





ZW), eurasisches (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

— vor Ablauf der für Änderungen der Ansprüche geltenden Frist; Veröffentlichung wird wiederholt, falls Änderungen eintreffen

**Veröffentlicht:**

— mit internationalem Recherchenbericht

Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

---

**(57) Zusammenfassung:** Ein Recherchensystem und ein Verfahren zur Ermittlung von Informationen aus einer Datenbank, insbesondere aus dem World-Wide-Web, sollen zur Erzielung von gleichbleibend hohen Recherchequalitäten in besonderem Masse geeignet sein. Dazu ist erfindungsgemäss die gemeinsame und/oder gleichzeitige Nutzung einer Mehrzahl von Suchmaschinen (2) vorgesehen, wobei in einem Integrationsmodul (6), das die von den Suchmaschinen (2) gelieferten, jeweils eine Mehrzahl von Ergebnisdatensätzen (E) umfassenden Ergebnisdaten in ein gemeinsames Datenformat konvertiert, unter Rückgriff auf in einem Speichermodul (14) hinterlegte Priorisierungsparameter (P) die Zuordnung eines Relevanzkennwerts zu jedem Ergebnisdatensatz (E) vorgesehen ist.

## Beschreibung

1  
2  
3  
4 Recherchensystem und Verfahren zur Ermittlung von Informationen aus einer  
5 Datenbank, insbesondere aus dem World-Wide-Web

6  
7 Die Erfindung bezieht sich auf ein Recherchensystem zur Ermittlung von  
8 Informationen aus einer Datenbank unter Nutzung einer Mehrzahl von  
9 Suchmaschinen. Sie betrifft weiter ein Verfahren zur Ermittlung von Informationen  
10 aus einer Datenbank, insbesondere aus dem World-Wide-Web, unter Nutzung  
11 eines derartigen Recherchensystems.

12  
13 In komplexen Datenbanken oder auch im weltumspannenden Computernetz  
14 („World-Wide-Web“) wird eine enorme Menge an Informationen bereitgehalten,  
15 die ein Benutzer zu Recherchezwecken mehr oder weniger gezielt abrufen kann.  
16 Um dabei Informationen aus der großen Informationsmenge heraus gezielt  
17 nutzbar machen zu können, kommen so genannte Suchmaschinen zum Einsatz,  
18 die gerade bei der Informationsbeschaffung aus dem World-Wide-Web teilweise  
19 erhebliche Verbreitung erfahren haben. Bei derartigen Suchmaschinen wird dem  
20 Benutzer üblicherweise über ein Ein-/Ausgabemodul ein Anfragenfenster zur Ver-  
21 fügung gestellt, über das gezielte Such- oder Recherchebegriffe vorgegeben wer-  
22 den können. Anschließend durchsucht die Suchmaschine die Informationsbasis  
23 der Datenbank bzw. des World-Wide-Web nach geeigneten Schlüssel- oder Key-  
24 wörtern. Die daraufhin aufgefundenen Antwortdatensätze werden üblicherweise  
25 von der jeweiligen Suchmaschine hinsichtlich ihrer Relevanz für den vorgegebenen  
26 Suchauftrag kategorisiert und dem Benutzer in der Art einer Trefferliste in  
27 nach ihrer Relevanz geordneter Reihenfolge zur Verfügung gestellt.

28  
29 Zwischenzeitlich haben sich eine Vielzahl derartiger Suchmaschinen etabliert, die  
30 sich hinsichtlich ihrer Leistungsfähigkeit, Zuverlässigkeit und/oder der  
31 berücksichtigten Informationsbasis in der Art einer thematischen Spezialisierung  
32 voneinander unterscheiden können. Gerade weil aber aufgrund der Vielzahl der

1 beispielsweise im World-Wide-Web angebotenen Informationen beim Betrieb  
2 einer derartigen Suchmaschine eine gewisse Schwerpunktbildung oder  
3 Spezialisierung einzelner Suchmaschinen erforderlich sein kann, kann die Qualität  
4 der Suchergebnisse abhängig von der vorgegebenen Suchanfrage für einzelne  
5 Suchmaschinen in gewissen Bandbreiten variieren.

6  
7 Der Erfindung liegt daher die Aufgabe zugrunde, ein Recherchensystem zur  
8 Ermittlung von Informationen aus einer Datenbank, insbesondere aus dem World-  
9 Wide-Web, anzugeben, das zur Erzielung einer gleichbleibend hohen  
10 Recherchequalität im besonderem Maße zur gemeinsamen und/oder  
11 gleichzeitigen Nutzung einer Mehrzahl von Suchmaschinen besonders geeignet  
12 ist. Des Weiteren soll ein Verfahren zur Ermittlung von Informationen aus einer  
13 Datenbank, insbesondere aus dem World-Wide-Web, unter Nutzung eines  
14 derartigen Recherchensystems angegeben werden.

15  
16 Bezüglich des Recherchensystems wird diese Aufgabe erfindungsgemäß gelöst  
17 mit einem Integrationsmodul, das die von den Suchmaschinen gelieferten, jeweils  
18 eine Mehrzahl von Ergebnisdatensätzen umfassenden Ergebnisdaten in ein  
19 gemeinsames Datenformat konvertiert und anschließend unter Rückgriff auf in  
20 einem im Speichermodul hinterlegte Priorisierungsparameter jedem  
21 Ergebnisdatensatz einen Relevanzkennwert zuordnet.

22  
23 Die Erfindung geht dabei von der Überlegung aus, dass zur Erreichung  
24 gleichbleibend hoher Recherchequalitäten grundsätzlich eine Mehrzahl von  
25 alternativen Suchmaschinen eingesetzt werden sollte, so dass  
26 suchfragenabhängige Qualitätsmängel wechselweise ausgeglichen werden  
27 können. Gerade beim Einsatz einer Mehrzahl von Suchmaschinen für eine  
28 Suchanfrage ist jedoch damit zu rechnen, dass eine große Vielzahl an  
29 Recherchetreffern oder Ergebnisdatensätzen von der Gesamtheit der  
30 Suchmaschinen generiert wird. Für eine benutzerfreundliche und gezielte  
31 Auswertung dieser Ergebnisse sollte eine allen Ergebnissen gemeinsame  
32 Relevanzkennung vergeben werden, so dass in für den Benutzer

1 nachvollziehbarerweise eine systematische Ordnung sämtlicher  
2 Rechercheergebnisse nach vorgegebenen Relevanzkriterien erfolgen kann.  
3 Insbesondere sollte dabei ermöglicht sein, die nach vorgegebenen Kriterien als  
4 relevantest erachteten Ergebnisse zuerst zu präsentieren, so dass der Benutzer  
5 sich auf die wesentlichen Rechercheergebnisse konzentrieren kann. Um dies zu  
6 ermöglichen, ist zunächst die Konvertierung der eingehenden Datensätze in ein  
7 gemeinsames Datenformat vorgesehen, so dass eine gemeinsame  
8 Datenauswertung und -strukturierung ermöglicht wird. Anschließend ist die  
9 Gliederung der aufgefundenen Ergebnisdatensätze im Sinne einer gemeinsamen  
10 Relevanzstruktur vorgesehen. Diese Funktion der Zusammenfügung und Konso-  
11 lidierung von Ergebnisdatensätzen aus verschiedenen Quellen wird durch das  
12 Integrationsmodul erfüllt.

13  
14 Die hierarchische Kategorisierung der aufgefundenen Ergebnisdatensätze durch  
15 Zuordnung eines spezifischen Relevanzkennwerts kann dabei nach vom Benutzer  
16 oder einem Betreiber vorgegebenen Priorisierungskriterien erfolgen. Insbesondere  
17 könnte vorgesehen sein, bei der Auswertung der gelieferten Ergebnisse bestimm-  
18 ten Suchmaschinen generell und standardisiert eine erhöhte Relevanz zuzuord-  
19 nen, beispielsweise weil für einen bestimmten Typus einer Suchanfrage bekannt  
20 ist, dass bestimmte Suchmaschinen in der Regel vergleichsweise hohe  
21 Recherchequalität liefern. Um dies zu ermöglichen, sind die  
22 Priorisierungsparameter vorteilhafterweise abhängig von der jeweiligen  
23 Suchmaschine gewählt, so dass beispielsweise anhand des für eine bestimmte  
24 Suchmaschine hinterlegten Priorisierungskennwerts festgestellt werden kann,  
25 dass ein von dieser Suchmaschine gelieferter Ergebnisdatensatz mit einem  
26 vergleichsweise hohen Relevanzkennwert versehen wird.

27  
28 Die Vorgabe einer spezifischen Priorisierung bei der Berücksichtigung der  
29 Suchmaschinen kann dabei vom Benutzer nach Art und Umfang vorgegeben sein.  
30 Innerhalb der von einer Suchmaschine gelieferten Ergebnisdatensätze kann die  
31 so genannte interne Priorisierung anschließend auf der Grundlage der von der  
32 Suchmaschine bereits gelieferten Priorisierung erfolgen. Dazu wird in weiterer

1 oder alternativer vorteilhafter Ausgestaltung der Relevanzkennwert für den je-  
2 weiligen Ergebnisdatensatz unter Berücksichtigung eines von der jeweiligen  
3 Suchmaschine bereitgestellten Ausgangsrelevanzkennwerts vom  
4 Integrationsmodul ermittelt.

5

6 Dabei kann insbesondere in der Art einer „dynamischen Priorisierung“ eine Nor-  
7 malisierung von den jeweiligen Suchmaschinen bereits mitgelieferter  
8 Ausgangsrelevanzkennwerte vorgesehen sein, bei der der von der jeweiligen  
9 Suchmaschine als relevantest angesehene Ergebnisdatensatz mit einem  
10 vorläufigen Relevanzkennwert von beispielsweise 100 % versehen wird, wobei die  
11 von der jeweiligen Suchmaschine bereits als vergleichsweise weniger relevant  
12 eingestuft weiteren Ergebnisdatensätze mit einem entsprechend abgestuften  
13 Relevanzkennwert versehen werden. Bei dieser Art der Auswertung und  
14 Priorisierung ist somit in der Art einer Gleichberechtigung sämtlicher eingesetzter  
15 Suchmaschinen ein direkter Quervergleich einzelner Ergebnisdatensätze  
16 ermöglicht. Zusätzlich oder alternativ kann aber auch vorgesehen sein, dass  
17 bereits bei der Normalisierung, also bei der Bildung der Bezugsgröße für den von  
18 der jeweiligen Suchmaschine als relevantest angesehenen Ergebnisdatensatz, ein  
19 suchmaschinenspezifischer Kennwert mit vorgegeben wird, durch den der  
20 Relevanzkennwert des von der jeweiligen Suchmaschine als relevantest  
21 angesehenen Ergebnisdatensatzes suchmaschinenspezifisch vorgegeben wird.  
22 Durch eine derartige suchmaschinenspezifische Vorgabe sind somit  
23 Relativwichtungen zwischen einzelnen Suchmaschinen möglich, in die  
24 insbesondere auch Erfahrungswerte über die generelle Zuverlässigkeit der  
25 jeweiligen Suchmaschine und/oder auch themen- oder suchanfragenabhängige  
26 Erkenntnisse über die jeweilige Suchmaschine mit einfließen können.

27

28 Das Integrationsmodul ist zur gemeinsamen Auswertung der von den  
29 verschiedenen Suchmaschinen gelieferten Ergebnisdatensätzen ausgelegt, wobei  
30 zunächst eine Konvertierung der einzelnen Ergebnisdatensätze in ein  
31 gemeinsames und somit gemeinsam auswertbares Datenformat vorgesehen ist.  
32 Dazu ist das Integrationsmodul dafür ausgelegt, die spezifische, von der

1 jeweiligen Suchmaschine eingesetzte Syntax oder deren Datenformat auswerten  
2 und gegebenenfalls umwandeln zu können. Um dies zu ermöglichen, ist das  
3 Integrationsmodul vorteilhafterweise mit einem zweiten Speichermodul  
4 verbunden, in dem Syntaxdaten für die einzelnen Suchmaschinen hinterlegt sind.  
5 Anhand dieser Syntaxdaten kann sodann die suchmaschinenspezifische  
6 Datenkonvertierung vorgenommen werden.

7  
8 In weiterer vorteilhafter Ausgestaltung ist zusätzlich zum Integrationsmodul auch  
9 ein Distributionsmodul vorgesehen, das einen im gemeinsamen Datenformat  
10 eingehenden Suchauftrag geeignet auf die einzelnen zum Einsatz vorgesehenen  
11 Suchmaschinen aufteilt. Dazu ist das Distributionsmodul, das zweckmäßigerweise  
12 über ein gemeinsames Pufferspeichermodul mit dem Integrationsmodul  
13 kommuniziert, vorteilhafterweise dafür ausgelegt, den jeweiligen Suchauftrag  
14 geeignet in eine für die jeweilige Suchmaschine verarbeitbare Syntax oder ein  
15 entsprechendes Datenformat umzuwandeln.

16  
17 Um eine zuverlässige Zuordnung der gelieferten Ergebnisse zum jeweiligen Such-  
18 auftrag auch dann zu ermöglichen, wenn in enger zeitlicher Reihenfolge eine ver-  
19 gleichsweise große Vielzahl an Suchaufträgen zu bearbeiten ist, was insbesonde-  
20 re bei unterschiedlichen Reaktionszeiten der Suchmaschine zu unterschiedlichen  
21 Wartezeiten auf die Ergebnisse und somit zu zeitlichem Überlapp zwischen ver-  
22 schiedenen Suchaufträgen führen könnte, ist dem Integrationsmodul und dem  
23 Distributionsmodul vorteilhafterweise ein gemeinsames Pufferspeichermodul zu-  
24 geordnet. Im Pufferspeichermodul werden zweckmäßigerweise für jeden Such-  
25 auftrag charakteristische Identifikationsdaten, beispielsweise eine Identifikations-  
26 kennummer, hinterlegt, wobei die einzelnen, an die Suchmaschinen direkt ausge-  
27 gebenen spezifizierten Suchaufträge mit einer entsprechenden Kennnummer ver-  
28 sehen sind. Bei den von den Suchmaschinen gelieferten Antworten ist dann im  
29 Integrationsmodul durch einen Rückgriff auf die im Pufferspeicher hinterlegten  
30 charakteristischen Identifikationsdaten eine eindeutige Zuordnung der Ergebnis-  
31 datensätze zum jeweiligen Suchauftrag ermöglicht, so dass eine konsolidierte und  
32 konsistente Auswertung der eingehenden Ergebnisse ermöglicht ist.

1  
2 Das Distributionsmodul ist vorteilhafterweise spezifisch dafür ausgelegt, den  
3 eingehenden Suchauftrag datenseitig in eine für die jeweiligen Suchmaschinen  
4 verarbeitbares Datenformat oder eine Syntax zu konvertieren. Dazu ist auch das  
5 Distributionsmodul vorteilhafterweise mit dem zweiten Speichermodul verbunden,  
6 in dem Syntaxdaten für die Suchmaschinen hinterlegt sind. Durch gezielte  
7 Wartung und Pflege des zweiten Speichermoduls ist somit auf besondere Weise  
8 ermöglicht, gegebenenfalls Aktualisierungen der von den Suchmaschinen  
9 verwendeten Syntax oder auch die Hinzufügung oder Löschung vollständiger  
10 Suchmaschinen auf besonders einfache Weise vorzunehmen.

11  
12 In weiterer vorteilhafter Ausgestaltung ist das Distributionsmodul eingangsseitig  
13 über ein Filtermodul mit einem Ein-/Ausgabemodul verbunden, über das vom  
14 Benutzer einzelne Suchaufträge vorgebar sind. Das Filtermodul kann dabei  
15 insbesondere dafür ausgelegt sein, anhand einstellbarer oder individuell  
16 vorgebar Filterkriterien bestimmte Suchanfragen ganz zu unterdrücken oder  
17 beispielsweise länder- oder kundenspezifische Individualvorgaben hinzuzufügen.  
18 Dabei kann insbesondere auch vorgesehen sein, anhand der Art der Anfrage  
19 gezielt einzelne, als besonders geeignet erkannte Suchmaschinen für die  
20 nachfolgende Bearbeitung des Suchauftrags vorzugeben. Das Filtermodul kann  
21 somit insbesondere auch dafür ausgelegt sein, abhängig von der jeweiligen  
22 Suchanfrage oder von deren thematischem Schwerpunkt eine Vor- oder  
23 Grobauswahl der einzusetzenden Suchmaschinen zu treffen und somit eine  
24 Grobstrategie für die nachfolgende Strategie vorzugeben.

25  
26 Grundsätzlich können das Distributionsmodul und das Integrationsmodul als  
27 eigenständige Einheiten, beispielsweise auf Software- oder auf Hardwarebasis,  
28 ausgeführt sein. Eine besonders kompakte und somit günstige Bauweise ist dabei  
29 erreichbar, indem das Distributionsmodul vorteilhafterweise mit dem  
30 Integrationsmodul in eine gemeinsame Einheit integriert ist.



1 Es ist weit verbreitete Praxis, gerade bei der Durchführung von Recherchen im  
2 Internet oder World-Wide-Web, dass dem Benutzer gemeinsam mit den auf seine  
3 Suchanfrage hin generierten Trefferlisten Werbeanzeigen, so genannte  
4 „Banners“, mit angezeigt werden. Dabei werden in Reaktion auf eine Suchanfrage  
5 von einem auch als „Ad-Server“ bezeichneten Werbemitteldatenbank gemeinsam  
6 mit den Ergebnissen zufällig aus den dort hinterlegten Banners ausgewählte  
7 Datensätze mit übermittelt, die anschließend auf dem Ein-/Ausgabemodul zur  
8 Anzeige gebracht werden. Bei einem derartigen System erfolgt die Hinzufügung  
9 der individuellen Banners jedoch üblicherweise unspezifiziert und insbesondere  
10 ohne Bezug auf das vom Benutzer durch die Suchanfrage vorgegebene Thema,  
11 so dass das Benutzerinteresse am jeweiligen Werbemittel nur vergleichsweise  
12 gering ist. In besonders vorteilhafter Weiterbildung des oben genannten Systems  
13 sind jedoch nunmehr das Distributions- und das Integrationsmodul datenseitig mit  
14 einer Werbemitteldatenbank verbunden, die als Antwortdatensatz auf eine  
15 eingehende Suchanfrage einen hinsichtlich eines Kennungssignals an die  
16 Suchanfrage angepassten Datensatz liefert. Mit anderen Worten: Aufgrund der  
17 ohnehin vorgesehenen, vergleichsweise qualifizierten Auswertung der  
18 vorgegebenen Suchanfrage, die über die entsprechende Konvertierung im  
19 Distributionsmodul ermöglicht wird, ist nunmehr die Erzeugung einer  
20 beispielsweise inhalts- oder themenbezogenen Kennung für die Suchanfrage  
21 ermöglicht, die bei einer Weiterleitung der Suchanfrage auf die  
22 Werbemitteldatenbank eine gezielte Selektion von hinsichtlich des  
23 Kennungssignals, also beispielsweise hinsichtlich des bearbeiteten Themas,  
24 besonders geeigneten Werbedatensatzes ermöglicht. Durch eine derartige  
25 gezielte Zuordnung der als Antwortsignal mitgegebenen Werbemitteldatensätze  
26 ist somit eine deutlich erhöhte Aufmerksamkeit beim Benutzer gegenüber der  
27 eingesetzten Werbemittel erreichbar.

28  
29 Bezüglich des Verfahrens wird die genannte Aufgabe unter Nutzung des  
30 genannten Recherchensystems gelöst, indem die von den Suchmaschinen  
31 gelieferten, jeweils eine Mehrzahl von Ergebnisdatensätzen umfassenden  
32 Ergebnisdaten in ein gemeinsames Datenformat konvertiert werden, wobei

1 anschließend unter Rückgriff auf in einem Speichermodul hinterlegte  
2 Priorisierungsparameter jedem Ergebnisdatensatz ein Relevanzkennwert  
3 zugeordnet wird. Über das Integrationsmodul ist somit vorgesehen, für die  
4 gelieferten Ergebnisdatensätze ein konsolidiertes und konsistentes  
5 Relevanzsystem zu erzeugen, in dem die einzelnen Ergebnisdatensätze relativ  
6 zueinander im Hinblick auf vorgegebene Relevanzkriterien geeignet positioniert  
7 sind.

8  
9 Vorteilhafterweise werden die Priorisierungsparameter dabei abhängig von der  
10 jeweiligen Suchmaschine gewählt, wobei der Relevanzkennwert in alternativer  
11 oder zusätzlicher vorteilhafter Weiterbildung für den jeweiligen Ergebnisdatensatz  
12 unter Berücksichtigung eines von der jeweiligen Suchmaschine bereitgestellten  
13 Ausgangsrelevanzkennwerts ermittelt wird.

14  
15 Vorteilhafterweise wird ein im gemeinsamen Datenformat eingehender  
16 Suchauftrag für jede ausgewählte Suchmaschine in deren spezifisches  
17 Datenformat umgewandelt und anschließend an diese ausgegeben. In alternativer  
18 oder zusätzlicher vorteilhafter Ausgestaltung wird ein über ein Ein-/Ausgabemodul  
19 vorgegebener Suchauftrag vor seiner weiteren Bearbeitung einer Filterung  
20 unterzogen.

21  
22 Vorteilhafterweise wird ein Suchauftrag zusätzlich zu den Suchmaschinen an eine  
23 Werbemitteldatenbank ausgegeben, wobei ein von dieser eingehender  
24 Antwortdatensatz mittels eines Kennungssignals der jeweiligen Suchanfrage  
25 zugeordnet wird. Durch das Kennungssignal ist dabei insbesondere ermöglicht,  
26 den Antwortdatensatz hinsichtlich vorgegebbarer Kriterien, beispielsweise  
27 hinsichtlich verwendeter Inhalte oder Themen, an die jeweilige Suchanfrage  
28 anzupassen, so dass im Ergebnis dem Benutzer gemeinsam mit seinen  
29 Suchergebnissen ein thematisch zur Suchanfrage passender Werbeträger  
30 angeboten wird.

1 Die mit der Erfindung erzielten Vorteile bestehen insbesondere darin, dass durch  
2 die Verwendung des Integrationsmoduls, dass eingehende Ergebnisdatensätze  
3 auf ein gemeinsames Datenformat konvertiert und anschließend eine relative  
4 Gewichtung und Priorisierung der Ergebnisdatensätze untereinander durch  
5 Vergabe eines „globalen“ Relevanzkennwerts eine konsistente gemeinsame  
6 Nutzung einer Mehrzahl von Suchmaschinen ermöglicht ist. Durch das mit dem  
7 Integrationsmodul kommunizierende Distributionsmodul ist weiterhin  
8 gewährleistet, dass unter Rückgriff auf im Speichermodul hinterlegte,  
9 suchmaschinenspezifische Daten wie beispielsweise Syntaxdaten eine gezielte  
10 Ansprache verschiedenartiger Suchmaschinen mit einer gemeinsamen  
11 Suchanfrage ermöglicht ist, wobei ebenso eine konsolidierte gemeinsame  
12 Auswertung der von den angesprochenen Suchmaschinen gelieferten Ergebnisse  
13 unter Berücksichtigung von deren spezifischer Syntax ermöglicht ist.

14  
15 Ein Ausführungsbeispiel der Erfindung wird anhand einer Zeichnung näher  
16 erläutert. Darin zeigen:

17  
18 Fig. 1 schematisch ein Recherchensystem zur Ermittlung von  
19 Informationen aus einer Datenbank,

20  
21 Fig. 2 eine konzeptionelle Darstellung des Recherchensystems nach Fig. 1,

22  
23 Fig. 3 eine schematische Aufstellung der im Recherchensystem nach  
24 Fig. 1 ablaufenden Prozesse,

25  
26 Fig. 4 eine alternative Darstellung für das Gesamtsystem nach Fig. 1,

27  
28 Fig. 5 schematisch die Struktur der Administration des Recherchensystems  
29 nach Fig. 1,

30  
31 Fig. 6 ein Schema für eine Anfrage an eine Mehrzahl von Suchmaschinen,  
32

- 1 Fig. 7 ein Schema für eine verteilte Abfrage und Zusammenführung der  
2 Ergebnisse,
- 3
- 4 Fig. 8 ein Schema für verschiedene Suchmodi im Recherchensystem nach  
5 Fig. 1,
- 6
- 7 Fig. 9 ein Schema für eine Filterung von Suchanfragen,
- 8
- 9 Fig. 10 eine schematische Darstellung der im Recherchensystem nach  
10 Fig. 1 verwendeten Module,
- 11
- 12 Fig. 11 schematisch eine Darstellung der Struktur eines Filter-  
13 /Interpretermoduls,
- 14
- 15 Fig. 12 schematisch eine Struktur eines Integrationsmoduls,
- 16
- 17 Fig. 13 schematisch die Struktur eines Server-Engine-Moduls,
- 18
- 19 Fig. 14 schematisch ein dynamisches Model für einen Suchprozess,
- 20
- 21 Fig. 15 ein Ablaufschema für eine Initialisierung einer Server-Engine,
- 22
- 23 Fig. 16 ein Ablaufschema für die Bearbeitung einer Suchanfrage,
- 24
- 25 Fig. 17 ein Ablaufschema für die Vorbereitung von Adaptern für eine  
26 Suchanfrage, und
- 27
- 28 Fig. 18 ein Schema für eine Admin-Anwendung des Recherchensystems  
29 nach Fig. 1.
- 30
- 31 Das Recherchensystem 1 gemäß Fig. 1 ist zur Ermittlung von Informationen aus  
32 einer komplexen Datenbank, insbesondere aus dem Internet, dem World-Wide-

1 Web oder einem anderen Index, mit besonders hoher Recherchenqualität  
2 ausgelegt. Dazu ist das Recherchensystem 1 datenseitig mit einer Mehrzahl von  
3 Suchmaschinen 2 verbunden, wie die derzeit zur gezielten Durchsuchung des  
4 Internets weit verbreitet im Einsatz sind.

5  
6 Um dabei den gleichzeitigen und gemeinsamen Einsatz einer Mehrzahl von Such-  
7 maschinen 2 zu ermöglichen, umfasst das Recherchensystem 1 in einer zentralen  
8 Auswerteeinheit 4 ein Integrationsmodul 6, dem die von den Suchmaschinen 2  
9 gelieferten Trefferlisten oder Ergebnisdaten in Form von Ergebnisdatensätzen E  
10 zuführbar sind. Das Integrationsmodul 6 stellt dabei sicher, dass eine  
11 gemeinsame, konsistent priorisierte Auswertung der von den verschiedenen  
12 Suchmaschinen 2 gelieferten Ergebnisdatensätze E ermöglicht ist.

13  
14 Dazu ist das Integrationsmodul 6 dazu ausgelegt, die eintreffende  
15 Ergebnisdatensätze E zunächst in ein gemeinsames Datenformat zu konvertieren,  
16 so dass eine konsolidierte Auswertung ermöglicht ist. Um dem Benutzer dabei  
17 eine gezielte Auswertung der möglicherweise großen Anzahl von Informationen zu  
18 ermöglichen, ist die konsolidierte Priorisierung und Relevanzkennung der  
19 einzelnen Ergebnisdatensätze E vorgesehen, wobei nach vorgegebenen  
20 Relevanzkriterien sämtlichen Ergebnisdatensätzen E untereinander vergleichbare  
21 Relevanzkennwerte zugeordnet werden. Um dies zu ermöglichen, ist das  
22 Integrationsmodul 6 mit einem Speichermodul 8 verbunden, in dem für die  
23 Zuweisung der Relevanzkennwerte relevante Priorisierungsparameter P hinterlegt  
24 sind.

25  
26 Bei der Priorisierung kann insbesondere vorgesehen sein, die nach Vorgabe  
27 durch einen Benutzer oder durch einen Betreiber ausgewählte Suchmaschinen 2  
28 grundsätzlich mit einer vergleichsweise hohen Priorisierung und somit einem  
29 vergleichsweise hohen Relevanzkennwert zu versehen. In diesem Fall erfolgt die  
30 Wahl der Priorisierungsparameter abhängig von der jeweiligen Suchmaschine 2.  
31 Alternativ oder zusätzlich kann vorgesehen sein, in der Art einer Normalisierung  
32 die von einer Suchmaschine 2 gelieferten Ergebnisdatensätze E hinsichtlich ihres

1 Relevanzkennwerts zunächst auf denjenigen Ergebnisdatensatz E zu normieren,  
2 der bereits von der jeweiligen Suchmaschine 2 aus mit dem höchsten  
3 Relevanzkennwert versehen wurde. In der Art einer prozentualen Darstellung  
4 könnte dies somit bedeuten, dass der von der jeweiligen Suchmaschine 2 als am  
5 relevantesten erkannte Ergebnisdatensatz E mit einem vorläufigen Relevanzkenn-  
6 wert von 100 % versehen wird, wobei die anderen Ergebnisdatensätze E dieser  
7 Suchmaschine 2 entsprechend ihrem von der Suchmaschine 2 zugewiesenen  
8 Ausgangsrelevanzkennwert einen entsprechend abgestaffelten Relevanzkennwert  
9 erhalten. Wenn dies für sämtliche Suchmaschinen 2 gleichgewichtig durchgeführt  
10 wird, erfolgt somit eine relative Priorisierung der Ergebnisdatensätze E  
11 unterschiedlicher Suchmaschinen 2 im Wesentlichen anhand ihrer relativen  
12 Relevanz im Hinblick auf den von der jeweiligen Suchmaschine 2 als am  
13 relevantesten eingestuften Ergebnisdatensatz E.

14  
15 Zusätzlich kann bei den Priorisierungsparametern P aber auch ein  
16 suchmaschinenspezifischer Kennwert hinterlegt sein, der allgemein oder abhängig  
17 von der jeweiligen Suchanfrage der jeweiligen Suchmaschine 2 einen  
18 Verlässlichkeitskennwert zuordnet.

19  
20 Die Auswerteeinheit 4 umfasst weiterhin ein Distributionsmodul 10, das über einen  
21 Pufferspeicher 12 mit dem Integrationsmodul 6 kommuniziert. Das  
22 Distributionsmodul 10 ist weiterhin mit einem zweiten Speichermodul 14  
23 verbunden, in dem für die Suchmaschinen 2 charakteristische Datensätze,  
24 insbesondere charakteristisch für deren Datenformat oder Syntax, hinterlegt sind.  
25 Eingangsseitig ist das Distributionsmodul 10 über ein Filtermodul 16 mit einem  
26 Ein-/Ausgabemodul 18 verbunden, das seinerseits eingangsseitig mit dem  
27 Integrationsmodul 6 verbunden ist. Das Ein-/Ausgabemodul 18 kann dabei  
28 insbesondere als herkömmliche Bedieneinheit, umfassend eine Tastatur und ein  
29 Ausgabemittel wie einen Bildschirm, ausgestaltet sein.

30  
31 Bei der Bearbeitung einer vom Benutzer über das Ein-/Ausgabemodul 18  
32 vorgegebenen Suchanfrage erfolgt zunächst eine Filterung dieser Anfrage im

1 Filterbaustein 16. Dabei kann beispielsweise die Elimination von Suchanfragen  
2 eines bestimmten Kriterium vorgesehen sein oder es können beispielsweise  
3 länder- oder regionenspezifische Vorgaben für die Bearbeitung der Suchanfrage  
4 berücksichtigt werden. Des Weiteren erfolgt im Filtermodul 16 eine Vorauswahl  
5 der für die jeweiligen Suchanfragen als geeignet erkannten Suchmaschinen 2, so  
6 dass das Filtermodul 16 in der Art einer Grobstrategie bereits die mit der  
7 Suchanfrage endgültig zu betrauenden Suchmaschinen 2 mit vorgibt.

8  
9 Die solchermaßen gefilterte Suchanfrage 2 wird anschließend dem  
10 Distributionsmodul 10 zugeführt, das unter Rückgriff auf die im zweiten  
11 Speichermodul hinterlegten suchmaschinenspezifischen Informationen und  
12 insbesondere Syntaxkennwerte die jeweilige Suchanfrage in eine Vielzahl  
13 individualisierter Suchanfragen in den für die ausgewählten Suchmaschinen 2  
14 passenden Datenformate umsetzt. Anschließend gibt das Distributionsmodul 10  
15 die individualisierten Suchanfragen an die ausgewählten Suchmaschinen 2 weiter,  
16 wobei in der Art einer Protokollierung und zur Verwaltungserleichterung im  
17 Pufferspeicher 12 ein Profil der Suchanfrage gemeinsam mit einer diese  
18 identifizierenden Kennung hinterlegt wird.

19  
20 Die durch die Zuführung der individualisierten Suchanfrage gestarteten  
21 Suchmaschinen 2 absolvieren anschließend ihr Suchprogramm im World-Wide-  
22 Web und stellen ihre Antwortsignale, beispielsweise in Form von Trefferlisten, in  
23 denen die aufgefundenen Ergebnisdaten als Ergebnisdatensätze E nach der  
24 durch die jeweilige Suchmaschine 2 erkannten Relevanz geordnet aufgelistet  
25 sind. Anhand der charakteristischen Identifikationsdaten für die jeweilige  
26 Suchanfrage werden die Ergebnisdatensätze E sodann im Integrationsmodul 6  
27 der jeweiligen Suchanfrage wieder zugeordnet, so dass eine gezielte Auswertung  
28 auch bei zeitlich sich überschneidenden Suchvorgängen ermöglicht ist.

29  
30 Im Integrationsmodul 6 erfolgt zudem, gegebenenfalls ebenfalls unter Rückgriff  
31 auf die im zweiten Speichermodul 14 hinterlegten Syntaxinformationen, eine  
32 Konvertierung der Ergebnisdatensätze E in das gemeinsame Datenformat.

1 Anschließend erfolgt unter Rückgriff auf die im Speichermodul 8 hinterlegten  
2 Priorisierungsparameter P die Erzeugung eines Relevanzkennwerts für jeden  
3 Ergebnisdatensatz E, wobei die solchermaßen erzeugten Relevanzkennwerte  
4 dazu dienen sollen, eine konsolidierte und konsistente Ergebnishierarchie  
5 zwischen den Ergebnisdatensätzen E herzustellen. Im Ergebnis wird somit dem  
6 Benutzer eine homogenisierte Ergebnis- oder Trefferliste erzeugt, bei der die als  
7 relevantest erkannten Ergebnisdatensätze E zuerst gelistet sind. Die Zuordnung  
8 des jeweiligen Relevanzkennwerts unter Rückgriff auf die Priorisierungsparameter  
9 P kann dabei insbesondere im Hinblick auf die jeweilige Suchmaschine 2 erfolgen,  
10 die den betroffenen Ergebnisdatensatz E geliefert hat. Alternativ oder zusätzlich  
11 kann bei der Zuordnung des Relevanzkennwerts auch ein von der jeweiligen  
12 Suchmaschine 2 bereitgestellter Ausgangsrelevanzkennwert mit berücksichtigt  
13 werden.

14  
15 Zusätzlich zu den Suchmaschinen 2 ist das Recherchensystem 1 datenseitig mit  
16 einer Werbemitteldatenbank 20 verbunden. In der Werbemitteldatenbank, auch  
17 als „Ad-Server“ bezeichnet, sind in Form von so genannten „Banners“  
18 Werbeelemente hinterlegt, die von der Werbemitteldatenbank 20 in Reaktion auf  
19 eine eingehende Suchanfrage an das Recherchensystem 1 zurückgeliefert  
20 werden. Das Recherchensystem 1 ist dabei derart ausgestaltet, dass die  
21 Werbemitteldatenbank 20 ausgewählte und spezifizierte Werbeelemente als  
22 Antwortdatensatz auf eine eingehende Suchanfrage zurücksendet, wobei die  
23 Auswahl und Spezifikation inhalts- oder themenbezogen angepasst an die  
24 jeweilige Suchanfrage erfolgt. Die Auswahl der zurückgesandten Werbeelemente  
25 erfolgt dabei dahingehend, dass das jeweilige Werbeelement inhaltlich oder  
26 thematisch einen Bezug zur Suchanfrage aufweist, so dass mit einem  
27 vergleichsweise erhöhtem Interesse des Benutzers an dem jeweiligen  
28 Werbeelement gerechnet werden kann. Zu diesem Zweck übermittelt das  
29 Recherchensystem 1 die jeweilige Suchanfrage gemeinsam mit einem  
30 Kennungssignal an die Werbemitteldatenbank 20, wobei anhand des  
31 mitgelieferten Kennungssignals die Auswahl möglicherweise für den  
32 Antwortdatensatz geeigneter Werbeelemente vorgenommen wird.



1  
2 Fig. 2 verdeutlicht die beteiligten Komponenten und Aktoren im Bezug auf die  
3 (Sub)Systemgrenzen.  
4

## 5 **Rollenbeschreibung**

### 6 **Kunde**

7 „Kunde“ ist jeder Benutzer, der eine Suchanfrage starten kann.  
8

### 9 **TSC Filtermanager**

10 Die Rolle „TSC Filtermanager“ beschreibt den Mitarbeiterkreis, in dessen  
11 Verantwortungsbereich die Filterdefinitionen liegen. Dieser Mitarbeiterkreis wertet  
12 außerdem die Statistikprotokollierung des Recherchensystems aus.  
13

### 14 **TSC Admin**

15 Die Rolle des „Admin“ beschreibt den Mitarbeiterkreis, der für den Betrieb des  
16 Recherchensystems verantwortlich ist.  
17 Er administriert die Konfigurationsdatei (tsc.cfg) des Recherchensystems und  
18 überwacht die Fehlerprotokolle (tsc\_error.log).  
19

## 20 **Beschreibung der externen Systeme**

### 21 **Inhalte**

#### 22 ***Kurzbeschreibung***

23 Hier werden Suchergebnisse bereitgestellt, die auf private und Business-Homepa-  
24 ges zeigen.

#### 25 ***Schnittstelle***

26 Die Schnittstelle ist analog zur WWW-Suche mit Fast.

## 1 **P4P-Server**

### 2 ***Kurzbeschreibung***

3 Die Suchergebnisse des Pay4Performance(P4P)-Servers werden hier  
4 bereitgestellt.

### 5 ***Schnittstelle***

6 Die Schnittstelle des GoTo-Suchcenters wird per HTTP-Request (inklusive  
7 Parameter) angesprochen. Die Ergebnisse werden in XML-Format geliefert.

## 8 **FAST Suchcenter**

### 9 10 ***Kurzbeschreibung***

11 Hier werden internet-weite Suchergebnisse bereitgestellt, die mittels der  
12 Suchmaschine „AllTheWeb“ gefunden wurden.

### 13 14 ***Schnittstelle***

15 Die Schnittstelle des FAST-Suchcenters wird per HTTP-Request (inklusive  
16 Parameter) angesprochen. Die Ergebnisse werden in XML-Format geliefert.

## 17 18 **Ad-Server**

### 19 20 ***Kurzbeschreibung***

21 Der Ad-Server liefert entsprechend der Suchanfrage eine Werbe-Komponente, die  
22 in die Gesamtdarstellung der Suchergebnisse einfließt.

## 23 24 **Recherchensystem Prozesse**

25 In Fig. 3 werden die Prozesse des Recherchensystem in der Übersicht dargestellt  
26 und der Prozessdurchlauf im „Gut“-Fall beschrieben.

## 27 **TSC Konfiguration**

28 Die Konfiguration des Recherchensystem wird über Parameter gesteuert, die in  
29 einer eigenen Datei abgelegt sind. Diese Konfigurationsdatei wird im XML-Format

1 im Dateisystem des Suchcenters abgelegt. Sie kann mit Hilfe eines Standard-Edi-  
2 tors bearbeitet werden.

3 Die geänderte Konfiguration kann nur mit Hilfe des Restarts der „TSC Suche“ Pro-  
4 zesse erreicht werden.

5  
6 Der Prozess „TSC Konfiguration“ beschreibt die notwendigen Arbeitsschritte für  
7 die Administration der Konfigurationsdatei des Recherchensystem.

Rolle	Aktivitäten
TSC Admin	1. Konfiguration (tsc.cfg) editieren 2. Konfiguration (tsc.cfg) auf alle Produktionssysteme verteilen 3. „TSC Restart“ Prozess durchführen

### 9 TSC Ausschlussfilter pflegen

10 Die TSC Ausschlussfilter definieren Worte und Phrasen für die ein Suchstring,  
11 wenn sie in diesem enthalten sind, für bestimmte Kanäle gesperrt wird. Diese  
12 Sperrungen sind landesspezifisch und können zeitabhängig definiert werden.

13 Gewöhnlich werden die Filter mit Hilfe einer eigenen Anwendung (der TSC Admin  
14 Anwendung) gepflegt. Letztendlich ist die Datei der TSC Ausschlussfilter das  
15 Produkt eines Exports aus der „Filter DB“ und liegt im CSV-Format vor. Diese  
16 Datei kann ggf. mit einem Standardeditor bearbeitet werden.

17 Die Aktivierung der geänderten Filter kann nur mit Hilfe des Restart der „TSC  
18 Suche“ Prozesse erreicht.

19 Folgende Arbeitsschritte werden für den Prozess „TSC Ausschlussfilter pflegen“  
20 definiert:

Rolle	Aktivitäten
TSC Filtermanager	<ol style="list-style-type: none"> <li>1. Anmelden an der Anwendung „TSC Admin“</li> <li>2. Möglichkeit zur Ausführung folgender Aktionen: <ol style="list-style-type: none"> <li>a. Filter Aktivieren</li> <li>b. Filter Erstellen</li> <li>c. Filter Ändern</li> <li>d. Filter Löschen</li> </ol> </li> <li>3. Aktivierung der geänderten Filter <ol style="list-style-type: none"> <li>a. Filterdatei (tsc_filter_table) aus „Filter DB“ generieren</li> <li>b. Filterdatei (tsc_filter_table) auf alle Produktionssysteme verteilen</li> <li>c. „Sukzessive“ Neustart der „TSC Suche“ Prozesse</li> </ol> </li> </ol>

1

2 **TSC Restart**

3 Der „TSC Restart“ Prozess kann manuell oder automatisch durch andere Skripte,  
4 wie z. B. im Kontext der Filteraktivierung, gestartet werden.

5

Rolle	Aktivitäten
TSC Admin, TSC Filtermanager	<ol style="list-style-type: none"> <li>1. tsc_restart.sh starten <ol style="list-style-type: none"> <li>a. Starten des tsc_restart.sh auf allen Produktionssystemen <ol style="list-style-type: none"> <li>i. Ermitteln der PIDs aller Prozesse „TSC Suche“</li> <li>ii. Sukzessive, zeitversetztes Senden des kill Signals an alle Prozesse</li> </ol> </li> </ol> </li> </ol>

6

7 Die „fast-cgi“ Umgebung des Apache Web-Servers sorgt automatisch für das Neu-  
8 starten der vorkonfigurierten Anzahl „TSC Suche“ Prozesse.

9

## 1 TSC Suchanfrage bearbeiten

2 Der Prozess „TSC Suchanfrage bearbeiten“ ist der zentrale Prozess, der durch  
3 das Recherchensystem realisiert wird.

4 Der folgende Ablauf skizziert diesen Prozesses auf hoher Ebene. Detailliert wird  
5 dieser Ablauf später im Dokument.

Rolle	Aktivitäten
Kunde	1. Suche eingeben und abschicken
TSC Server Engine	2. Suchanfrage entgegennehmen 3. Suchstring lesen 4. Attribute der Suchanfrage lesen
BRISBANE-Interpreter	5. Suchstring in elementare Bestandteile zerlegen 6. Stopp-Wort-Filterung durchführen 7. Suchstring für die verschiedenen Kanäle aufbereiten
BRISBANE-Integrator	8. Suchanfrage je „Server“ starten 9. Warten auf alle Ergebnisse der Suchanfrage (ggf. Timeout) 10. Transformation der Ergebnisse (→ XML) 11. Protokollierung der Statistikinformationen
TSC Server Engine	12. Ergebnis zurückliefern

## 8 Auswertung TSC Statistik

9 Die Statistikdatei kann mittels eines Texteditors eingesehen werden oder in  
10 andere Programme importiert werden, da die Datei im CSV-Format vorliegt.

## 1 Auswertung TSC Fehlerprotokoll

2 Mittels eines Texteditors kann das Fehlerprotokoll eingesehen werden.

Rolle	Aktivitäten
BRISBA NE-SC Admin	tbd.

## 6 TSC PlattformAnalysemodell

### 7 Use-Case-Diagramm, Grobanalyse

8 Fig. 4 beschreibt die groben Zusammenhänge der am Recherchensystem (TSC)  
9 beteiligten Akteure und Anwendungsfälle.

### 10 Use-Case: TSC Konfiguration

11 Der „TSC Admin“ erhält die Möglichkeit, die in einer XML-Datei abgelegte  
12 Konfiguration des Suchcenters einzusehen und zu verändern. Dazu bedient er  
13 das Admin-Tool, das auf einem separaten, nicht zur Produktivumgebung  
14 gehörenden Rechner läuft. Damit eine geänderte Konfiguration wirksam wird, ist  
15 ein Neustart der einzelnen Prozesse notwendig.

### 16 Business Process Diagramm

17  
18 Die in Fig. 5 schematisch dargestellte Administration des Recherchensystems be-  
19 inhaltet verschiedene Aktionen. Dazu zählen die Änderungen der Ausschlussfilter  
20 und das Aktivieren der Änderungen. Zur Aktivierung wird ein Script angestoßen,  
21 das die aktuelle Konfiguration auf die Rechner des Produktivsystems kopiert. Im  
22 Diagramm sind die Rechner des Produktivsystems an der unteren rechten Ecke

1 zu sehen. Der Administrator selbst tritt als „worker“ auf und erscheint hier links  
2 oben im Diagramm.

### 3 **Use-Case: TSC Ausschlussfilter pflegen**

4 Der „TSC Filtermanager“ pflegt die Wörter, die von der Suche ausgeschlossen  
5 werden sollen. Diese Wörter können pro Land (Portal) als auch zeitlich  
6 eingeschränkt definiert werden.

7 Nach Änderung der Ausschlussfilter ist jeweils das Recherchensystem (TSC)  
8 neuzustarten, damit die Änderungen aktiv werden können.

### 9 **Use-Case: Konfiguration und Filter aktivieren**

10 Die Konfiguration wird mit Hilfe des Admin-Tools vom Admin-Rechner auf alle am  
11 Suchcenter beteiligten Rechner übertragen und dann ein Neustart des Systems  
12 durchgeführt. Durch einen Neustart werden u. a. neue Prozesse gestartet, die  
13 während ihrer Initialisierung die aktuelle Konfiguration und die Filtertabellen laden.

### 14 **Use-Case: TSC Restart**

15 Das Suchcenter wird mit Hilfe des Admin-Tools neu gestartet. Dabei werden alle  
16 Prozesse auf allen am Suchcenter beteiligten Rechnern geordnet beendet und  
17 neugestartet. Ein Neustart ist notwendig, damit Änderungen der Ausschlussfilter  
18 oder der Konfiguration wirksam werden.

### 19 **Use-Case: Suchanfrage bearbeiten**

20 Der „Kunde“ gibt in der Suchmaske des Ein-/Ausgabesystems seinen  
21 Suchausdruck ein. Dieser kann aus mehreren mit UND-, ODER- und NICHT-  
22 Operatoren verknüpften Wörtern bestehen. Dieser Suchausdruck wird nach  
23 Absenden an die Server-Engine von dieser interpretiert, an die Anforderungen der  
24 verschiedenen (z. T. externen) Suchmaschinen angepasst und an diese selbst  
25 geschickt. Unter Berücksichtigung von konfigurierbaren Timeouts werden die  
26 verschiedenen Suchergebnisse dann zu einer Gesamtansicht integriert und zum  
27 Browser des Kunden zurückgeliefert.

### 1 **Use-Case: Auswertung Statistik**

2 Der „TSC Filtermanager“ kann über einen Texteditor die Suchbegriffe und  
3 Verarbeitungszeiten einsehen. Zusätzlich sollen dem Auswerter Möglichkeiten  
4 bereitgestellt werden, die eine Ad-hoc-Standardauswertung nach noch zu  
5 definierenden Kriterien ermöglicht.

### 6 **Use-Case: Auswertung Fehlerprotokoll**

7 Der „TSC Admin“ kann über einen Texteditor die in einer Textdatei abgelegten  
8 Fehlermeldungen einsehen.

9

### 10 **Use-Case-Diagramme, Feinanalyse**

11 In Fig. 6 ist der Ablauf einer Anfrage an die Suchmaschinen schematisch  
12 dargestellt. „Interpretierte Anfrage senden“ bearbeitet den eintreffenden  
13 Suchauftrag, passt diesen an die Syntax der beteiligten externen Suchmaschinen  
14 an und sendet ihn an diese weiter.

15 „Ergebnisse Integrieren“ empfängt die Suchergebnisse der Suchmaschinen und  
16 präsentiert sie dem Nutzer im gewünschten Format.

17

18 Fig. 7 zeigt schematisch den Ablauf einer verteilten Abfrage mit anschließendem  
19 Zusammenführen der Ergebnisse. „Ergebnisse sammeln“ akkumuliert die von den  
20 Suchmaschinen eintreffenden Suchergebnisse. Ein Suchergebnis muss innerhalb  
21 eines konfigurierbaren Zeitlimits (Timeout) eintreffen. Ist dem nicht so, wird dies  
22 als Fehler interpretiert.

23 „Errorhandling“ übernimmt das Protokollieren von Fehlerzuständen in eine  
24 Fehlerdatei.

25 „Priorisieren“ wichtet die von den verschiedenen Suchmaschinen gelieferten  
26 Ergebnisse nach bestimmten Geschäftsregeln.

27 Fehlerbehandelte und priorisierte Suchergebnisse werden „in Ergebnisseite  
28 integrieren“ in das gewünschte Ausgabeformat überführt.



1  
2 Der Nutzer/Kunde hat, wie in Fig. 8 dargestellt, zwei Möglichkeiten der Suche.  
3 Bei „einfach Suchen“ kann der Nutzer einen oder mehrere Suchbegriffe und even-  
4 tuell Suchoperatoren eingeben, die der FAST-CGI Syntax entsprechen müssen.

5  
6 Bei „erweiterter Suche“ kann der Nutzer zuerst den Typ des gesuchten  
7 Webinhalts auswählen ( Webseite, Bild, Musik etc.). Je nach Inhaltstyp stehen  
8 dann angepasste Suchoptionen zur Verfügung.

9  
10 Der in Fig. 9 schematisch dargestellte Filter oder Interpreter bekommt zusätzlich  
11 zu den Konfigurationsdaten den aktuellen „query string“. Diese Zeichenkette  
12 enthält Parameter und die Suchanfrage des Nutzers. Die Suchanfrage wird  
13 entsprechend der Filterregeln des jeweiligen Systems (Suchmaschine) modifiziert.  
14 Ein weiterer Schritt baut aus den jeweiligen „query strings“ und einigen  
15 Parametern eine komplette URL für die Suchmaschine. Diese URL ist im o. g.  
16 Diagramm als „Ziel\_URL“ identifizierbar.

## 17 **Modulüberblick**

18 Das Gesamtsystem unterteilt sich, wie in Fig. 10 gezeigt, im wesentlichen in drei  
19 Module. Diese werden im Folgenden näher beschrieben. In Interpreter und Inte-  
20 grator existieren Adapterklassen, die jeweils auf genau ein Zielsystem zugeschnit-  
21 ten sind. Alle Adapterklassen besitzen einen statischen und einen dynamischen  
22 Teil. Der statische Teil wird durch Konfigurationsdaten repräsentiert. Der  
23 dynamische Teil hängt von der jeweiligen Suchanfrage ab.

## 25 **Modul Interpreter**

26 Das Modul BRISBANE-Interpreter enthält, wie in Fig. 11 gezeigt, Klassen, die der  
27 Filterung und sonstiger Aufbereitung der Suchanfragen dienen. Allgemein kann  
28 man sagen, dass der Interpreter die Basisdaten für den Integrator bereitstellt. Die  
29 Funktionalität der Filterung ist stark vom Zielsystem (Suchmaschine) abhängig  
30 und verbirgt sich in den entsprechenden Adapterklassen.

## Modul Integrator

Die in Fig. 12 gezeigten Klassen im Modul Integrator sind verantwortlich für Weiterleitung der modifizierten Suchanfragen an die Zielsysteme. Weitere Aufgaben des Integrators sind das Zusammenstellen der Suchergebnisse und deren Konvertierung von XML in HTML, das Darstellen der Ergebnisseite, sowie die Bereitstellung der Statistikdaten.

Für jede Anfrage an ein Zielsystem wird eine Instanz der Klasse Request erzeugt. Diese Objekte stellen die Kommunikationsschnittstelle zwischen Suchcenter und Zielsystemen dar.

Die Kommunikation mit den Zielsystemen wird durch den RequestManager verwaltet. Er ist jedoch nicht für die Darstellung der Ergebnisse verantwortlich.

## Modul Server-Engine

Dieses in Fig. 13 anhand seiner Klassen gezeigte Modul repräsentiert den Hauptprozess des Suchcenters. Er besitzt eine Instanz der Klasse „Configuration“, die den Zugriff auf alle Konfigurationsdaten ermöglicht. Innerhalb des Prozesses werden Konfigurationsdaten in unterschiedlichen Strukturen wie z. B. AdapterMap abgelegt.

Die Server-Engine ist auch für das Instanzieren von Interpreter und Integrator sowie das geordnete Beenden des Prozesses ( bei Eintreffen eines Signals) verantwortlich.

Die Klasse Logging ist für die Protokollierung von Fehlern und Statistikdaten verantwortlich.

## Dynamisches Modell

In Fig. 14 ist das dynamische Verhalten der einzelnen Komponenten und der Informationsfluss zwischen ihnen dargestellt.

Nach dem Start der Suchcenter-Software wird die Konfiguration ausgelesen. Je nach Konfiguration werden dann entsprechende Ausschlussfilter und eine

1 länderspezifische XSL-Datei geladen. Der Prozess wartet nun auf eine  
2 Suchanfrage. Wenn diese eintrifft, wird der Suchausdruck interpretiert. In  
3 Ausschlussfiltern vorkommende Wörter werden aus der Suchanfrage entfernt.  
4 Zusätzlich wird die Anfrage an die Syntax der beteiligten Suchmaschinen  
5 angepasst. Die modifizierten Suchausdrücke werden nun an die Suchmaschinen  
6 gesendet. Abhängig vom Format der zurückgelieferten Ergebnisse ist eventuell  
7 eine Transformation in das HTML- Format notwendig (wenn beispielsweise die  
8 Ergebnisse im XML- Format vorliegen). Die Ergebnisse werden an den Browser  
9 des Kunden gesendet.

10 Der Prozess geht wieder in Warteposition, bis der nächste Request eintrifft.

11 Die einzelnen Suchanfragen werden mit Suchausdruck, Zeitstempel und Dauer  
12 der Verarbeitung in der Statistikdatei „tsc\_statistics“ abgelegt. Zudem werden  
13 Fehlerzustände während des gesamten Verlaufs in der Datei „tsc\_error\_log“  
14 eingetragen.

## 16 **Modul Server Engine**

### 17 **Initialisierungsphase**

18  
19 Während der Initialisierung werden, wie in Fig. 15 gezeigt, abhängig von aktiven  
20 Kanälen, Konfigurationsobjekte und Instanzen aller Adapter erzeugt. Die Kanäle  
21 entsprechen den vorher angesprochenen Channels und repräsentieren ein ganz  
22 bestimmtes Zielsystem.

### 23 **Verarbeitung einer Suchanfrage**

24 Bei Eintreffen eines Requests werden, wie in Fig. 16 dargestellt, Instanzen von  
25 Interpreter und Integrator angelegt. Alle notwendigen Informationen für die  
26 Instanzen werden von der Server-Engine bereitgestellt. Der Interpreter bereitet nun  
27 die Modifikationen der Suchanfrage vor, der Integrator sendet diese an die  
28 Zielsysteme und liefert die formatierten Suchergebnisse zurück. Die Instanzen von  
29 Interpreter und Integrator werden anschließend zerstört.

## 1 **Modul Interpreter**

2  
3 Der Interpreter veranlasst, wie in Fig. 17 gezeigt, alle für den Request benötigten  
4 Adapter, jeweils ein AdapterResult zu erzeugen. Jeder einzelne Adapter  
5 modifiziert die originale Suchanfrage unter Zuhilfenahme der entsprechenden  
6 Filterregeln. Diese Suchanfrage und weitere Parameter sind für das Erzeugen  
7 eines AdapterResult-Objektes notwendig. Das erzeugte AdapterResult-Objekt  
8 wird dem Interpreter zurückgegeben, der es einer Liste hinzufügt.

## 9 **Modul Integrator**

10 Nach Erzeugung der Integrator-Instanz befindet diese sich in einem initialen  
11 Zustand. Die Verarbeitung der Requests erfolgt durch den Aufruf der Methode  
12 „doRequests()“.

## 14 **Main des fast- Programms**

15  
16 Das Fast-CGI Programm enthält eine Funktion „main()“. Dieses wird vom Webser-  
17 ver aufgerufen. Voraussetzung für das ordnungsgemäße Arbeiten des Programms  
18 ist die korrekte Initialisierung und das Eintreten in die Warteschleife. Die Konfigu-  
19 ration wird in einer Datei gehalten. Die Angabe des kompletten Pfades dieser Da-  
20 tei erfolgt über eine Umgebungsvariable.

## 21 **Konfiguration des Recherchensystem**

22 Das Suchcenter wird über eine einzige Datei konfiguriert (tsc.cfg). Diese Datei be-  
23 findet sich auf einem separaten Rechner, der nicht als Produktivrechner verwen-  
24 det wird. Die Konfigurationsdatei dient als Vorlage für alle Konfigurationsdateien  
25 pro Rechner in der Produktionsumgebung. Bei Änderungen der Konfiguration  
26 (z. B.: Modifikation des Verzeichnisses für die Statistikdateien) wird die Konfigura-  
27 tionsdatei auf dem außerhalb des Produktivsystems befindlichen Rechner modifi-  
28 ziert und per Script auf alle Rechner im Produktivsystem kopiert.

## 1 **Format der Filtertabelle**

2 Auf Anforderung im Admintool wird aus der Datenbank (mysql) eine aktuelle  
3 Filterdatei erstellt, die über die Verteilmechanismen an die Suchcenter-Rechner  
4 übertragen wird.

5 Diese Filterdatei hat zeilenweise Einträge, innerhalb einer Zeile sind die Werte  
6 durch Semikolon getrennt.

7 Eine Zeile ist folgendermaßen sematisch aufgebaut (Werte in [] sind optional.):

8 Ländercode; zu filterndes Wort oder Phrase; [Datum Beginn]; Uhrzeit  
9 Beginn; [Datum Ende]; Uhrzeit Ende; Kanal 1; Kanal 2; ...;Kanal n

10 z. B.

11 DE;Neu;2000-10-10;12:00:00;2001-12-31;23:35:00;WWW;;;;;

12 DE;verboten;2001-10-20;01:00:00;2001-12-

13 31;20:00:00;WWW;P4P;BRISBANE;PHP;SP;AD

14 UK;Bier;;00:00:00;;00:00:00;WWW;;BRISBANE;;SP;

15 CH;Alkohol;2001-09-30;15:00:00;;00:00:00;WWW;;BRISBANE;PHP;;

## 17 **Aktivierung von Konfiguration und Filtertabelle**

18 Die Aktivierung von Konfiguration und Filtertabelle erfolgt durch das Starten von  
19 Scripts.

## 21 **Format der Protokolldatei**

22  
23 Der Name der Statistikdatei wird in der Konfigurationsdatei (tsc.cfg) angegeben,  
24 z. B. tsc\_statistics\_<PID>. <PID> wird dann im Betrieb durch die Prozess-ID  
25 ersetzt. Da gleichzeitig mehrere Prozesse laufen, gibt es auch dementsprechend  
26 viele Protokolldateien.

27  
28 Alle Statistikdateien liegen in einem zentralen Verzeichnis.

29 Diese Statistik-Dateien werden im CSV-Format gespeichert. Zur Auswertung  
30 existiert ein separater Prozess, der aufgrund des großen Datenvolumens auf

1 einem eigens dafür eingerichteten Server laufen sollte. Die Such-Prozesse  
2 schreiben entweder ihre Daten direkt auf den Auswertungsserver oder der  
3 Auswertungsprozess sammelt in regelmäßigen Abständen die Statistik-Daten aller  
4 Such-Prozesse ein.

### 5 6 **Aufbau der Datei**

7  
8 Die Statistik-Datei besteht aus 4 Grundspalten und pro realem Kanal kommen  
9 weitere 4 Spalten hinzu.

#### 10 11 **Grundspalten:**

- 12 1. Eingegebene Suchbegriffe
- 13 2. Uhrzeit der Suche
- 14 3. Suchmodus
- 15 4. Gesamtzeit in ms oder Timeout

#### 16 17 **Spalten pro Kanal:**

- 18 1. Name
- 19 2. verwendete Suchbegriffe
- 20 3. Zeitdauer in ms oder Timeout
- 21 4. Anzahl der Ergebnisse

### 22 23 **Mengengerüst**

24  
25 Grundspalten ca. 50 Byte

26 Kanalspalten je ca. 50 Byte

27  
28 Maximallast 500 Anfragen pro Sekunden, ca. 250 Byte pro Anfrage:

29 → 125 KByte pro Sekunde → 450 MByte pro Stunde → 10,8 GByte pro Tag

## 1 **Format der Fehlerprotokollierung**

2 Der Name der Fehlerprotokolldatei wird in der Konfigurationsdatei (tsc.cfg)  
3 angegeben, z. B. tsc\_err\_<PID>. <PID> wird dann im Betrieb durch die Prozess-  
4 ID ersetzt. Da gleichzeitig mehrere Prozesse laufen, gibt es auch  
5 dementsprechend viele Fehlerprotokolldateien.

6 Einträge in das Fehlerprotokoll geschehen zeilenweise. Innerhalb eines Eintrages  
7 sind die Werte kommasepariert und hat folgende sematische Bedeutung  
8 Uhrzeit, Original-URL, angepasste URL, Zeit für den reinen Request, Anzahl  
9 Treffer, Beschreibung des aufgetretenen Fehlers

10 z. B.:

11 12.10.2001 09:41:35,http://192.168.70.42:80/fast-  
12 cgi/ubrpsc?lang=any&q=test&start=0&num=10&context=www&type=all&Port  
13 alLanguage=de&context=www,http://66.77.74.21:8075/max\_xml\_b00?hits=  
14 10&query=test&offset=0,HTTP-Call:145 ms,Total Hits:15737604,ERROR in  
15 BrisbaneIntegrator::getQueryResults() 'ghttp\_process () failed,Error:  
16 getQueryResults

17  
18 Wenn ein allgemeiner Fehler auftritt, wird dieser in der Form, Uhrzeit,  
19 Beschreibung des aufgetretenen Fehlers protokolliert, z. B.

20 12.10.2001 09:41:35,ERROR in BrisbaneIntegrator::initXSLT() 'mpXMLStyleptr:  
21 xmlParseMemory error',Error: can't initialize xslt-process. Process Exit

## 22 **Beschreibung der Stylesheets**

23 Die XSL-Stylesheets dienen innerhalb des Suchcenters dazu, die gesammelten  
24 Antworten in XML-Form zu einer HTML-Seite entsprechend der aktuellen  
25 Layoutvorgaben umzusetzen. Die Daten für eine Webseite werden dem  
26 Stylesheet im XML, bzw. durch Stylesheetparameter übergeben. Pro Sprache  
27 existiert im Suchcenter ein Stylesheet. Innerhalb des Stylesheets werden die  
28 Suchergebnisse in HTML aufbereitet sowie die Links für das Vor- und  
29 Zurückblättern erzeugt. Entsprechend den Voreinstellungen wird dann noch die

- 1 entsprechende Reiterstruktur über den Suchergebnissen und die Werbebanner
- 2 bzw. die Formular für die neue Suche erstellt.

### 3 **Administration der Filtertabelle**

#### 4 **Kontext**

- 5 **Der Kontext der Admin Anwendung ist in Fig. 18 dargestellt.**



## Bezugszeichenliste

	1	Recherchensystem
5	2	Suchmaschinen
	4	zentrale Auswerteeinheit
	6	Integrationsmodul
	8	Speichermodul
	10	Distributionsmodul
10	12	Pufferspeicher
	14	Speichermodul
	16	Filtermodul
	18	Ein-/Ausgabemodul
	20	Werbemitteldatenbank
15	E	Ergebnisdatensätze
	P	Priorisierungsparameter
20	1	

- 32 -

## Ansprüche

- 5 1. Recherchensystem (1) zur Ermittlung von Informationen aus einer Datenbank unter Nutzung einer Mehrzahl von Suchmaschinen (2) mit einem Integrationsmodul (6), das die von den Suchmaschinen (2) gelieferten, jeweils eine Mehrzahl von Ergebnisdatensätzen (E) umfassenden Ergebnisdaten in ein gemeinsames Datenformat konvertiert und anschließend unter Rückgriff auf in einem Speichermodul (8) hinterlegte Priorisierungsparameter (P) jedem  
10 Ergebnisdatensatz (E) einen Relevanzkennwert zuordnet.
2. Recherchensystem (1) nach Anspruch 1, bei dem die Priorisierungsparameter (P) abhängig von der jeweiligen Suchmaschine (2) gewählt sind.
- 15 3. Recherchensystem (1) nach Anspruch 1 oder 2, dessen Integrationsmodul (6) den Relevanzkennwert für den jeweiligen Ergebnisdatensatz (E) unter Berücksichtigung eines von der jeweiligen Suchmaschine (2) bereitgestellten Ausgangsrelevanzkennwerts ermittelt.
- 20 4. Recherchensystem (1) nach einem der Ansprüche 1 bis 3, dessen Integrationsmodul (6) über ein gemeinsames Pufferspeichermodul (12) mit einem Distributionsmodul (10) kommuniziert, das einen im gemeinsamen Datenformat eingehenden Suchauftrag für jede ausgewählte Suchmaschine (2) in deren spezifisches Datenformat umwandelt und anschließend an diese ausgibt.  
25
5. Recherchensystem (1) nach Anspruch 4, dessen Distributionsmodul (10) mit einem zweiten Speichermodul (14) verbunden ist, in dem Syntaxdaten für die Suchmaschinen (2) hinterlegt sind.
- 30 6. Recherchensystem (1) nach Anspruch 4 oder 5, in dessen Pufferspeichermodul (12) für einzelne Suchaufträge charakteristische Identifikationsdaten hinterlegbar sind.

7. Recherchensystem (1) nach einem der Ansprüche 4 bis 6, dessen Distributionsmodul (10) eingangsseitig über ein Filtermodul (16) mit einem Ein-/Ausgabemodul (18) verbunden ist.
- 5 8. Recherchensystem (1) nach einem der Ansprüche 4 bis 7, dessen Distributionsmodul (10) mit dem Integrationsmodul (6) in eine gemeinsame Einheit integriert ist.
9. Recherchensystem (1) nach einem der Ansprüche 4 bis 8, bei dem das  
10 Distributions- und das Integrationsmodul (6) datenseitig mit einer Werbemitteldatenbank (20) verbunden sind, die als Antwortdatensatz auf eine eingehende Suchanfrage einen hinsichtlich eines Kennungssignals an die Suchanfrage angepassten Datensatz liefert.
- 15 10. Verfahren zur Ermittlung von Informationen aus einer Datenbank, insbesondere aus dem World-Wide-Web, unter Nutzung eines Recherchensystems (1) nach einem der Ansprüche 1 bis 9, bei dem die von den Suchmaschinen (2) gelieferten, jeweils eine Mehrzahl von Ergebnisdatensätzen (E) umfassenden Ergebnisdaten in ein gemeinsames Datenformat konvertiert werden, wobei anschließend unter  
20 Rückgriff auf in einem Speichermodul (8) hinterlegte Priorisierungsparameter (P) jedem Ergebnisdatensatz (E) einen Relevanzkennwert zugeordnet wird.
11. Verfahren nach Anspruch 10, bei dem die Priorisierungsparameter (P) abhängig von der jeweiligen Suchmaschine (2) gewählt werden.
- 25 12. Verfahren nach Anspruch 10 oder 11, bei dem der Relevanzkennwert für den jeweiligen Ergebnisdatensatz (E) unter Berücksichtigung eines von der jeweiligen Suchmaschine (2) bereitgestellten Ausgangsrelevanzkennwerts ermittelt wird.
- 30 13. Verfahren nach einem der Ansprüche 10 bis 12, bei dem ein im gemeinsamen Datenformat eingehender Suchauftrag für jede ausgewählte Suchmaschine (2) in deren spezifisches Datenformat umgewandelt und anschließend an diese ausgegeben wird.

- 34 -

14. Verfahren nach einem der Ansprüche 10 bis 13, bei dem ein über ein Ein-/Ausgabemodul (18) vorgegebener Suchauftrag vor seiner weiteren Bearbeitung einer Filterung unterzogen wird.

- 5 15. Verfahren nach einem der Ansprüche 10 bis 14, bei dem ein Suchauftrag zusätzlich an eine Werbemitteldatenbank (20) ausgegeben wird, wobei ein von dieser eingehender Antwortdatensatz mittels eines Kennungssignals der jeweiligen Suchanfrage zugeordnet wird.

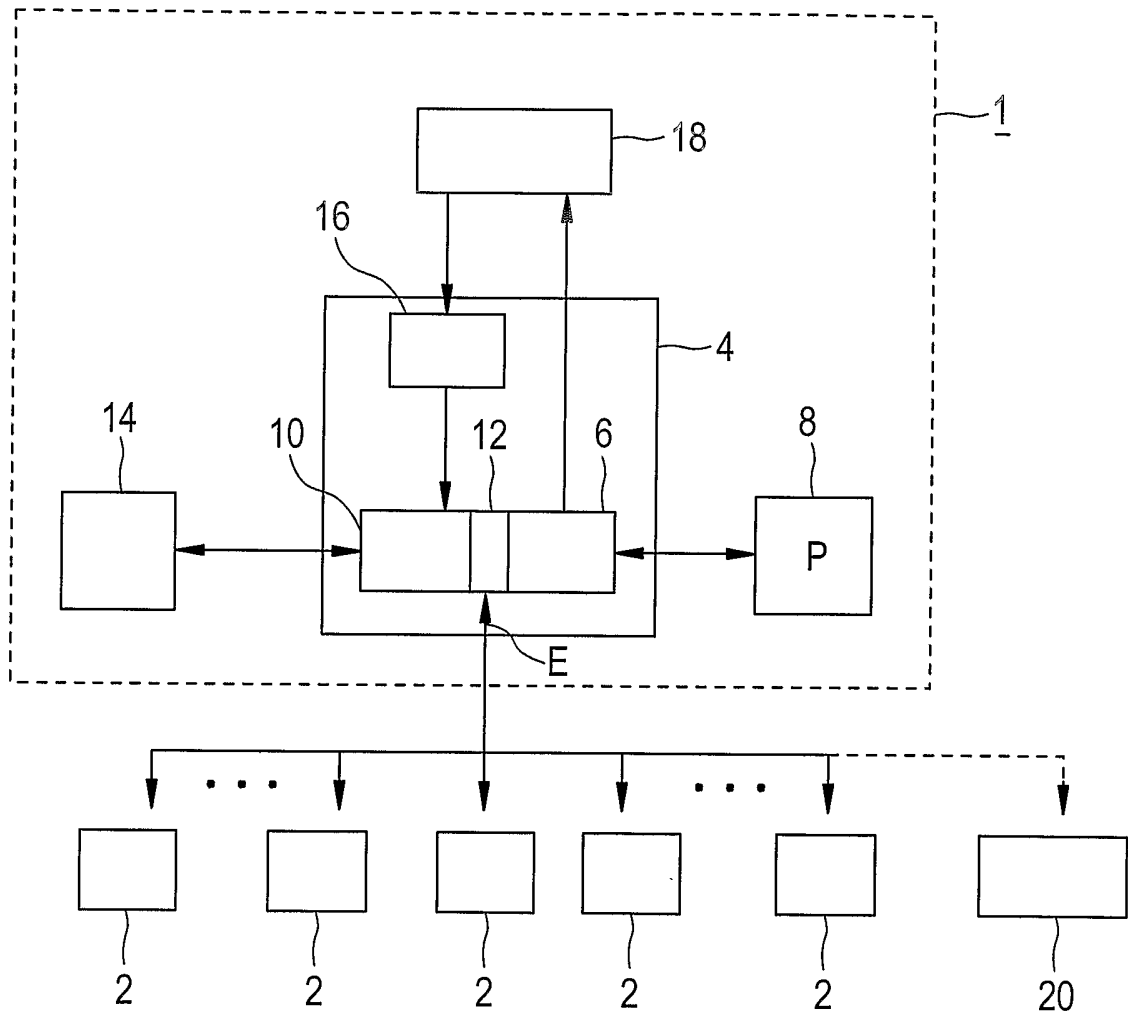


FIG. 1

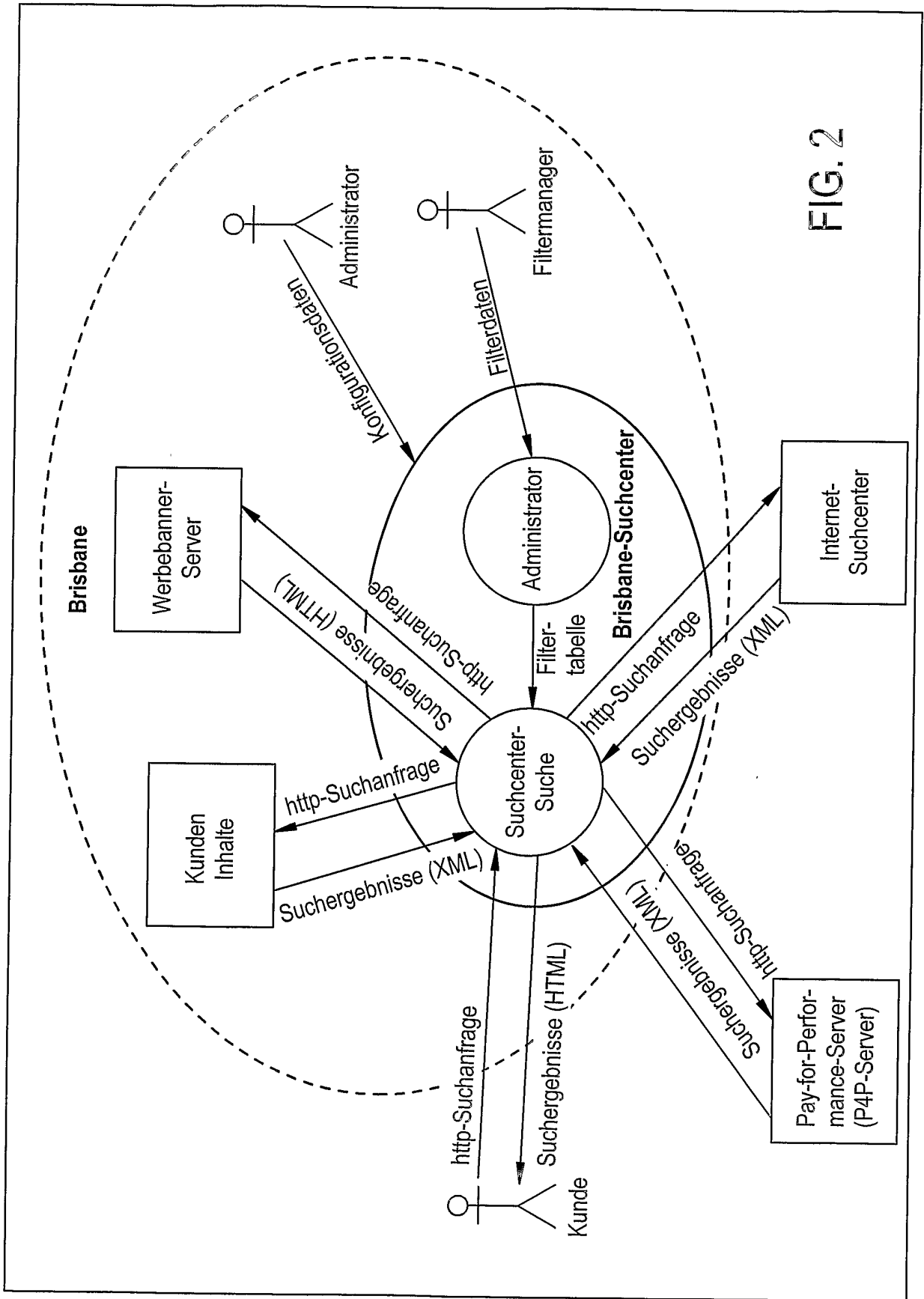


FIG. 2

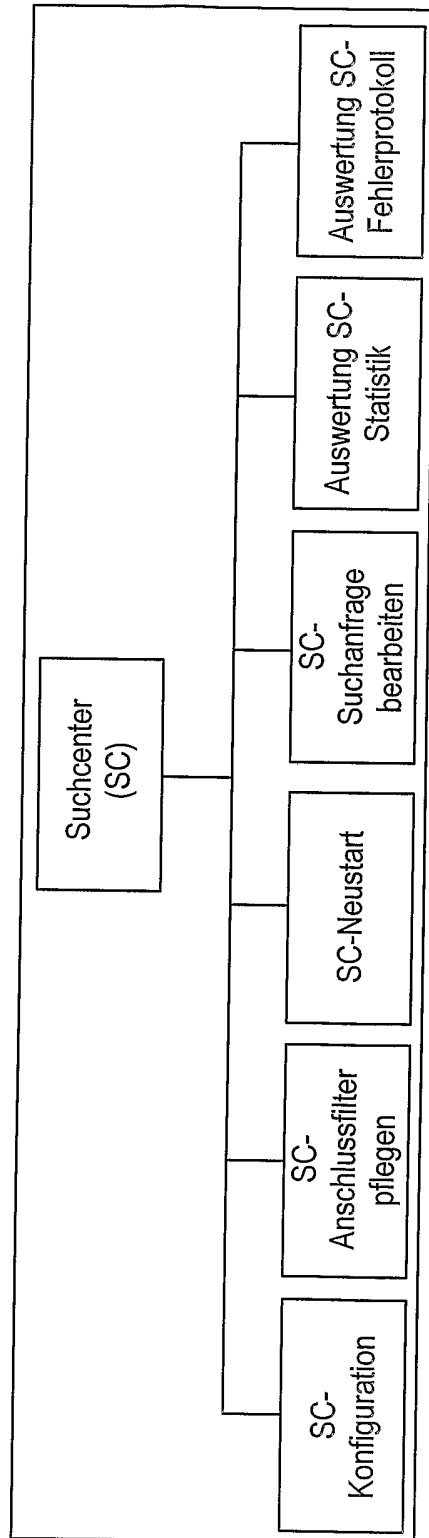
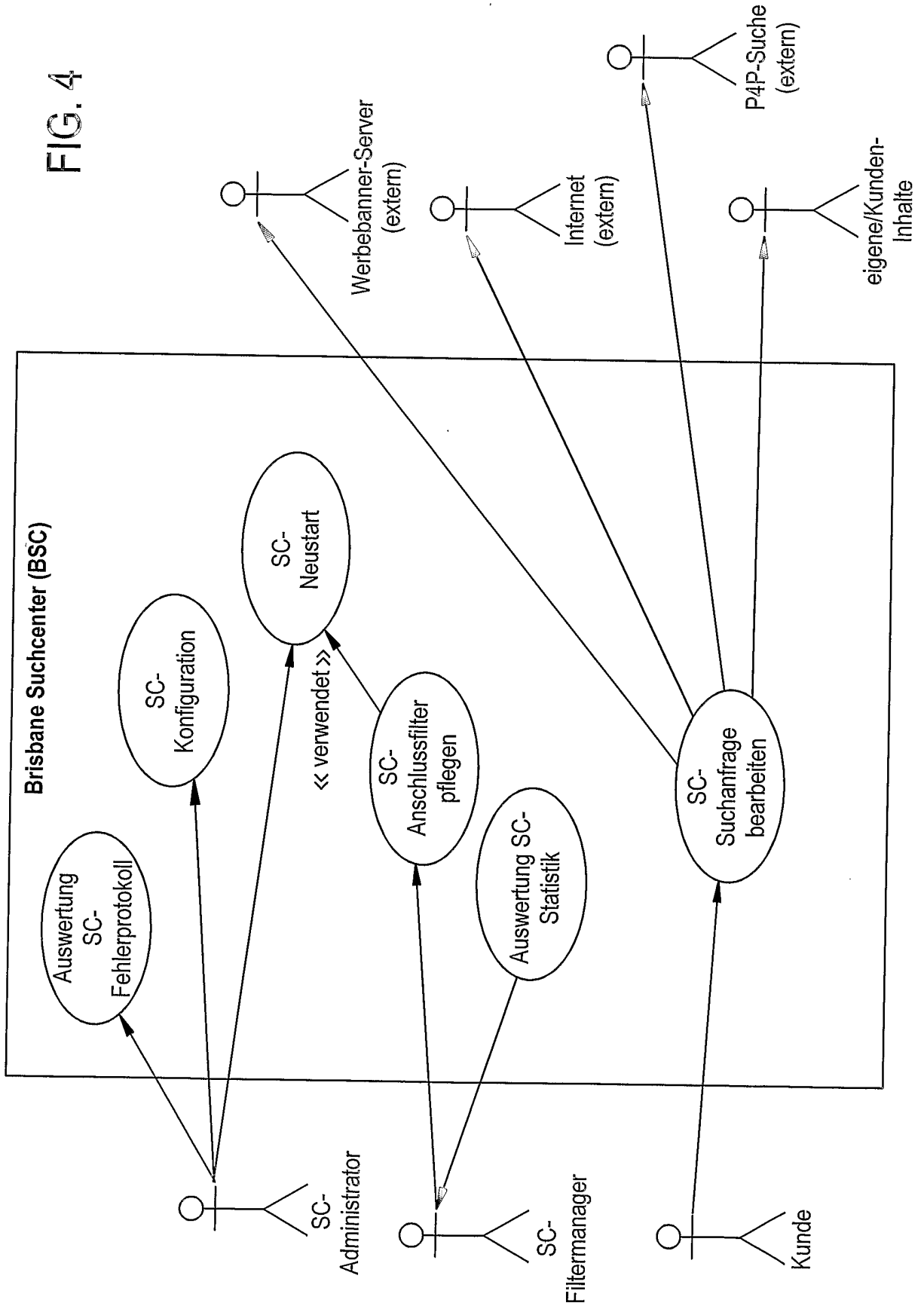


FIG. 3

FIG. 4





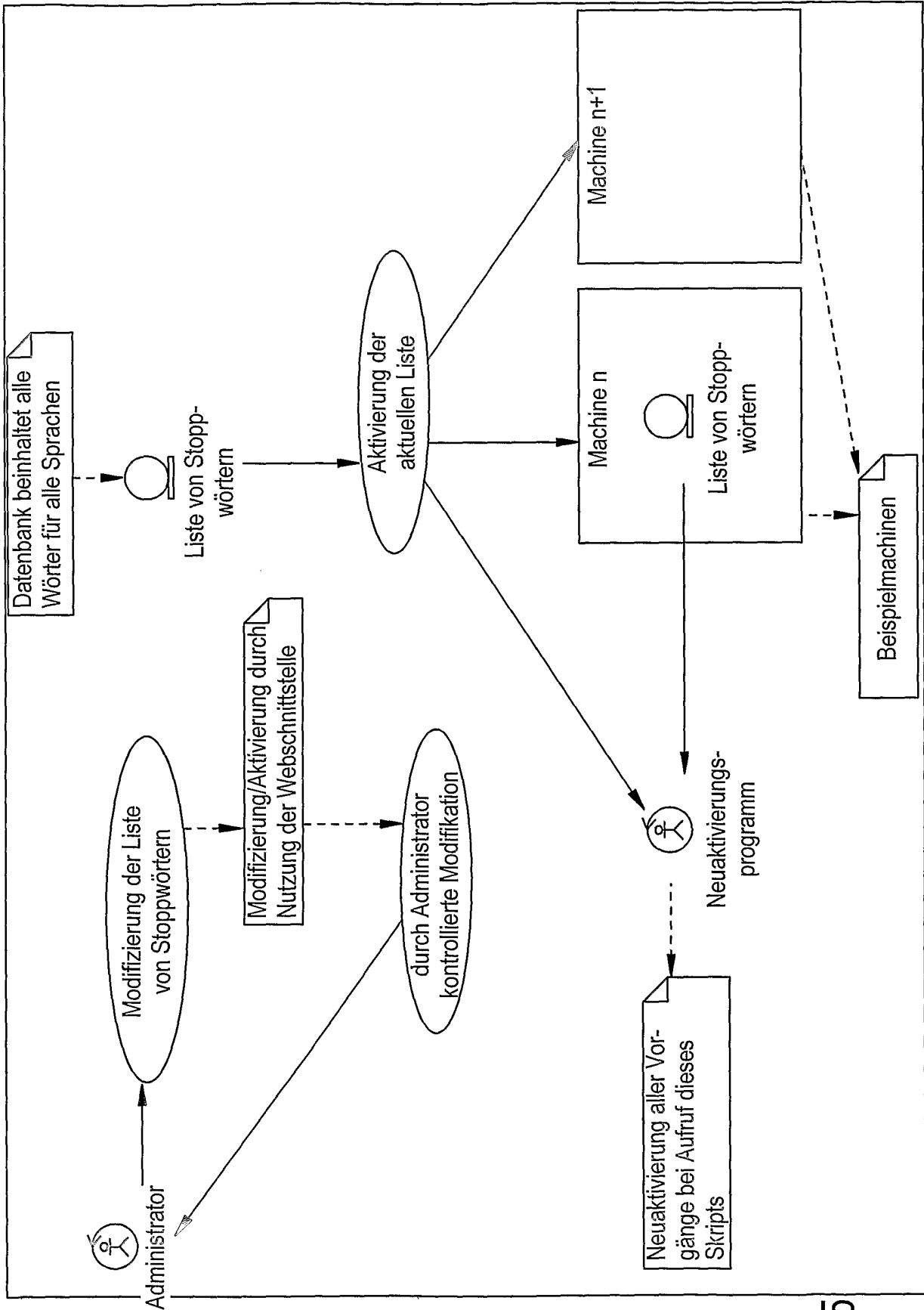


FIG. 5

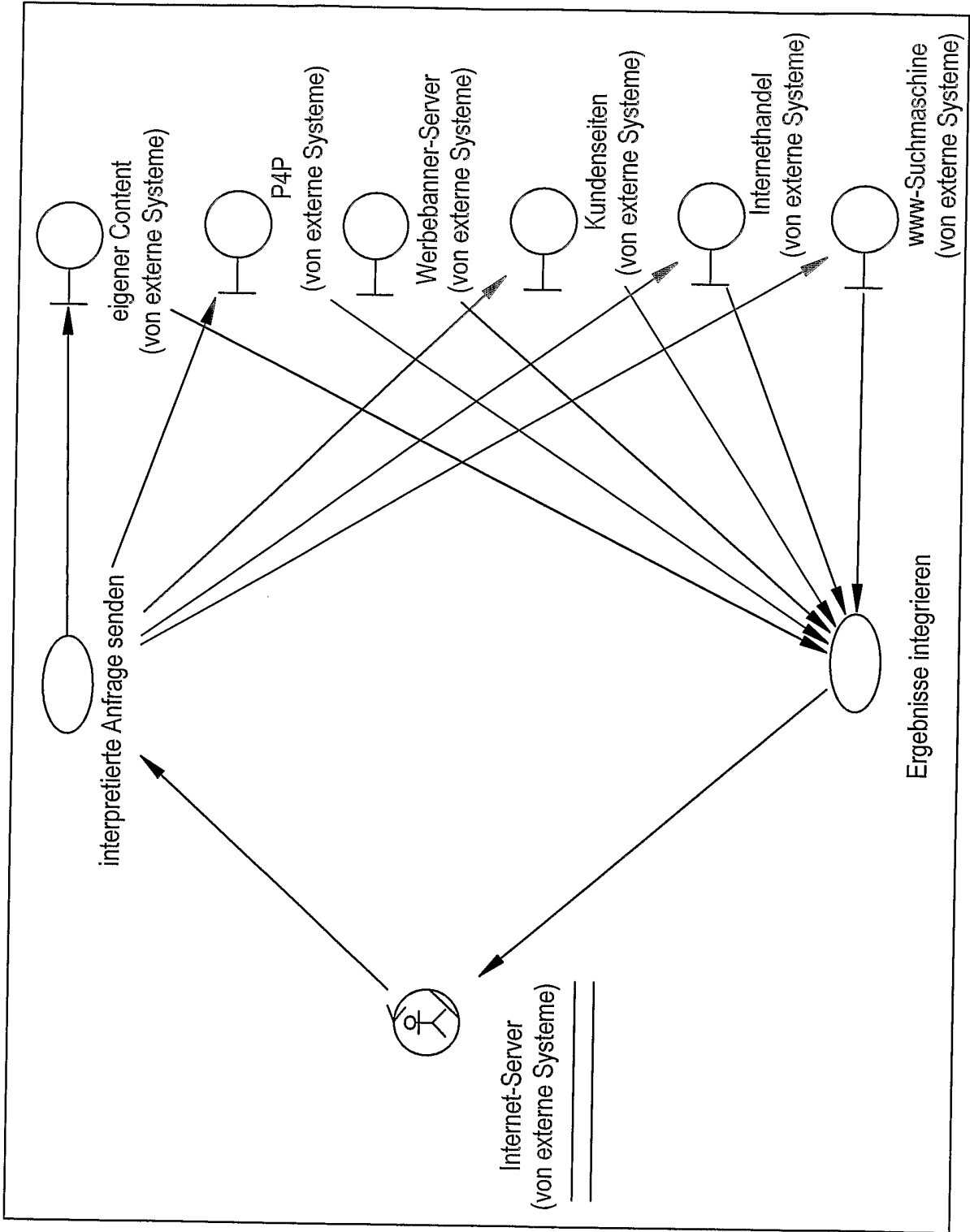


FIG. 6

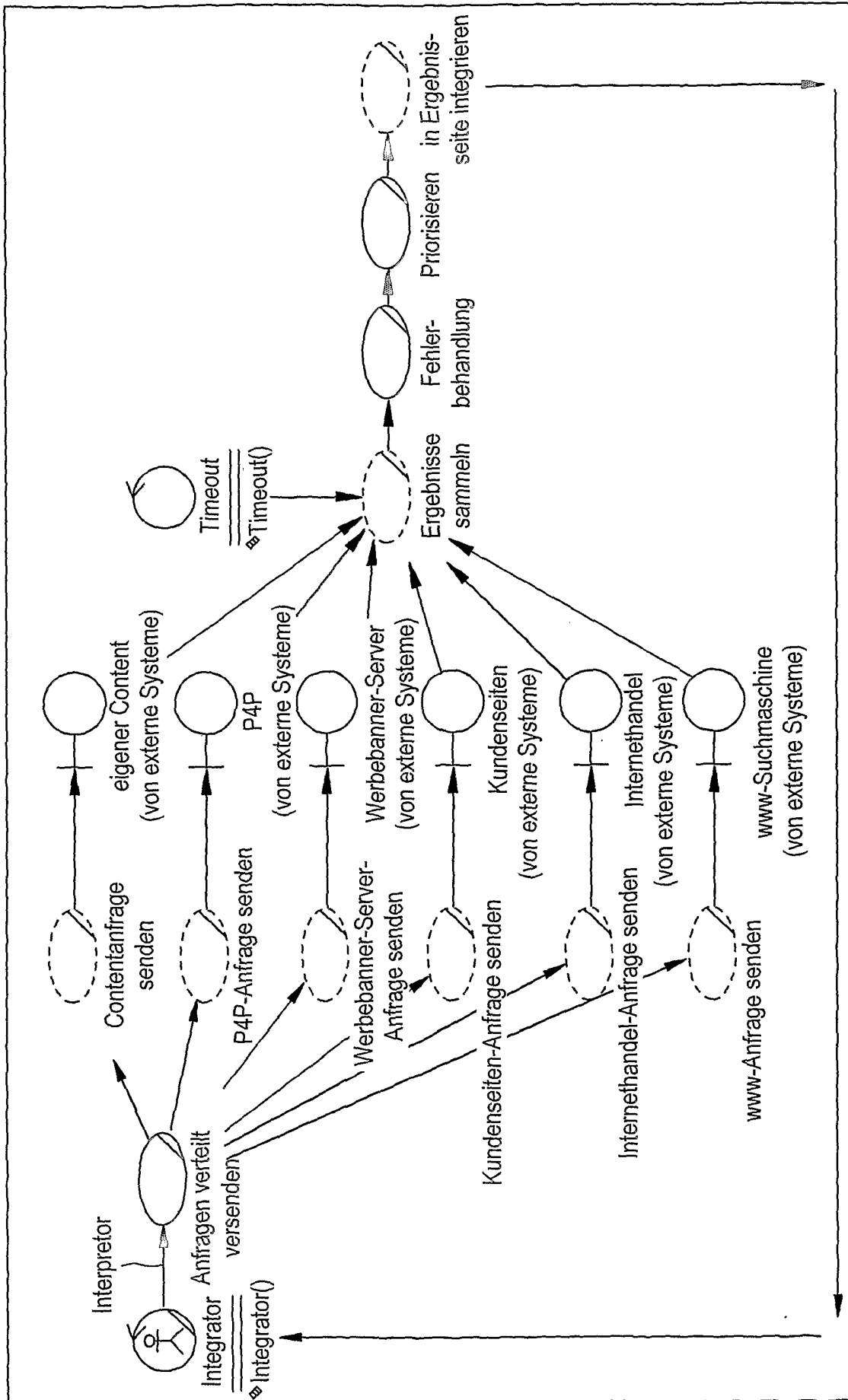


FIG. 7

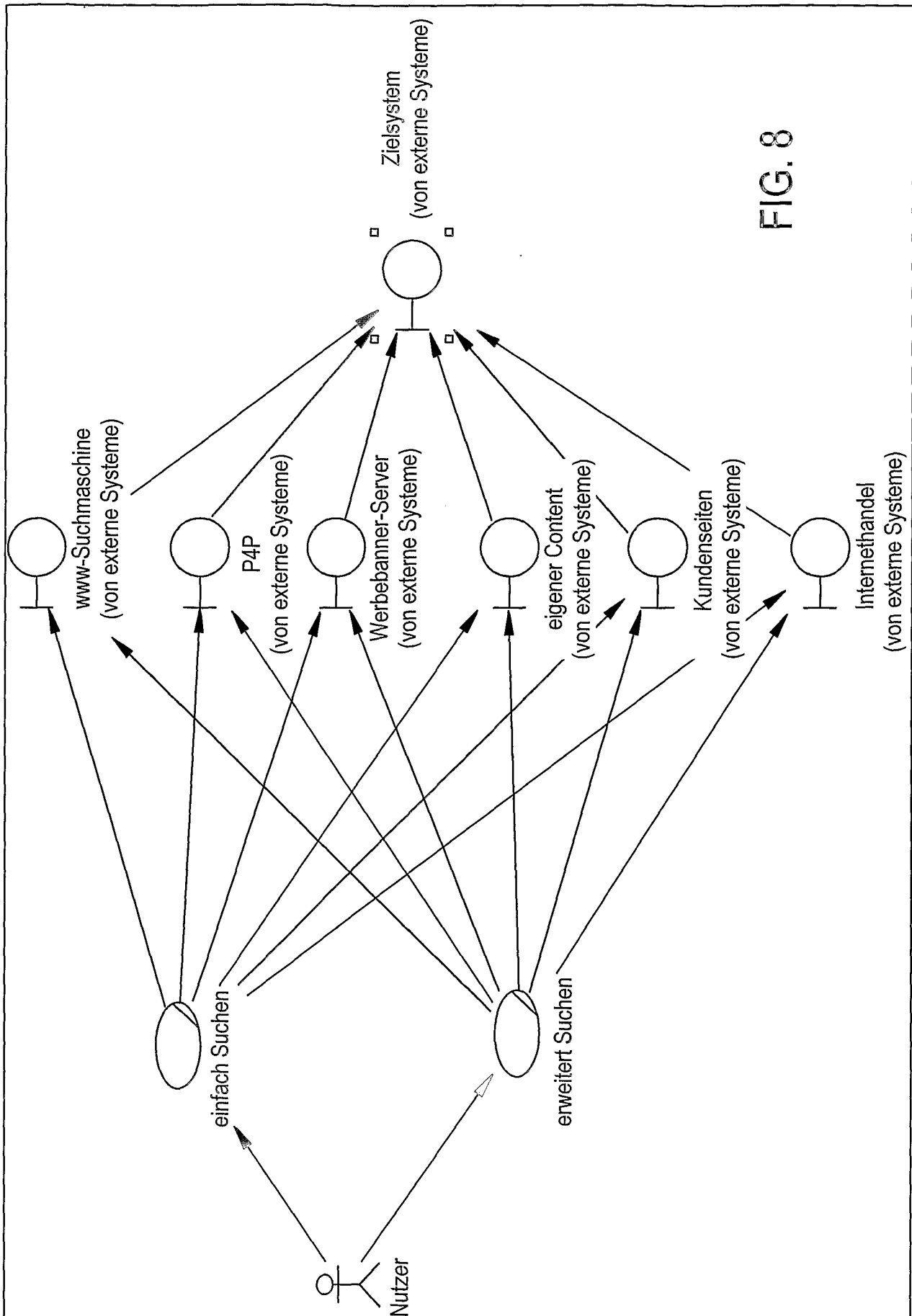


FIG. 8

9/17

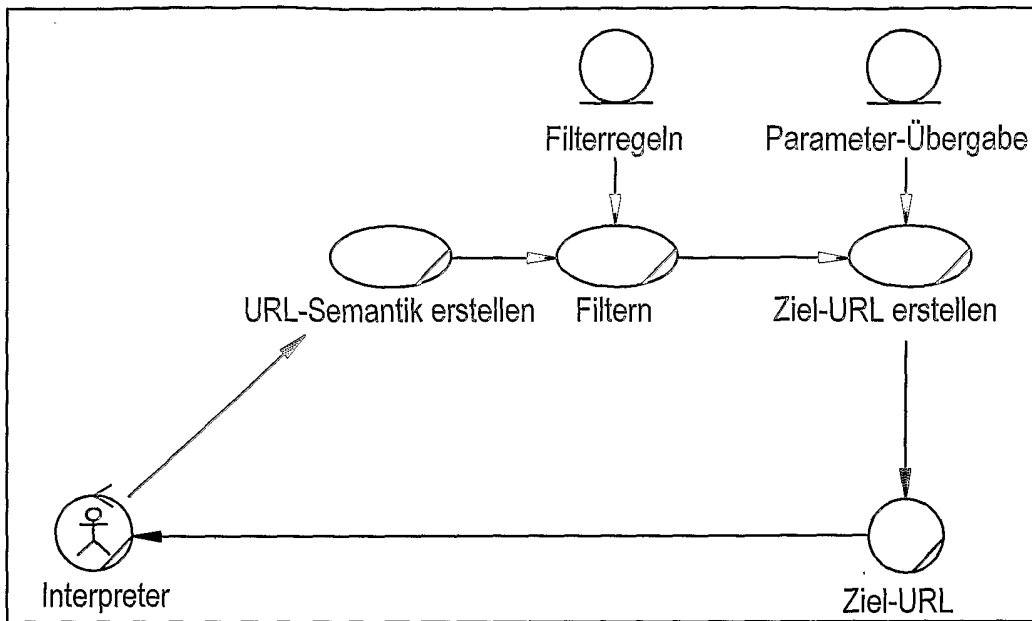
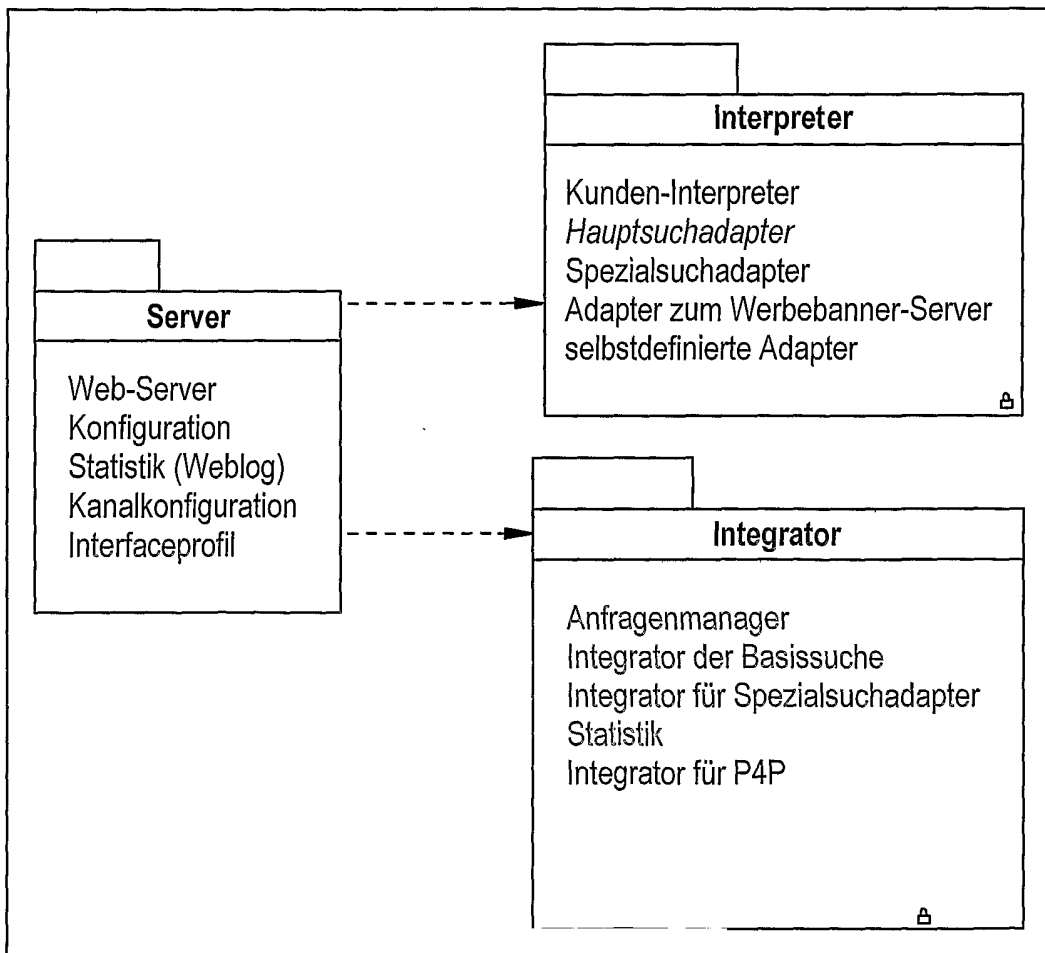


FIG. 9

FIG. 10



10/17

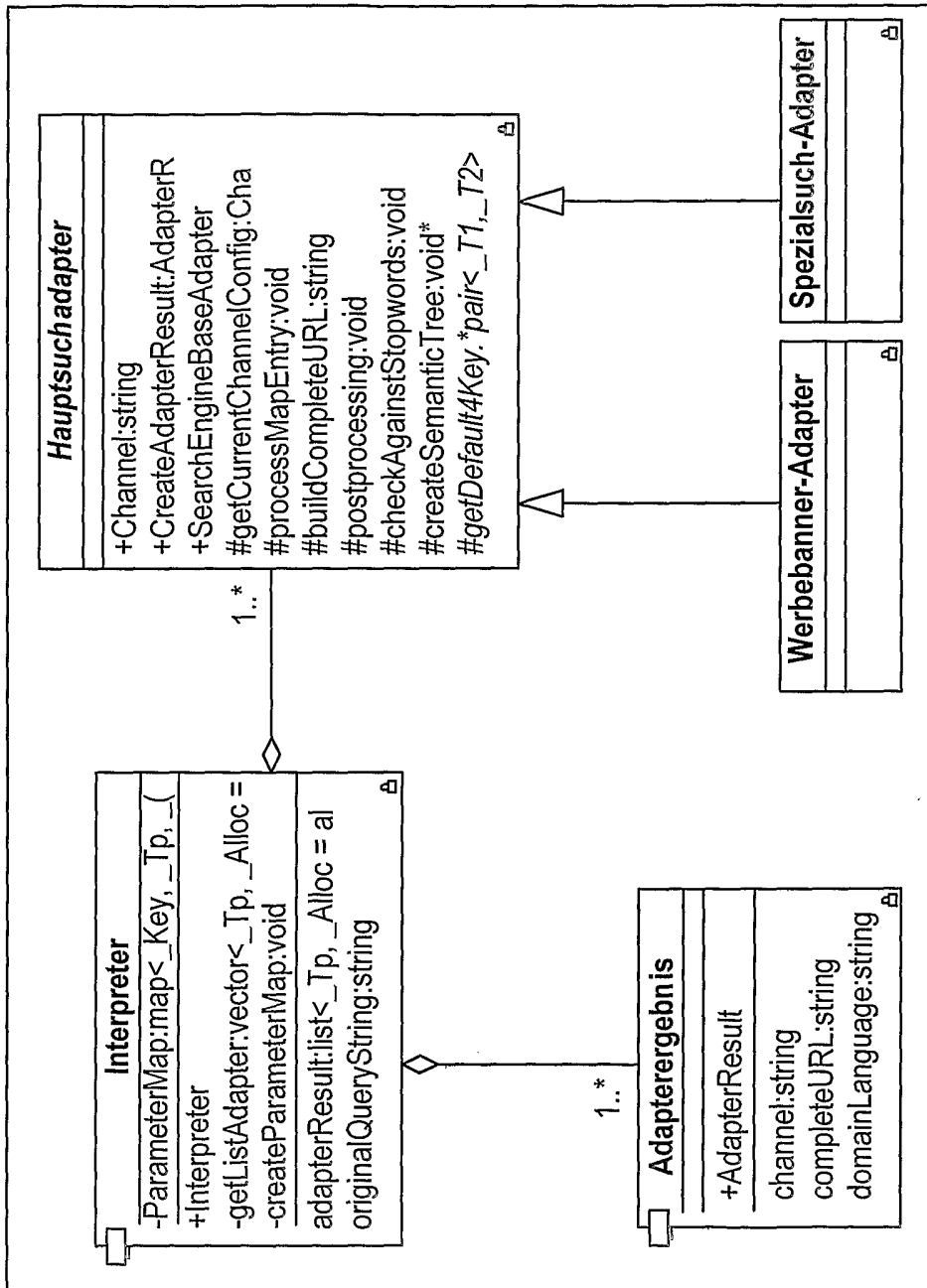


FIG. 11

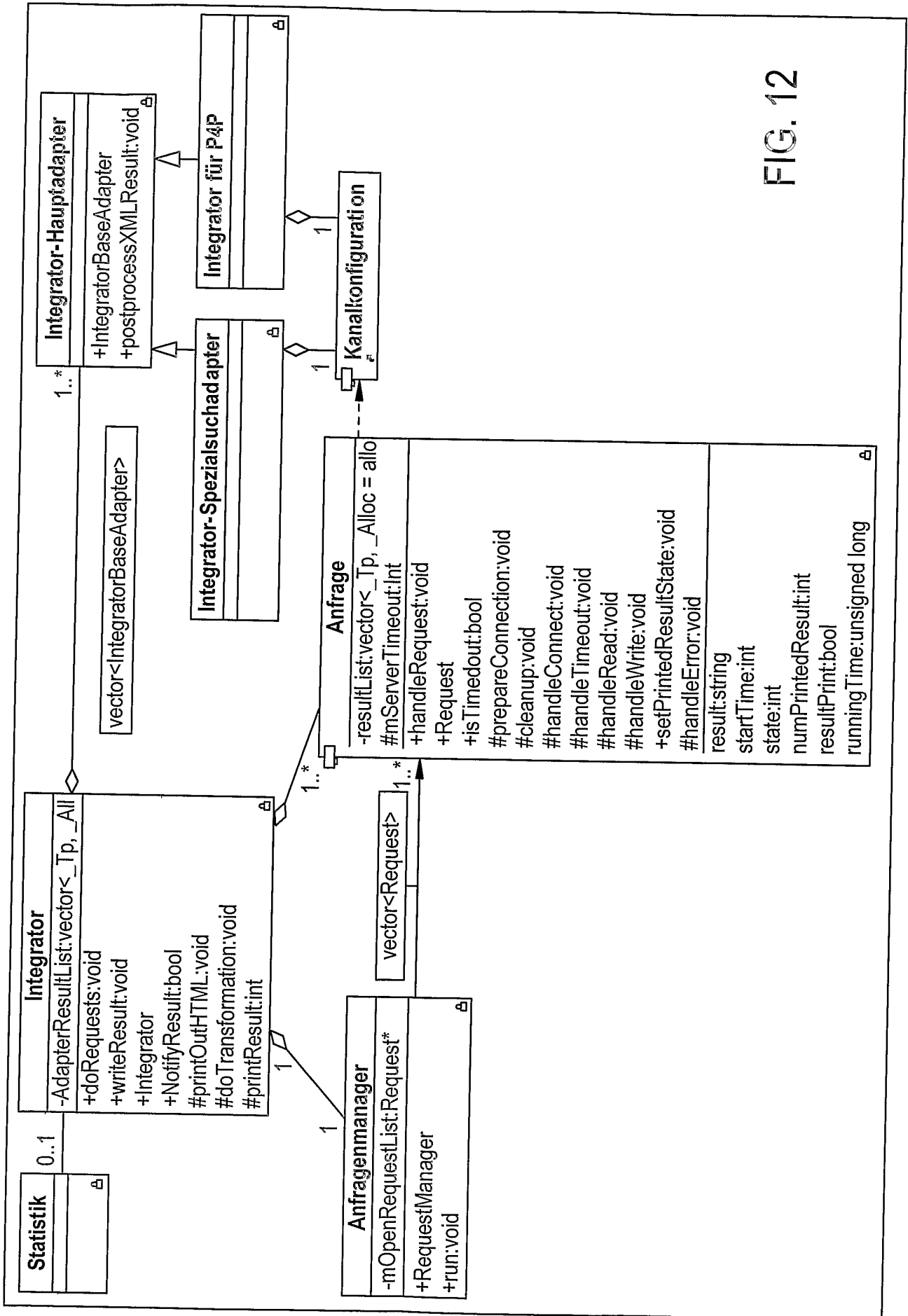


FIG. 12

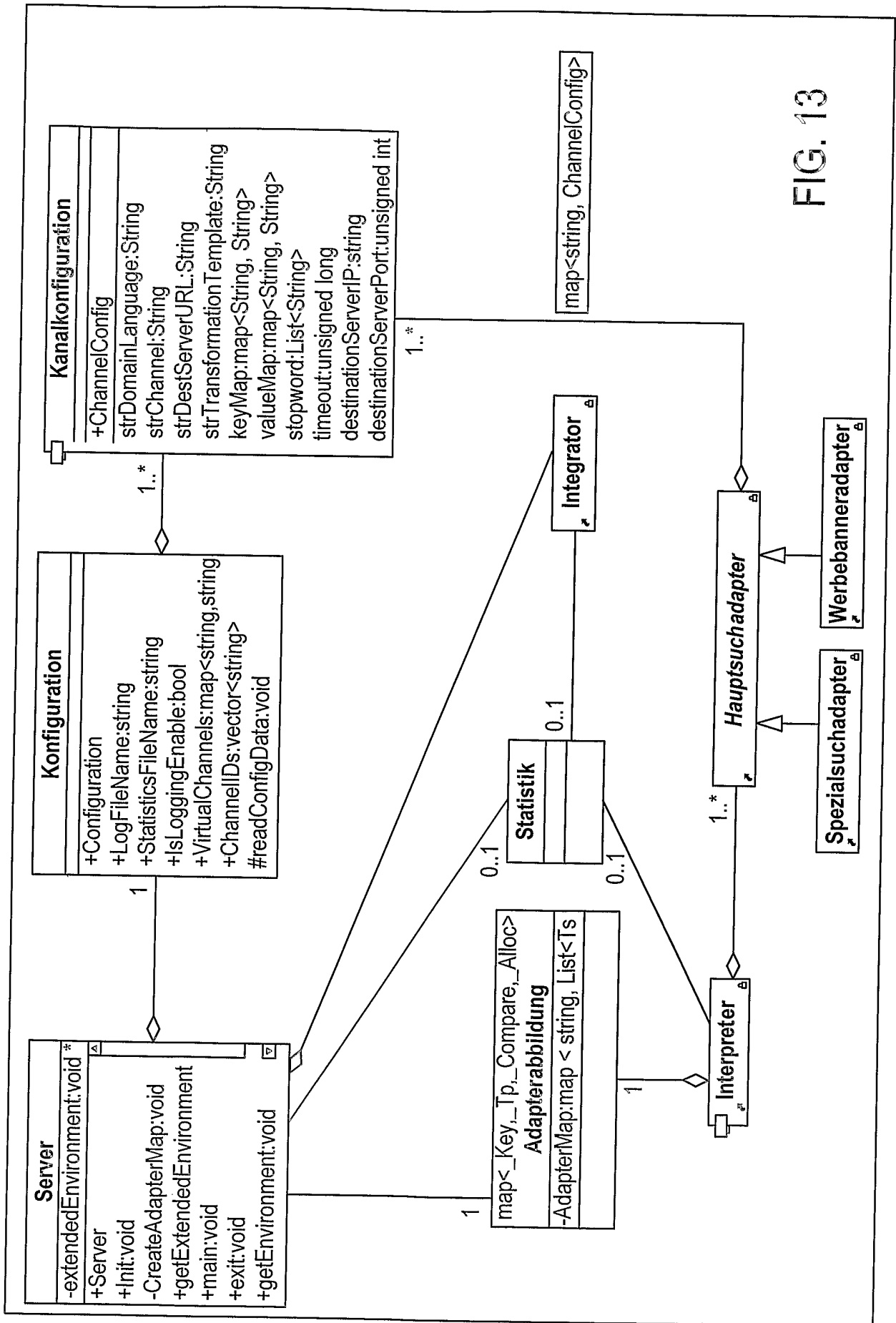


FIG. 13



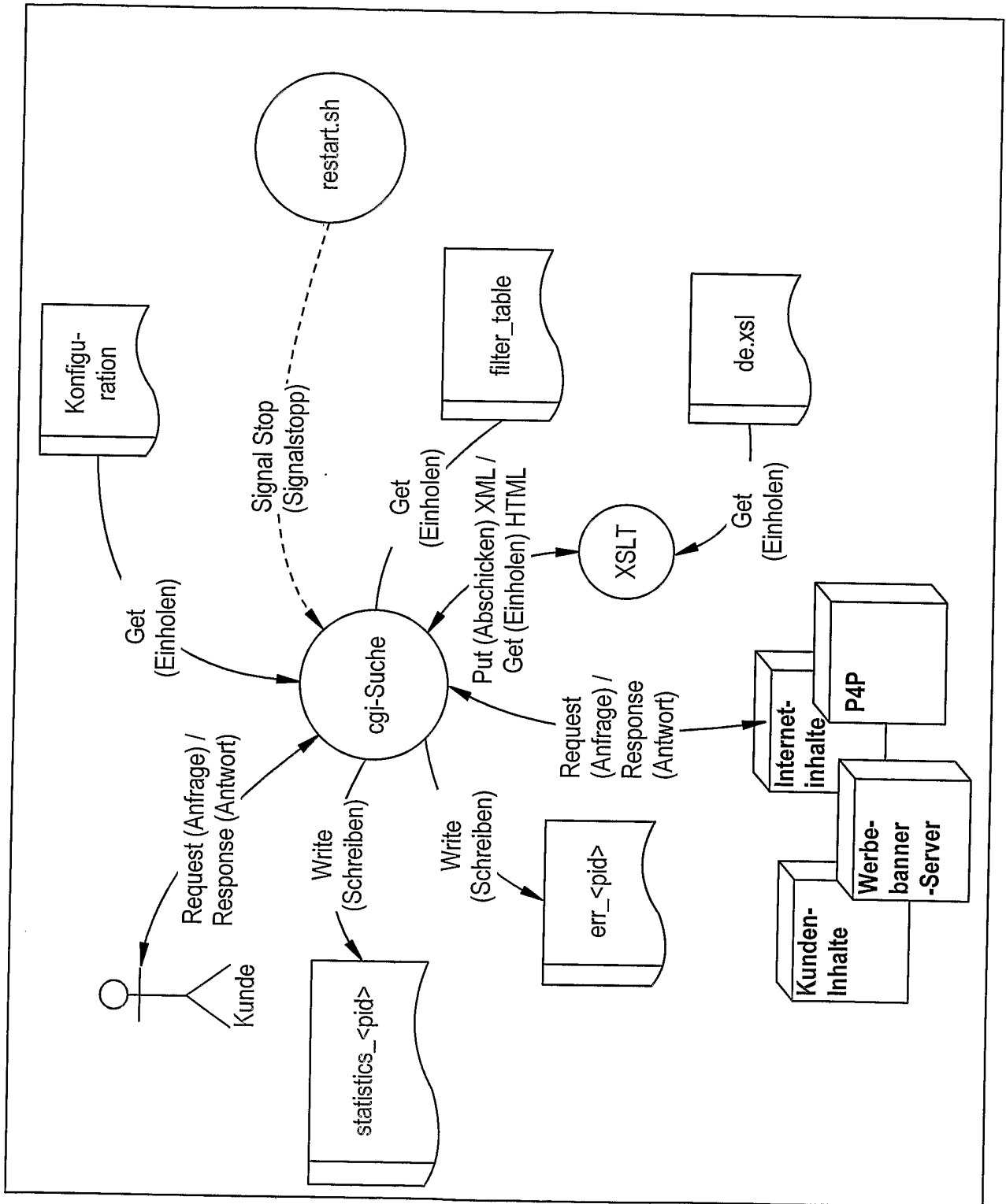


FIG. 14

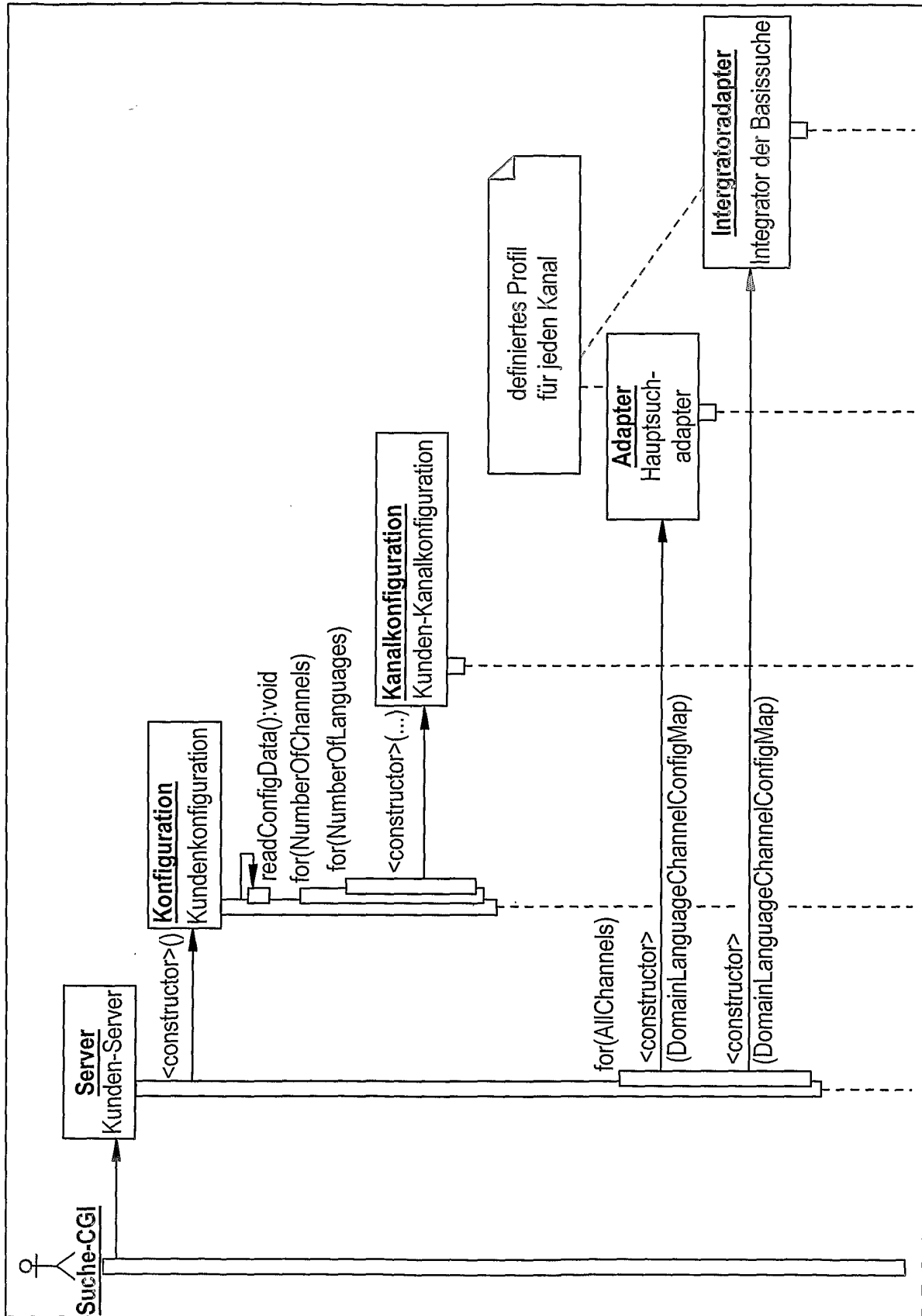


FIG. 15

FIG. 16

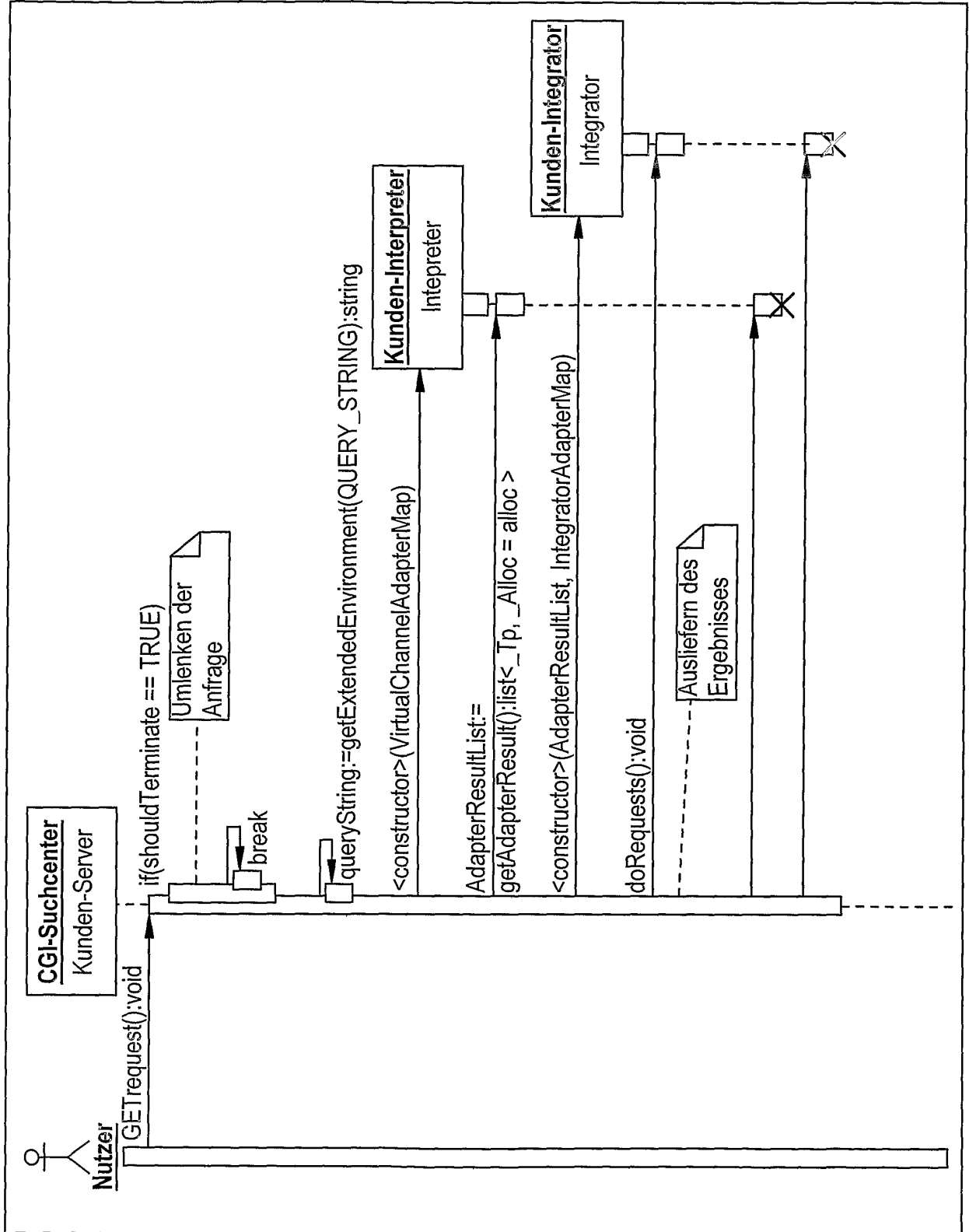
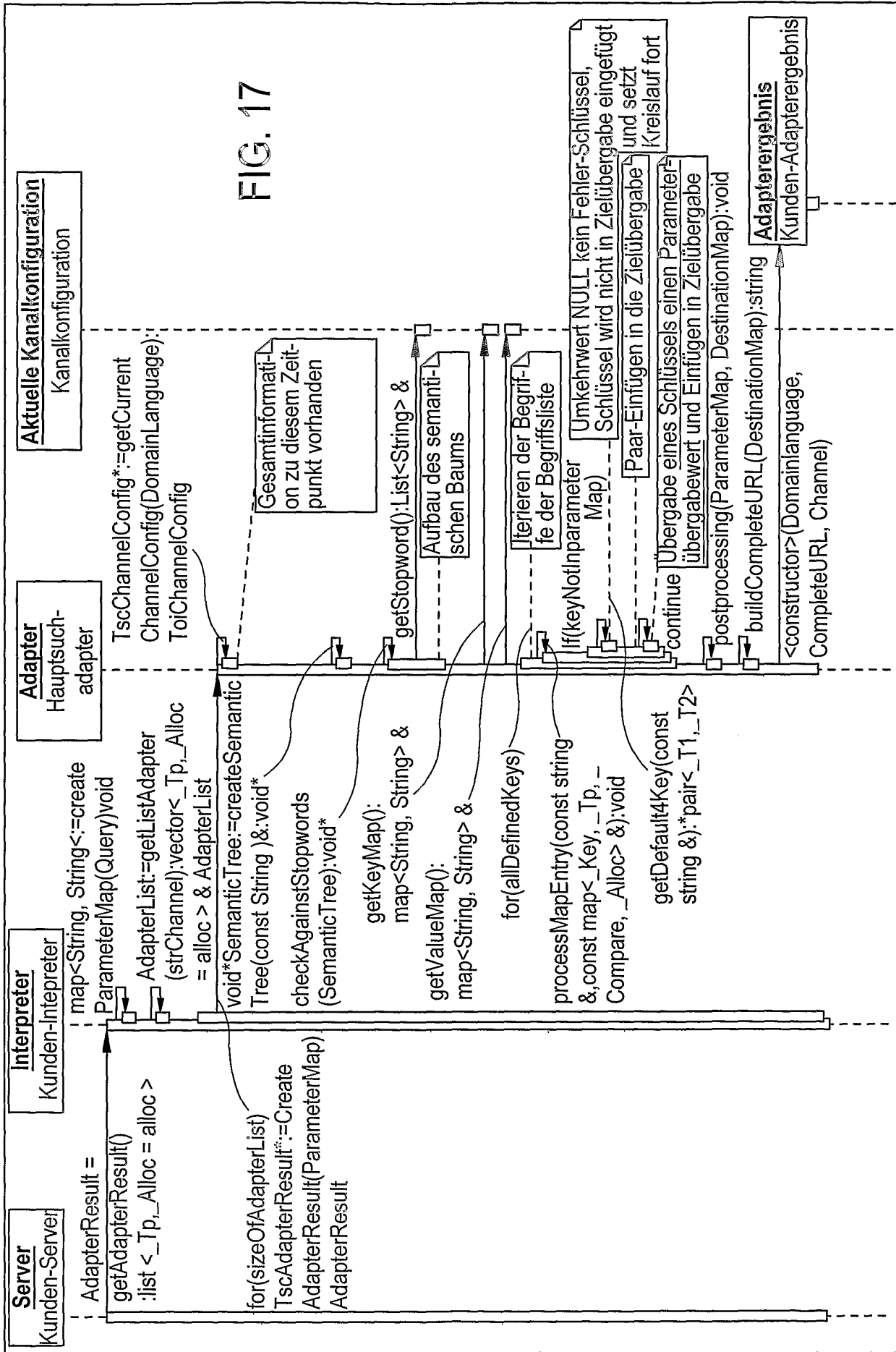


FIG. 17



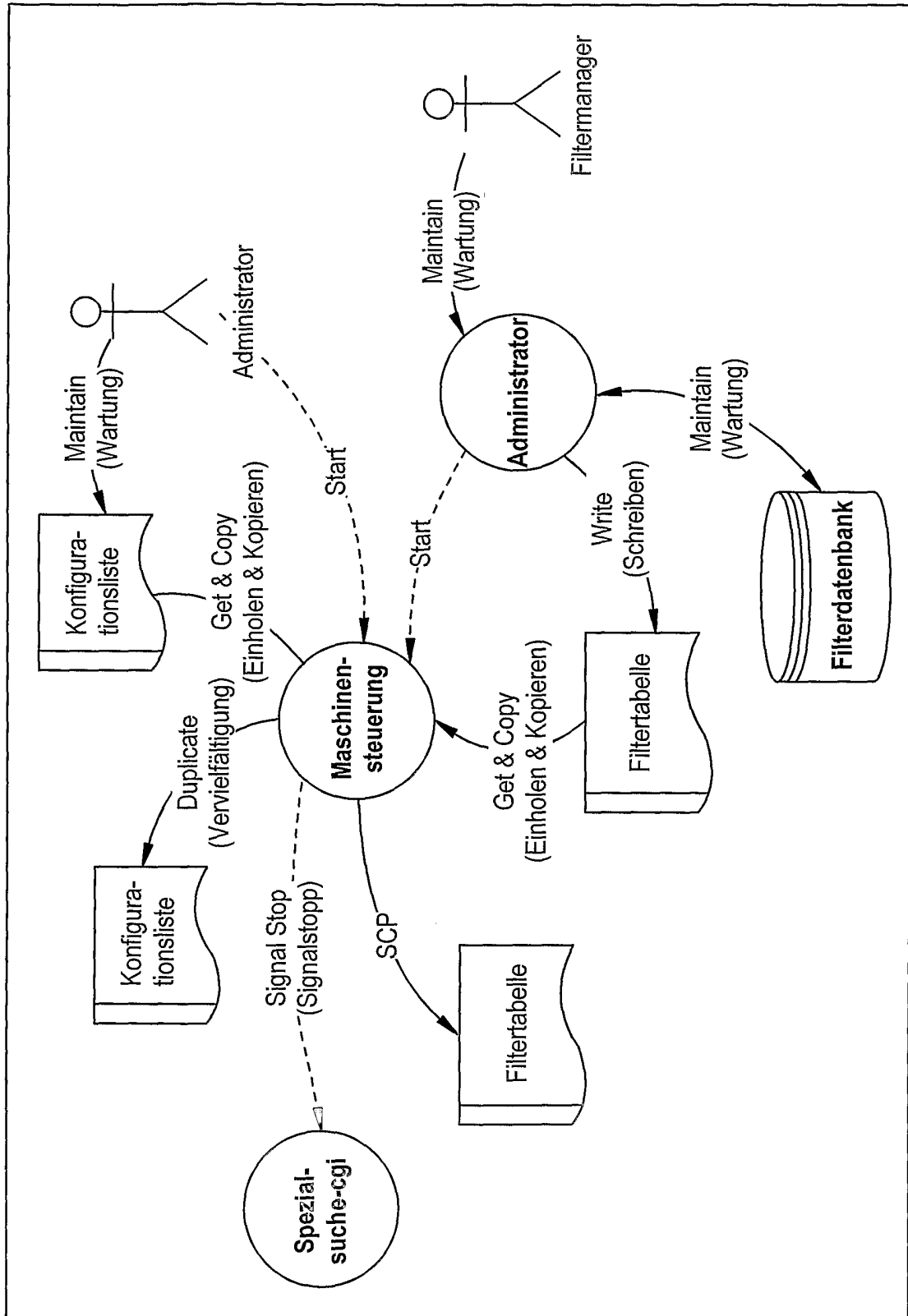


FIG. 18

**INTERNATIONAL SEARCH REPORT**

International Application No  
PCT/EP2004/002915

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC 7 G06F17/30		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) IPC 7 G06F		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal, INSPEC		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6 102 969 A (LAUCKHART GREGORY ET AL) 15 August 2000 (2000-08-15) the whole document	1-15
X Y	DE 100 56 681 A (HEWLETT PACKARD CO) 23 May 2002 (2002-05-23)  column 2, line 13 - column 4, line 6 column 5, line 11 - column 7, line 66 claims 23,24	1, 3, 4, 10, 12, 13 2, 5-9, 11, 14, 15
----- -/--		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C.		
<input checked="" type="checkbox"/> Patent family members are listed in annex.		
° Special categories of cited documents :		
*A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed		
*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. *&* document member of the same patent family		
Date of the actual completion of the international search  7 September 2004	Date of mailing of the international search report  17/09/2004	
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2260 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer  Abbing, R	

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/EP2004/002915

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1 050 830 A (XEROX CORP) 8 November 2000 (2000-11-08) column 1, paragraph 4 - column 3, paragraph 11 column 4, paragraph 17 column 10, paragraph 42 - column 11, paragraph 44 column 12, paragraph 52 - column 13, paragraph 54	1-3, 10-12
A	GRAVANO L ET AL: "STARTS: STANFORD PROPOSAL FOR INTERNET META-SEARCHING" SIGMOD RECORD, ASSOCIATION FOR COMPUTING MACHINERY, NEW YORK, US, vol. 26, no. 2, 1 June 1997 (1997-06-01), pages 207-218, XP000730508	1,10
Y	the whole document	2,5-8, 11,14
Y	WO 01/63484 A (INKTOMI CORP) 30 August 2001 (2001-08-30) abstract	9,15
A	US 2002/049756 A1 (CHANDRASHEKAR SRIDHAR ET AL) 25 April 2002 (2002-04-25) abstract page 1, column 1, paragraph 6 - page 1, column 2, paragraph 8 page 1, column 2, paragraph 16 page 2, column 2, paragraph 24 - page 3, column 2, paragraph 35	1-3, 10-12
A	US 6 430 553 B1 (FERRET BERNARD A) 6 August 2002 (2002-08-06) column 1, line 45 - column 2, line 3 column 2, line 40 - column 3, line 3 column 5, line 49 - column 7, line 39	1-3, 10-12
A	US 2002/143744 A1 (KIM AHRITTA J ET AL) 3 October 2002 (2002-10-03) abstract page 1, column 1, paragraph 6 - page 2, column 1, paragraph 14 page 2, column 2, paragraph 20	1-3, 10-12

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP2004/002915

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 6102969	A	15-08-2000	US 6085186 A	04-07-2000
			AT 255246 T	15-12-2003
			AU 731420 B2	29-03-2001
			AU 4499497 A	14-04-1998
			CA 2269227 A1	26-03-1998
			CN 1249042 A	29-03-2000
			EP 1015964 A2	05-07-2000
			IL 129066 A	04-01-2004
			JP 2001501003 T	23-01-2001
			KR 2000048514 A	25-07-2000
			NZ 335247 A	28-04-2000
			WO 9812881 A2	26-03-1998
DE 10056681	A	23-05-2002	DE 10056681 A1	23-05-2002
EP 1050830	A	08-11-2000	US 6327590 B1	04-12-2001
			EP 1050830 A2	08-11-2000
WO 0163484	A	30-08-2001	US 6564208 B1	13-05-2003
			AU 4321701 A	03-09-2001
			WO 0163484 A2	30-08-2001
			US 2003120646 A1	26-06-2003
US 2002049756	A1	25-04-2002	NONE	
US 6430553	B1	06-08-2002	NONE	
US 2002143744	A1	03-10-2002	US 2004030690 A1	12-02-2004



INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen  
PCT/EP2004/002915

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES  
IPK 7 G06F17/30

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RECHERCHIERTER GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)  
IPK 7 G06F

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)  
EPO-Internal, INSPEC

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	US 6 102 969 A (LAUCKHART GREGORY ET AL) 15. August 2000 (2000-08-15) das ganze Dokument	1-15
X Y	DE 100 56 681 A (HEWLETT PACKARD CO) 23. Mai 2002 (2002-05-23)  Spalte 2, Zeile 13 - Spalte 4, Zeile 6 Spalte 5, Zeile 11 - Spalte 7, Zeile 66 Ansprüche 23,24	1, 3, 4, 10, 12, 13 2, 5-9, 11, 14, 15

Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen

Siehe Anhang Patentfamilie

- \* Besondere Kategorien von angegebenen Veröffentlichungen :
- \*A\* Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist
- \*E\* älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist
- \*L\* Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)
- \*O\* Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht
- \*P\* Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist
- \*T\* Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist
- \*X\* Veröffentlichung von besonderer Bedeutung, die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden
- \*Y\* Veröffentlichung von besonderer Bedeutung, die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist
- \*Z\* Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche  7. September 2004	Absenddatum des internationalen Recherchenberichts  17/09/2004
Name und Postanschrift der Internationalen Recherchenbehörde Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL - 2260 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Bevollmächtigter Bediensteter  Abbing, R

## INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen  
PCT/EP2004/002915

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	EP 1 050 830 A (XEROX CORP) 8. November 2000 (2000-11-08) Spalte 1, Absatz 4 - Spalte 3, Absatz 11 Spalte 4, Absatz 17 Spalte 10, Absatz 42 - Spalte 11, Absatz 44 Spalte 12, Absatz 52 - Spalte 13, Absatz 54	1-3, 10-12
A	----- GRAVANO L ET AL: "STARTS: STANFORD PROPOSAL FOR INTERNET META-SEARCHING" SIGMOD RECORD, ASSOCIATION FOR COMPUTING MACHINERY, NEW YORK, US, Bd. 26, Nr. 2, 1. Juni 1997 (1997-06-01), Seiten 207-218, XP000730508	1,10
Y	das ganze Dokument	2,5-8, 11,14
Y	----- WO 01/63484 A (INKTOMI CORP) 30. August 2001 (2001-08-30) Zusammenfassung	9,15
A	----- US 2002/049756 A1 (CHANDRASHEKAR SRIDHAR ET AL) 25. April 2002 (2002-04-25) Zusammenfassung Seite 1, Spalte 1, Absatz 6 - Seite 1, Spalte 2, Absatz 8 Seite 1, Spalte 2, Absatz 16 Seite 2, Spalte 2, Absatz 24 - Seite 3, Spalte 2, Absatz 35	1-3, 10-12
A	----- US 6 430 553 B1 (FERRET BERNARD A) 6. August 2002 (2002-08-06) Spalte 1, Zeile 45 - Spalte 2, Zeile 3 Spalte 2, Zeile 40 - Spalte 3, Zeile 3 Spalte 5, Zeile 49 - Spalte 7, Zeile 39	1-3, 10-12
A	----- US 2002/143744 A1 (KIM AHRITTA J ET AL) 3. Oktober 2002 (2002-10-03) Zusammenfassung Seite 1, Spalte 1, Absatz 6 - Seite 2, Spalte 1, Absatz 14 Seite 2, Spalte 2, Absatz 20	1-3, 10-12

**INTERNATIONALER RECHERCHENBERICHT**

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/EP2004/002915

Im Recherchenbericht angeführtes Patentdokument		Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
US 6102969	A	15-08-2000	US 6085186 A	04-07-2000
			AT 255246 T	15-12-2003
			AU 731420 B2	29-03-2001
			AU 4499497 A	14-04-1998
			CA 2269227 A1	26-03-1998
			CN 1249042 A	29-03-2000
			EP 1015964 A2	05-07-2000
			IL 129066 A	04-01-2004
			JP 2001501003 T	23-01-2001
			KR 2000048514 A	25-07-2000
			NZ 335247 A	28-04-2000
			WO 9812881 A2	26-03-1998
-----				
DE 10056681	A	23-05-2002	DE 10056681 A1	23-05-2002
-----				
EP 1050830	A	08-11-2000	US 6327590 B1	04-12-2001
			EP 1050830 A2	08-11-2000
-----				
WO 0163484	A	30-08-2001	US 6564208 B1	13-05-2003
			AU 4321701 A	03-09-2001
			WO 0163484 A2	30-08-2001
			US 2003120646 A1	26-06-2003
-----				
US 2002049756	A1	25-04-2002	KEINE	
-----				
US 6430553	B1	06-08-2002	KEINE	
-----				
US 2002143744	A1	03-10-2002	US 2004030690 A1	12-02-2004
-----				