauffindung, wie sie bei

der vorliegenden Erfindung ausgeführt werden kann. Unter Bezugnahme auf **Fig. 8** wird bei **801** eine Suche nach Dokumenten abgefragt, die über den Mehrwort-Suchbegriff "ABCD" verfügen. Bei Schritt **802** wird der Suchbegriff in "Elementarwörter" aufgeteilt und werden "Stopwörter" (z.B. "und" oder "der, die das") unterdrückt: Hier sind A, B und D Elementarwörter und C ein Stopwort.

[0034] Somit stehen nur die Elementarwörter A, B und D für eine weitere Verarbeitung zur Verfügung. Bei Schritt **803** werden die Elementarwörter A, B und D anschließend übersetzt und deren Stamm gebildet, was zu einer Gruppe von Kombinationen abgestammter und übersetzter Elementarwörter führt.

[0035] Das Übersetzen und die Stammbildung sind in **Fig. 9** weiter dargestellt. Unter Bezugnahme auf **Fig. 9** wird das Elementarwort A beim Verfahrensschritt **901** übersetzt, was zu Übersetzungen A1, A2 und A3 führt. Für jede der einzelnen Übersetzungen A1, A2 wird beim Verfahrensschritt **902** der Stamm gebildet. Stammbildung ist ein Vorgang, bei dem sprachliche Ableitungen eines Wortes gebildet werden. Beispielsweise kann die Stammbildung des Wortes "Focus" zu den Wörtern "focused", "focusing", etc. führen. In jedem Fall führen, wie es in **Fig. 9** dargestellt ist, der Stamm und die übersetzten Elementarwörter zu mehreren Begriffen, die Teil der Suchbegriffe werden.

[0036] Beim Stammbilden und Übersetzen der vorliegenden Erfindung können die Techniken zur Anwendung kommen, wie sie in US-A-5.450.598, 5.581.780, 5.564.058, 5.613.145 und 5.642.522 beschrieben sind.

[0037] Wendet man sich wieder **Fig. 8** zu, so werden Suchen nach Dokumenten durchgeführt, die wenigstens eine der Kombinationen von stammgebildeten und übersetzten Elementarwörtern enthalten. Vorzugsweise werden gleichzeitige Suchen für jede Kombination ausgeführt. Schließlich wird der Verifizierungsvorgangsschritt **804** an den Ergebnissen jeder Suche ausgeführt. Dieser Verifizierungsvorgangsschritt wird verwendet, um sicherzustellen, daß die Suchbegriffe der zurückgesendeten Dokumente dieselbe linguistische Struktur haben, wie der Originalsuchbegriff. Die Techniken, die beim Verifizierungsschritt verwendet werden, können unter Anwendung der Verfahren ausgeführt werden, die im US-Patent No. 5.426.583 für Uribe-Echebarria Diaz De Mendibil mit dem Titel "Automatic Interlingual Translation System" (das 583-Patent) beschrieben sind. Jedoch ist der Verifizierungsschritt der vorliegenden Erfindung weit weniger anspruchsvoll als ein automatisches Sprachübersetzungssystem, weshalb die vorliegende Erfindung darauf ausgelegt ist, weniger vollständige oder komplexe Merkmale zu handhaben, wie sie bei der automatischen Sprachübersetzung im 583-Patent er-

forderlich sind. Es wird darauf hingewiesen, daß die Vorgangsschritte, die im Bezug auf **Fig. 8** und **Fig. 9** ausgeführt werden, normalerweise entweder beim Bedarfsvermittler oder beim Zwischenvermittler oder Kombinationen derselben durchgeführt werden, wie es oben unter Bezugnahme auf **Fig. 1** beschrieben ist.

[0038] Eine der Hauptschwierigkeiten, die sich durch die vorliegende Erfindung ergeben, ist die Verwaltung der zahlreichen Kombinationen, die aus den Stammbildungs- und Übersetzungsschritten resultieren. Das Verwalten derartiger Kombinationen, die intern als voneinander abhängige komplexe Abfragen modelliert werden, ist eines der inhärenten Merkmale des Informations-Wiederauffindungssystems, das von der vorliegenden Erfindung verwendet wird (siehe die zuvor erwähnten PCT-Veröffentlichungen WO 98/48359 (29. Oktober 1998) und WO 98/48361 (29. Oktober 1998)). Beispielsweise kann jede der unterschiedlichen Kombinationen durch das Informations-Wiederauffindungssystem, in dem die vorliegende Erfindung ausgeführt wird, als "Unterabfrage" für eine Hauptabfrage behandelt werden (die eine logische "OR"-Funktion ausführen kann, um die Ergebnisse aller Unteranfragen anzuhäufen).

[0039] Da das Informations-Wiederauffindungssystem der bevorzugten Ausführungsform Abfragen auf begrenzte Fachgebiete (wie etwa Computerwissenschaft) verfolgt, werden im Bezug auf die Zahl der erzeugten Kombinationen fachspezifische Wörterbücher verwendet, um die Menge der möglichen Übersetzungen und somit finaler Kombinationen zu verringern. Derartige fachspezifische Wörterbücher können lediglich Definitionen und Übersetzungen beinhalten, die für dieses spezielle Fachgebiet spezifisch sind. Zudem gestatten Wörterbücher idiomatischer Ausdrücke die Verbesserung und Fokussierung der Übersetzung und verringern somit möglicherweise die Zahl der erzeugten Kombinationen.

[0040] Eine weitere bedeutende Angelegenheit bei der Informations-Wiederauffindung ist die Qualität der erhaltenen Ergebnisse, die durch zwei Standardparameter gemessen wird: Präzision und Erinnerung. Die Präzision ist als die Zahl der wiederaufgefundenen relevanten Dokumente dividiert durch die Gesamtzahl der wiederaufgefundenen Dokumente definiert; die Erinnerung als die Zahl der wiederaufgefundenen relevanten Dokumente dividiert durch die Gesamtzahl der relevanten Dokumente in der Dokumentenbank. Bei der mehrsprachigen Informations-Wiederauffindung leidet die Präzision des finalen Suchbegriffes an sich, dargestellt durch die Kombination abgestammter Übersetzungen, bereits unter den zahlreichen Ergebnissen der Übersetzungs- und Stammbildungsschritte. Ohne Nachbearbeitung nach der Wiederauffindung der Dokumente ist wenigstens die Präzision des Wiederauffindungsergebnisses of-

fensichtlich gering. Daher spielt der letzte Schritt des Verfahrens der vorliegenden Erfindung eine bedeutende Rolle beim Verbessern der Präzision der Abfragen. Von besonderer Bedeutung ist die Tatsache, daß die Verifizierung der korrekten linguistischen Struktur der Suchschlüsselwörter in den wiederaufgefundenen Dokumenten hilfreich ist, unpassende Ergebnisse zu eliminieren.

[0041] Eine weitere Technik zum Verbessern der Präzision wäre die Steuerung der Ergebnisse der Übersetzungs- und Stammbildungsschritte durch den Benutzer. Mit Hilfe dieser Technik kann eine erste Abfrage mehrere Suchmuster erzeugen, die jeweils einer der Kombinationen der übersetzten Stämme entspricht. Es kann ein Suchkanal geöffnet werden, mit dem ein Benutzer die Suchergebnisse untersuchen kann. Sofern der Benutzer mit der Zielsprache vertraut ist, können sie einige der erzeugten Suchbegriffe auf der Basis der empfangenen Ergebnisse unterdrücken, wie etwa durch Steuern und manuelles Bestätigen der erzeugten Suchmuster oder Unterdrücken derselben, wenn diese nicht passend sind.

[0042] Das Computersystem, auf dem die bevorzugte Ausführungsform der vorliegenden Erfindung ausgeführt werden kann, ist unter Bezugnahme auf **Fig. 10** beschrieben. Das Computersystem **10** kann beispielsweise ein Benutzersystem sein, auf dem ein Bedarfsvermittler arbeitet und die Benutzerschnittstelle einem Benutzer dargeboten wird, oder es kann ein Serversystem sein, auf dem unterschiedliche Zwischen- und Versorgungsvermittler arbeiten.

[0043] In jedem Fall besteht unter Bezugnahme auf **Fig. 10** das Computersystem aus mehreren Komponenten, die über einen Bus **1001** gekoppelt sind. Der Bus **1001**, der hier dargestellt ist, ist aus Gründen des besseren Verständnisses vereinfacht. Der Bus **1001** kann aus mehreren parallelen Bussen (z.B. Adreß-, Daten- und Statusbussen) wie auch aus einer Hierarchie von Bussen (z.B. Prozessorbuss, lokaler Bus und einem I/O-Bus) bestehen. In jedem Fall besteht das Computersystem weiterhin aus einem Prozessor **1002** zum Ausführen von Anweisungen, die über den Bus **1001** vom internen Speicher **1003** bereitgestellt werden (es wird darauf hingewiesen, daß der interne Speicher **1003** normalerweise eine Kombination von RAMs und ROMs ist). Der Prozessor **1002** und der Internspeicher-ROM **1003** können einzelne Komponenten oder eine einzige integrierte Vorrichtung, wie etwa ein Application-Specification-Integrated-Circuit-(ASIC-) Chip sein, der entwickelt wurde, um die Funktionsmerkmale der vorliegenden Erfindung zu enthalten.

[0044] Mit dem Bus **1001** sind zudem eine Tastatur **1004** zum Eingeben einer alphanumerischen Eingabe, ein externer Speicher **1005** zum Speichern von

Daten, eine Cursorsteuervorrichtung **1006** zum Betätigen eines Cursors und eine Anzeigeeinrichtung **1007** zum Anzeigen einer visuellen Ausgabe verbunden. Die Tastatur **1004** ist normalerweise eine Standard-QWERTZ-Tastatur, kann jedoch auch ein telefontähnliches Tastenfeld sein. Der externe Speicher **1005** kann ein fest installiertes oder entnehmbares magnetisches oder optisches Plattenlaufwerk sein. Die Cursorsteuervorrichtung **1006** hat normalerweise eine Taste oder einen Schalter, der mit ihr verbunden ist, über den die Ausführung bestimmter Funktionen programmiert werden kann. Die Netzwerkverbindung **1008** kann verwendet werden, um Daten, die eine Bitmap-Darstellung eines Mediums enthalten, zu empfangen und/oder zu senden.

Patentansprüche

1. Verfahren zum mehrsprachigen Abfragen in einem Informations-Wiederauffindungssystem, wobei das Verfahren folgende Schritte enthält:

- a) Empfangen eines Suchbegriffes;
- b) Extrahieren eines oder mehrerer elementarer Wörter aus dem Suchbegriff (**801**);
- c) Bestimmen linguistischer Ableitungen von Übersetzungen für jedes des einzelnen oder der mehreren elementaren Wörter (**802**);
- d) Suchen nach und Wiederauffinden von Dokumenten, die Kombinationen der linguistischen Ableitungen von Übersetzungen enthalten (**803**); und
- e) Verifizieren, dass die linguistischen Ableitungen von Übersetzungen von elementaren Wörtern in den wiederaufgefundenen Dokumenten dieselbe linguistische Struktur haben wie die elementaren Wörter des Suchbegriffes (**804**);
- f) Verwerfen wiederaufgefundener Dokumente, wenn die linguistischen Ableitungen von Übersetzungen elementarer Wörter in den Dokumenten nicht dieselbe linguistische Struktur hat wie die elementaren Wörter des Suchbegriffes.

2. Verfahren nach Anspruch 1, bei dem vor dem Such- und Wiederauffindungsschritt der Schritt des Erzeugens von Kombinationen der linguistischen Ableitungen von Übersetzungen ausgeführt wird; und der Such- und Wiederauffindungsschritt aus der Suche nach Dokumenten besteht, die über jede der erzeugten Kombinationen linguistischer Ableitungen von Übersetzungen verfügen.

3. Verfahren nach Anspruch 2, bei dem gleichzeitig mit dem Such- und Wiederauffindungsschritt folgende Schritte ausgeführt werden:

Versorgen des Benutzers mit vorläufigen Suchergebnissen, die einer speziellen Kombination linguistischer Ableitungen von Übersetzungen zugeordnet sind, die man während des Such- und Wiederauffindungsschrittes ermittelt; und Empfangen eines Signals vom Benutzer, um diesen Such- und Wiederauffindungsschritt für die spezielle

Kombination linguistischer Ableitungen von Übersetzungen abubrechen.

4. Verfahren nach einem der vorhergehenden Ansprüche, bei dem der Schritt des Bestimmens linguistischer Ableitungen von Übersetzungen für jedes der elementaren Wörter aus folgenden Schritten besteht:

- c1) Übersetzen eines elementaren Wortes in eine Zielsprache; und
- c2) Ermitteln linguistischer Ableitungen des übersetzten elementaren Wortes.

5. System für mehrsprachige Informationswiederauffindung, enthaltend:

einen Bedarfsvermittler (**102**), der bereitstellt: eine Benutzerschnittstelle, um es einem Benutzer zu ermöglichen, einen Suchbegriff anzugeben; eine Abfrage-Angabeschnittstelle, die es dem Benutzer gestattet anzugeben, dass eine Abfrage von einer Ursprungssprache in eine Zielsprache übersetzt werden soll; einen Zwischenvermittler (**103**), der bereitstellt: ein Begriffs-Zerlegeelement zum Zerlegen eines Suchbegriffes in elementare Wörter und Stopwörter; ein Übersetzungselement zum Übersetzen elementarer Wörter in eine Zielsprache; ein Element linguistischer Ableitung zum Ermitteln linguistischer Ableitungen jedes der übersetzten elementaren Wörter; ein Suchen-Spezifikationselement zum Erzeugen von Suchanfragen für jede Kombination linguistischer Ableitungen übersetzter elementarer Wörter; und einen Versorgungsvermittler (**105**) zum Empfangen der Suchanfragen und Anfragen sowie Ermitteln von Dokumenten von den Informationsquellen in Abhängigkeit von den Suchanfragen; und ein Verifikationselement zum Verifizieren, daß die linguistischen Ableitungen übersetzter elementarer Wörter in den ermittelten Dokumenten dieselbe linguistische Struktur wie die elementaren Wörter des durch den Benutzer festgelegten Suchbegriffes haben, dadurch gekennzeichnet, dass das Verifikationselement so eingerichtet ist, dass es ermittelte Dokumente, deren linguistischen Ableitungen übersetzter elementarer Wörter nicht dieselbe linguistische Struktur wie die elementaren Wörter des durch den Benutzer festgelegten Suchbegriffes haben.

6. Programmspeichervorrichtung, die durch eine Maschine lesbar ist, in der ein Programm von Befehlen enthalten ist, die von der Maschine ausgeführt werden können, um ein Verfahren gemäß einem der Ansprüche 1 bis 4 durchzuführen.

Es folgen 9 Blatt Zeichnungen

Anhängende Zeichnungen

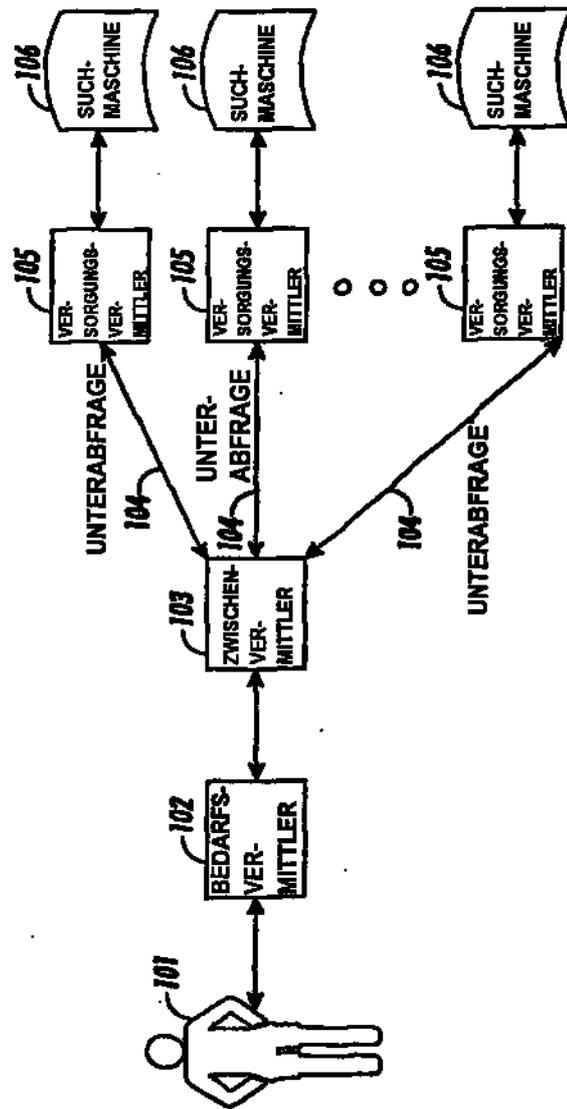


FIG. 1

The image shows a search interface with the following elements:

- 200**: Main search area header.
- 201**: Search criteria field: "ABFRAGE BETRIFFT BUECHER/AUFSAETZE".
- 202**: Search terms field: "TITEL".
- 203**: Search operators: "ENTHAELT".
- 204**: Logical connector: "-UND/ODER-".
- 205**: Action buttons: "RUECKGEBEN" and "HINZUFUEGEN ZUR SPEZIFIKATION".
- 206**: Search results list area containing:
 - TITEL ENTHAELT VERMITTLER
 - TITEL ENTHAELT NETZ ODER INTERNET
 - DATUM NACH 08.95
- 208**: "BEANWORTEN" button.
- 209**: "ENTFERNEN" button.
- 210**: "LOESCHEN" button.
- 211**: "UNTERABFRAGE ABBRECHEN" button.
- 212**: "SENDEN" button.
- 213**: "ABBRUCH" button.

ALLGEMEINE OPTIONEN: FALLBEZOGENE SUCHE
 SUCHE AUCH UEBER WEBCRAWLER

© 1998 XEROX CORPORATION (17 U.S.C. 401)

FIG. 2

301

HAUPTANFRAGE		UNTERANFRAGE 1	
ABFRAGE BETRIFFT	OPERN		
KOMPONIST	ENTHAELT	WAGNER	
RUECKSETZEN		HINZUFUEGEN ZUR SPEZIFIKATION	
BEARBEITEN	ENTFERNEN	LOESCHEN	UNTERABFRAGE ABBRECHEN
ALLGEMEINE OPTIONEN:		<input type="checkbox"/> FALLBEZOGENE SUCHE	
		<input type="checkbox"/> SUCHE AUCH UEBER WEBCRAWLER	
SENDEN		ABBRUCH	

© 1998 XEROX CORPORATION (17 U.S.C. 401)

FIG. 3

ERGEBNISSE DER ABFRAGE 1 (BUECHER/AUFSATZE): DATUM NACH 9506...

ZEIGE UEBERSICHT NACH TITEL TREFFERZAHL: 24

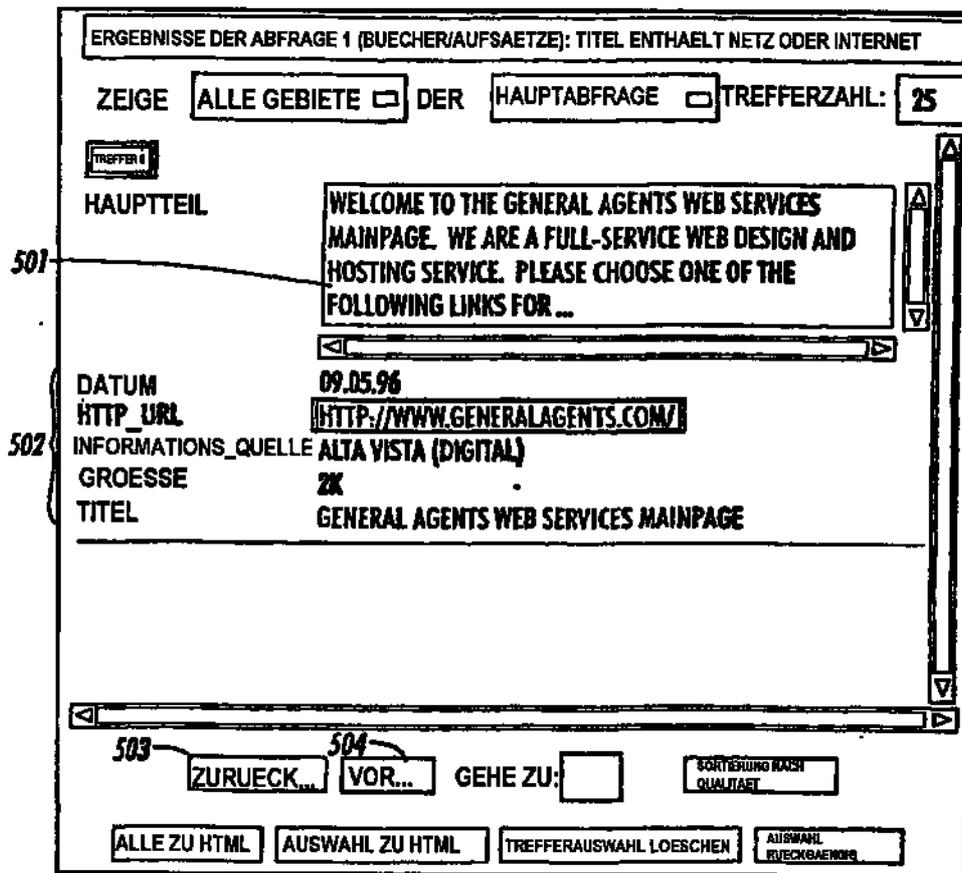
TREFFER 1: GENERAL AGENTS WEB SERVICE MAINPAGE
TREFFER 2: IWAH - INTELLIGENT WEB AGENTS/HOUSTON
TREFFER 3: PERSISTENT JAVA AND MOBILE WEB AGENTS
TREFFER 4: REAL ESTATE AGENTS WEB PAGES CREATED BY INNOVATIVE.
TREFFER 5: INTERNET AGENTS: SPIDERS, WANDERERS, BROKERS, AND 'BOTS/FAH-CHENG
TREFFER 6: PROGRAMMING INTELLIGENT AGENTS FOR THE INTERNET/MARK WATSON
TREFFER 7: GENERAL AGENTS WEB SERVICES MAINPAGE
TREFFER 8: INTERACTIVE HOME BUYING ON WEB - REAL ESTATE AGENTS & COMPANIES
TREFFER 9: HTML WEB PAGE DESIGN AND MARKETING AGENTS
TREFFER 10: ETHICAL WEB AGENTS
TREFFER 11: TAHITIWEB - TAHITI AND ITS ISLANDS WEB PAGES - TRAVEL AGENTS
TREFFER 12: WWW-VRML OCT-DEC1995: INTERACTIVE AGENTS AND THE WEB
TREFFER 13: WWW-VRML OCT-DEC 1995: RE: INTERACTIVE AGENTS AND THE WEB
TREFFER 14: DARTFLOW: A WORKFLOW MANAGEMENT SYSTEM ON THE WEB USING TRANSP
TREFFER 15: TRAVEL_AGENTS INTERNET MAILING LIST
TREFFER 16: INTERNET AGENTS (BOOK OUTLINE)

ALPHABETISCHE SORTIERUNG SORTIERUNG NACH QUALITAET

ALLE ZU HTML AUSWAHL ZU HTML TREFFERAUSWAHL LOESCHEN AUSWAHL RUECKGAENKIG

© 1998 XEROX CORPORATION® (17 U.S.C. 401)

FIG. 4



© 1998 XEROX CORPORATION (17 U.S.C. 401)

FIG. 5

≡
□ □
WISSENSVERMITTLER BASIEREND AUF EINSCHRAENKUNGEN - SITZUNGSFENSTER

STEUERUNG ANSICHT ABFRAGE ERGEBNISSE
HILFE

WISSENSVERMITTLER BASIEREND AUF EINSCHRAENKUNGEN - SITZUNGSFENSTER

ANONYM: ABFRAGE IST NOCH NICHT GEGENDET

601
602
603

ABFRAGEOPTIONEN
ERGEBNISANZEIGE/SORTIERUNG
ALLGEMEINE OPTIONEN

GEWAEHLTE OPTIONEN FUER ABFRAGE 1

ABFRAGESPRACHE:

ENGLISCH
 FRANZOESISCH
 BELIEBIGE SPRACHE

ABFRAGE-VORVERARBEITUNG:

KEINE

STAMMBILDUNG 606

UEBERSETZUNG IN 607

ENGLISCH
 FRANZOESISCH

608

SUCHOPTIONEN:

FALLBEZOGENE SUCHE 611
 WORTUEBEREINSTIMMUNGSGRENZEN 612
 SUCHE AUCH MIT WEBCRAWLERN 613

ABFRAGE-NACHBEARBEITUNG:

ERMITTLE SPRACHE FUER ERGEBNISSE

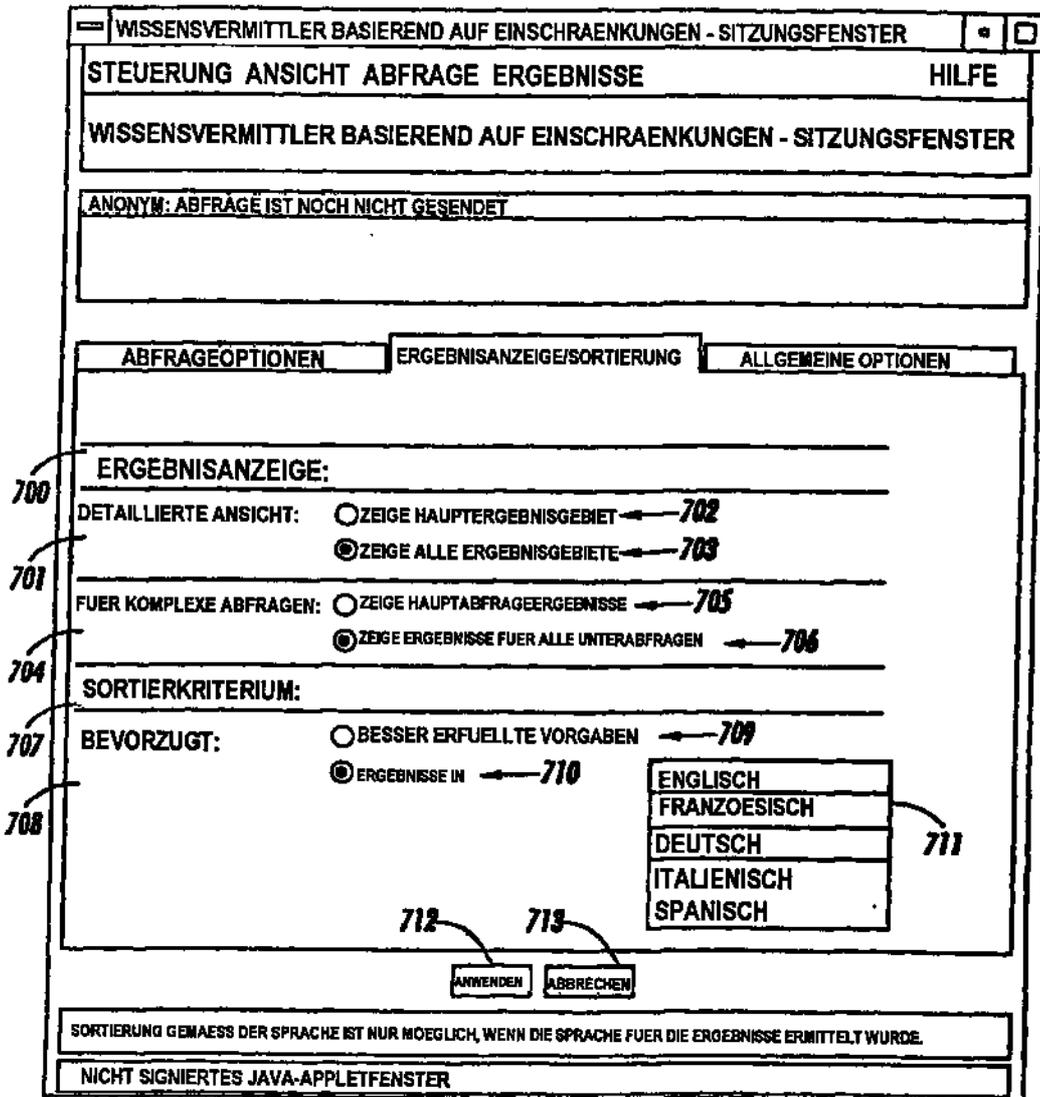
ANWENDEN
ABBRECHEN

ABFRAGEOPTIONEN KOENNEN FUER JEDE ABFRAGE SPEZIFIZIERT WERDEN

NICHT SIGNIERTES JAVA-APPLETFENSTER

© 1998 XEROX CORPORATION (17 U.S.C. 401)

FIG. 6



© 1998 XEROX CORPORATION (17 U.S.C. 401)

FIG. 7

FIG. 8

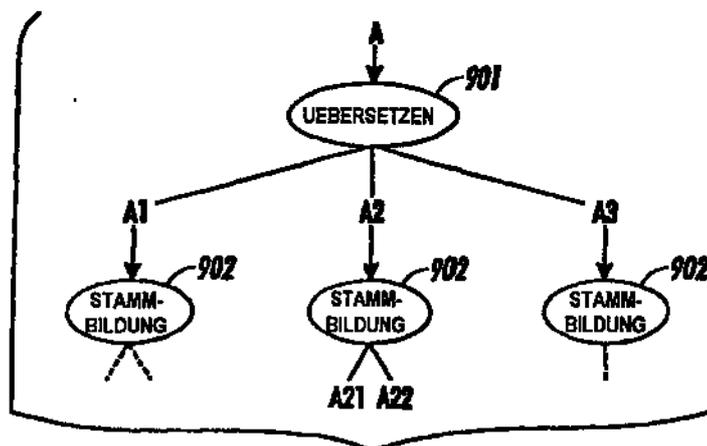
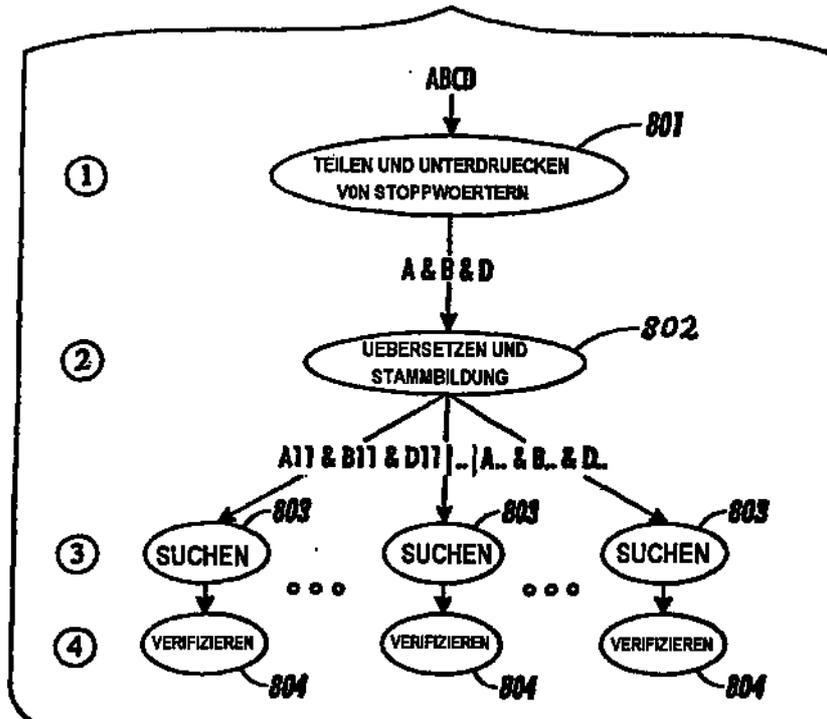


FIG. 9

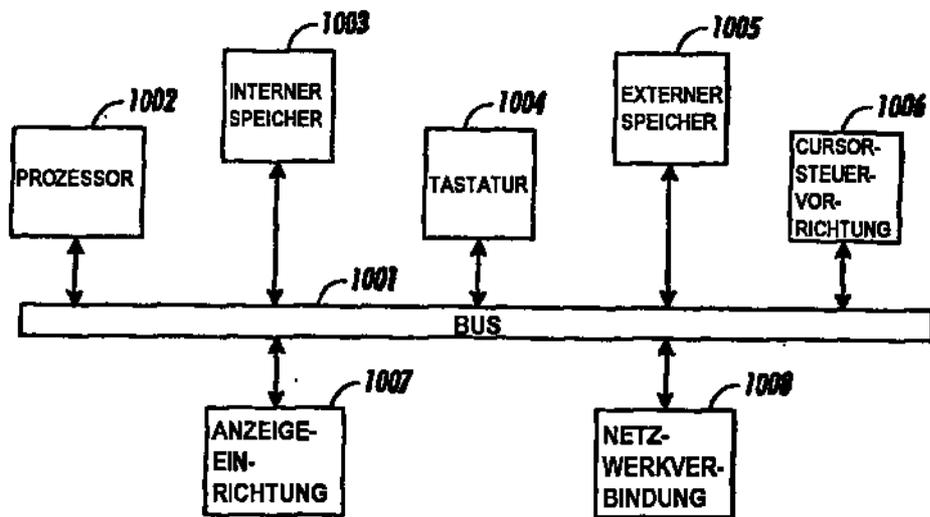


FIG. 10