



(19)
 Bundesrepublik Deutschland
 Deutsches Patent- und Markenamt

(10) DE 10 2008 030 125 A1 2010.01.14

(12)

Offenlegungsschrift

(21) Aktenzeichen: 10 2008 030 125.6

(22) Anmeldetag: 27.06.2008

(43) Offenlegungstag: 14.01.2010

(51) Int Cl.⁸: **G06N 3/02** (2006.01)

G06F 17/30 (2006.01)

G06F 17/27 (2006.01)

(71) Anmelder:

Klüver, Christina, Dr., 45130 Essen, DE; Klüver, Jürgen, Prof. Dr., 45130 Essen, DE

(74) Vertreter:

COHAUSZ DAWIDOWICZ HANNIG & SOZIEN, 40237 Düsseldorf

(72) Erfinder:

gleich Anmelder

(56) Für die Beurteilung der Patentfähigkeit in Betracht gezogene Druckschriften:

US 2007/01 92 863 A1

US 2008/01 12 620 A1

KOOTHs, S.: Management Support und Business Intelligence. In: MICE 2004, S.1-58 & Bib. www.kooths.de/mice/download/mice_pr-kooths_OPC-MSS-Muenster2004_de.pdf

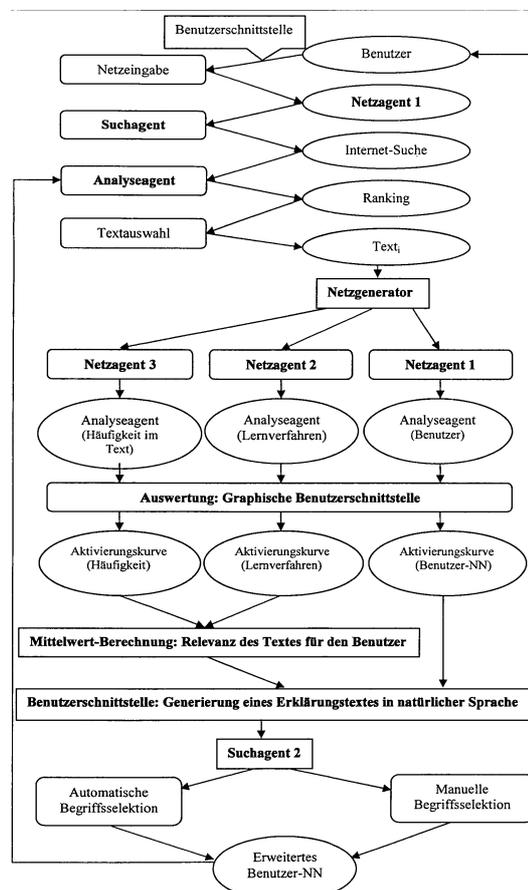
A. Schmidt: "A Modular Neural Network Architecture with Additional Generalization Abilities for High Dimensional Input Vectors", Manchester Metropolitan University, 1996, Department of Computing, S. i-x, 1-113

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

Prüfungsantrag gemäß § 44 PatG ist gestellt.

(54) Bezeichnung: **Such- und Auswerteverfahren sowie Such- und Auswertesystem**

(57) Zusammenfassung: Such- und Auswerteverfahren sowie Such- und Auswertesystem zur Auswertung von elektronisch gespeicherten Texten mittels eines Computersystems und neuronaler Netze unter Verwendung von vorgebbaren oder zuvor ermittelten Suchbegriffen, wobei ein jedes Neuron eine Aktivierung aufweist und über eine gewichtete Verbindung mit zumindest einem anderen Neuron verbunden ist, wobei das Produkt aus der Aktivierung des Neurons mit dem Verbindungsgewicht eine Endaktivierung bildet, wobei mehrere Suchbegriffe vorgegeben werden und zumindest zwei neuronale Netze generiert werden, wobei jeder Suchbegriff jeweils einem Neuron eines jeden Netzes zugeordnet wird, wobei bei dem ersten neuronalen Netz die Verbindungsgewichte konstant sind und die Aktivierung eines jeden Neurons in Abhängigkeit der Häufigkeit des Suchbegriffes im auszuwertenden Text erfolgt und wobei bei dem zweiten neuronalen Netz die Verbindungsgewichte in Abhängigkeit der Häufigkeit des Suchbegriffes im auszuwertenden Text sind und die Aktivierung konstant ist, und eine Berechnung der jeweiligen Endaktivierungen beider Netze erfolgt.



Beschreibung

[0001] Die Erfindung betrifft ein Such- und Auswerteverfahren zur Auswertung von elektronisch gespeicherten Texten mittels eines Computersystems und neuronaler Netze unter Verwendung von vorgebbaren oder zuvor ermittelten Suchbegriffen, wobei ein jedes Neuron eine Aktivierung aufweist und über eine gewichtete Verbindung mit zumindest einem anderen Neuron verbunden ist, wobei das Produkt aus der Aktivierung des Neurons mit dem Verbindungsgewicht eine Endaktivierung bildet. Des Weiteren betrifft die Erfindung ein Such- und Auswertesystem zur Auswertung von elektronisch gespeicherten Texten, insbesondere zur Anwendung des Verfahrens, aufweisend eine Benutzerschnittstelle zum Erhalt von Suchbegriffen, sowie eine Suchmaschinenschnittstelle zur Veranlassung einer Suche unter Verwendung der Suchbegriffe und zum Erhalt des Resultates der Suche und der aufgefundenen Texte.

[0002] Es sind verschiedenste Such- und Auswerteverfahren zur Auswertung von elektronisch gespeicherten Texten mittels eines Computersystems bekannt. Üblicherweise werden dabei unter Verwendung von einem oder mehreren Suchbegriffen elektronisch gespeicherte Texte nach diesen Begriffen durchsucht und anschließend in Ergebnislisten aufgeführt.

[0003] Ebenfalls bekannt sind neuronale Netze, bei denen ein jedes Neuron eine Aktivierung aufweist und über eine gewichtete Verbindung mit zumindest einem anderen Neuron verbunden ist, wobei das Produkt aus der Aktivierung des Neurons mit dem Verbindungsgewicht eine Endaktivierung bildet.

[0004] Nachteilig bei den bekannten Suchverfahren ist, es dass lediglich die Tatsache, dass ein Suchbegriff in dem durchsuchten Text enthalten ist, aus der Ergebnisliste erschlossen werden kann, es erfolgt jedoch keine Auswertung der Texte hinsichtlich ihrer Relevanz in Bezug auf die verwendeten Suchbegriffe. Dadurch dass keine Sortierung hinsichtlich der Relevanz erfolgt, werden zum Teil unbrauchbare Ergebnislisten mit einer unüberschaubar großen Trefferanzahl erzeugt, die aufgrund der Listenlänge unbrauchbar sind.

[0005] Die Aufgabe der Erfindung ist es, ein Verfahren und ein System zur Suche und Auswertung elektronisch gespeicherter Texte bereitzustellen, mittels dessen eine Ermittlung und Darstellung der Relevanz eines durchsuchten und ausgewerteten Textes in Bezug auf die verwendeten Suchbegriffe ermöglicht wird, um insbesondere eine Sortierung der durchsuchten und ausgewerteten Texte nach Relevanz zu ermöglichen.

[0006] Diese Aufgabe wird erfindungsgemäß durch

ein Verfahren gemäß Anspruch 1 sowie ein System gemäß Anspruch 8 gelöst. Vorteilhafte Weiterbildungen des erfindungsgemäßen Verfahrens sowie des erfindungsgemäßen Systems sind in den abhängigen Ansprüchen angegeben.

[0007] Besonders vorteilhaft bei dem erfindungsgemäßen Such- und Auswerteverfahren zur Auswertung von elektronisch gespeicherten Texten mittels eines Computersystems und neuronaler Netze unter Verwendung von vorgebbaren oder zuvor ermittelten Suchbegriffen, wobei ein jedes Neuron eine Aktivierung aufweist und über eine gewichtete Verbindung mit zumindest einem anderen Neuron verbunden ist, wobei das Produkt aus der Aktivierung des Neurons mit dem Verbindungsgewicht eine Endaktivierung bildet, ist es, dass mehrere Suchbegriffe vorgegeben werden und zumindest zwei neuronale Netze generiert werden, wobei jeder Suchbegriff jeweils einem Neuron eines jeden Netzes zugeordnet wird, wobei bei dem ersten neuronalen Netz die Verbindungsgewichte konstant sind und die Aktivierung eines jeden Neurons in Abhängigkeit der Häufigkeit des Suchbegriffes im auszuwertenden Text erfolgt, und wobei bei dem zweiten neuronalen Netz die Verbindungsgewichte in Abhängigkeit der Häufigkeit des Suchbegriffes im auszuwertenden Text sind und die Aktivierung konstant ist, und eine Berechnung der jeweiligen Endaktivierungen beider Netze erfolgt.

[0008] Unter dem Begriff eines elektronisch gespeicherten Textes ist dabei jede elektronisch gespeicherte Datei enthaltend Daten und/oder Zeichenfolgen zu subsumieren. Unter dem Begriff eines Suchbegriffes sind dabei Wörter oder Wortteile einer natürlichen Sprache oder jedwede sonstige beliebige Zeichenfolge zu subsumieren. Da auch Wortteile einer natürlichen Sprache Verwendung finden können, können somit auch beliebige Trunkierungen bei der Suche und Auswertung zugelassen werden, d. h. dass die benutzten Suchbegriffe insbesondere am Beginn und/oder am Ende mit einem Platzhalter versehen sein können.

[0009] Vorzugsweise wird für jeden Suchbegriff der arithmetische Mittelwert aus beiden Endaktivierungen ermittelt. Dieser Mittelwert kann als Maß für die Bedeutung des Begriffes für die analysierte Datei respektive umgekehrt als Maß für die Relevanz des Textes in Bezug auf den Suchbegriff dienen und insbesondere für weitere Auswertungen und/oder Ergebnisdarstellungen herangezogen werden.

[0010] Durch die Berücksichtigung beider Endaktivierungen kann ein zuverlässiges Maß für die Relevanz der analysierten Datei in Bezug auf den verwendeten Suchbegriff abgeleitet werden.

[0011] Bevorzugt wird eine Sortierung der Suchbegriffe nach absteigender Endaktivierung durchge-

führt, d. h. dass jenen Suchbegriffen eine größere Beadeutung zugemessen wird, deren jeweilige Endaktivierung größer ist, gegenüber solchen Suchbegriffen mit einer kleineren Endaktivierung.

[0012] In einer bevorzugten Ausführungsform des Verfahrens handelt es sich bei den Suchbegriffen um mittels einer vorhergehenden Suche unter Verwendung eines ersten Suchbegriffes ermittelten korrespondierende weitere Begriffe.

[0013] Vorzugsweise wird nach der Auswertung von mehreren Texten eine Rangliste der Texte nach Relevanz gebildet, wobei die Relevanz aus der Endaktivierung eines oder beider neuronaler Netze und/oder aus einem Mittelwert der Endaktivierungen beider neuronaler Netze abgeleitet wird.

[0014] Bevorzugt sind mittels einer vorgebbaren Negativliste bestimmte Begriffe von einer Suche und/oder Auswertung ausnehmbar. Hierdurch können triviale Resultate durch das vorherige Ausnehmen bestimmter Begriffe oder Zeichenfolge vermieden werden.

[0015] Vorzugsweise werden mittels einer ersten Anwendung des Verfahrens korrespondierende Suchbegriffe zur Erweiterung der Suche ermittelt, mittels derer unter erneuter Anwendung des Verfahrens iterativ eines oder mehrere weitere Paare von neuronalen Netzen erzeugt und deren Endaktivierungen ermittelt werden.

[0016] Durch eine iterative Anwendung des Verfahrens können zum einen korrespondierende Suchbegriffe, die im Kontext der primär verwendeten Suchparameter stehen identifiziert werden, und zum anderen die Qualität der Suche und Auswertung enorm verbessert werden.

[0017] Besonders vorteilhaft bei dem erfindungsgemäßen Such- und Auswertesystem zur Auswertung von elektronisch gespeicherten Texten, insbesondere zur Anwendung des erfindungsgemäßen Verfahrens, aufweisend eine Benutzerschnittstelle zum Erhalt von Suchbegriffen, sowie eine Suchmaschinenschnittstelle zur Veranlassung einer Suche unter Verwendung der Suchbegriffe und zum Erhalt des Resultates der Suche und der aufgefundenen Texte, ist es, dass das System einen Netzgenerator aufweist, mittels dessen zumindest zwei neuronale Netze generiert werden, wobei ein jedes Neuron eine Aktivierung aufweist und über eine gewichtete Verbindung mit zumindest einem anderen Neuron verbunden ist, wobei das Produkt aus der Aktivierung des Neurons mit dem Verbindungsgewicht eine Endaktivierung bildet, wobei jeder Suchbegriff jeweils einem Neuron eines jeden Netzes zugeordnet wird, und wobei eine Textauswerteeinheit vorgesehen ist, mittels derer die absoluten und/oder relativen Häufigkeiten der Suchbe-

griffe in einem Text ermittelt und an den Netzgenerator übergeben werden, wobei mittels des Netzgenerators bei dem ersten neuronalen Netz die Aktivierung eines jeden Neurons in Abhängigkeit der Häufigkeit des zugeordneten Suchbegriffes bei konstanten Verbindungsgewichten und bei dem zweiten neuronalen Netz die Verbindungsgewichte in Abhängigkeit der Häufigkeit des zugeordneten Suchbegriffes bei konstanter Aktivierung durchgeführt wird und die jeweiligen Endaktivierungen ermittelt werden, und aufweisend eine Ausgabereinheit, mittels derer die Endaktivierungen der neuronalen Netze ausgegeben werden.

[0018] Vorzugsweise weist das System eine graphische Benutzerschnittstelle auf, mittels derer die Endaktivierungen der beiden neuronalen Netze in Bezug auf die verwendeten Suchbegriffe graphisch dargestellt werden bzw. darstellbar sind.

[0019] Hierdurch kann die Relevanz der analysierten Texte und Dateien sowie die Relevanz der verwendeten Suchbegriffe zueinander in vorteilhafter und leicht verständlicher Weise abgebildet werden.

[0020] Besonders bevorzugt weist das System eine Benutzerschnittstelle auf, mittels derer eine Erläuterung in einer natürlichen Sprache über die Relevanz der Suchbegriffe in Bezug auf den ausgewerteten Text ausgegeben wird, wobei die Erläuterung mittels der Textauswerteeinheit aus den Endaktivierungen der beiden neuronalen Netze in Bezug auf die verwendeten Suchbegriffe generiert und.

[0021] Durch das automatische Generieren eines Erklärungstextes in natürlicher Sprache erfolgt quasi eine Übersetzung des numerischen Analyseergebnisses in ein unmittelbar verständliches Resultat, dem die Relevanz der verwendeten Suchbegriffe entnehmbar ist.

[0022] Bevorzugt wird mittels der Ausgabereinheit eine Rangliste einer Mehrzahl von ausgewerteten Texten gebildet und ausgegeben, wobei die Rangliste sich nach der Relevanz der Texte in Bezug auf die Suchbegriffe richtet.

[0023] Vorzugsweise weist das System Datenspeicherungsmittel zur Abspeicherung und/oder Zwischenspeicherung der Suchbegriffe und/oder der aufgefundenen und/oder ausgewerteten Texte und/oder der neuronalen Netze sowie Netzparameter und/oder von Ranglisten auf. Durch die Anordnung von flüchtigen und/oder dauerhaften Datenspeichern können die verschiedenen Dateien, Parameter, Ergebnisse, etc. stets schnell verfügbar vorgehalten werden.

[0024] Besonders bevorzugt ist eine Schnittstelle zum Erhalt von durch eine vorhergehende Suche er-

mittelten zu den ersten Suchbegriffen korrespondierenden weiteren Begriffen vorgesehen, wobei unter Verwendung der weiteren Begriffe das erste und zweite gebildete neuronale Netz erweitert und/oder ein drittes und viertes neuronales Netz mittels dem Netzgenerator gebildet wird. Es besteht hierdurch die Möglichkeit einer iterativen Anwendung und/oder vertieften und/oder erweiterten Suche mittels des erfindungsgemäßen Systems.

[0025] Vorzugsweise ist ein Filter vorgesehen, mittels dessen eine vorgebbare Negativliste von Begriffen von einer Suche und/oder Auswertung ausnehmbar ist, um dadurch Trivialergebnisse zu vermeiden, wie beispielsweise die Berücksichtigung eines Artikels oder eines Bindewortes einer natürlichen Sprache, da hierdurch das Ergebnis der Datenanalyse stark verfälscht werden könnte.

[0026] Mittels des erfindungsgemäßen Verfahrens und des erfindungsgemäßen Systems können somit insbesondere die Suchergebnisse von Internet-Suchmaschinen ausgewertet werden, d. h. dass das Verfahren und das System insbesondere in Form einer so genannten Meta-Suchmaschine, die für eine erste Suche auf andere, externe Suchmaschinen zurückgreift, betrieben werden kann. Es kann jedoch auch eine Primärsuchfunktion integriert sein, so dass nicht zunächst auf externe Suchmaschinen zurückgegriffen werden muss, sondern bereits die Primärsuche eigenständig durchgeführt wird.

[0027] Bei dem dargestellten Ausführungsbeispiel bildet das erfindungsgemäße System eine Meta-Suchmaschine. Eine Meta-Suchmaschine ist ein Programm, das auf bereits vorhandenen Suchmaschinen wie beispielsweise Google oder Yahoo oder dergleichen operiert. Ziel der Suche ist die Angabe von Texten. Vorzugsweise wird das Verfahren und das System in einem Computernetzwerk wie das Internet ausgeführt bzw. integriert. Die erforderlichen Schnittstellen zur Eingabe und Ausgabe von Daten stehen somit zur Verfügung.

[0028] Wenn also ein Benutzer einen oder mehrere Suchbegriffe in das erfindungsgemäße System eingibt, dann wird automatisch mindestens eine Suchmaschine, gewöhnlich Google, mit der Erstsuche beauftragt. Das Auswertesystem übernimmt die von der Suchmaschine angegebenen Texte und beginnt mit seiner eigentlichen Arbeit, nämlich der Auswertung der gefundenen Texte. Die Suchstrategie von z. B. Google wird dabei durch eine Suchanweisung des Programms modifiziert, da Google sich streng an den Vorgaben des Benutzers orientiert und nicht wie üblich an der Häufigkeit, mit der bestimmte Texte bereits aufgerufen wurden. Die vorliegende Meta-Suchmaschine hat zum allgemeinen Ziel, Informationen aus den von den Suchmaschinen gefundenen Texten zu filtern und zu kondensieren.

[0029] Konkret operiert das System wie folgt: Ein Benutzer des Systems wird zu Beginn aufgefordert, zu dem Thema, zu dem er relevante Texte sucht, mehrere Begriffe in einer Reihenfolge einzugeben, die die Relevanz der Begriffe für den Benutzer repräsentiert. Dazu steht eine Schnittstelle zur Verfügung.

[0030] Eine zweite Eingabemethode ist die der Prioritätslisten. Hier kann jedem Begriff gleichzeitig eine Priorität mitgegeben werden, um so eine erste Struktur festzulegen. Es können somit zu jedem Begriff wahlweise Unterbegriffe zugeordnet werden. Die Darstellung der Begriffe erfolgt in einem Baum dessen Blätter die Begriffe darstellen.

[0031] Alternativ oder kumulativ können die der Analyse zugrunde liegenden Suchbegriffe jedoch auch automatisch aus einer elektronischen Datei und/oder von einem Speichermedium ausgelesen und übernommen werden. Der Begriff des Benutzers ist somit lediglich als Synonym für eine Vorgabe beim Start des Prozesses aufzufassen.

[0032] Mittels des erfindungsgemäßen Verfahrens und des erfindungsgemäßen Systems wird nunmehr aus dieser Begriffsliste ein künstliches neuronales Netz konstruiert. Dieses repräsentiert die Zusammenhangsstruktur der Begriffe aus der Sicht des Benutzers und wird im Folgenden als Benutzernetz bzw. Netzagent 1 bezeichnet. Die Einheiten dieses Netzagenten 1 sind die Begriffe, die der Benutzer vorgibt; die Verbindungen zwischen den Einheiten werden als numerische Werte (Gewichte) entweder nach der Reihenfolge der Benutzereingaben oder gemäß der Prioritätenliste in das Netz eingefügt.

[0033] Anschließend ordnet das System die von der Suchmaschine gefundenen Texte in eine Rankingliste. Für dies Ranking wird ein weiterer Agent aktiviert, der als Suchagent bezeichnet wird, da er nach bestimmten Kriterien die Texte untersucht. Der Suchagent orientiert sein Ranking daran, wie repräsentativ die vom Benutzer vorgegebenen Begriffe für den jeweiligen Text sind, nämlich a) zu wie viel Prozent die Begriffe relativ zur Gesamtzahl aller Wörter im Text enthalten sind, b) wie häufig die Begriffe relativ zur Gesamtzahl der Begriffe in einem Text enthalten sind und c) wie gleich verteilt die Begriffe im Text vorkommen. Für das Kriterium c) gilt, dass ein Text umso höher bewertet wird, je homogener die Begriffe im Text vorkommen. Danach beginnt das System mit der Auswertung des Textes, der im Ranking an oberster Stelle steht.

[0034] Nach der Auswertung wird der Titel des Dokumentes aufgeführt, eine Kurzbeschreibung (die ersten 300 Zeichen des Dokumentes), die Suchmaschine, die das Ergebnis geliefert hat und der Rankingwert in Form eines grünen Balkens.

[0035] Um das originale Dokument zu betrachten kann der Benutzer entweder auf den Titel klicken oder auf die Schablone „Anzeigen“. Es kann auch jeweils eine automatische Anzeige erfolgen. In allen Fällen öffnet sich ein Browserfenster, indem das Dokument dargestellt wird. Unter der Schaltfläche „Details“ können detaillierte Informationen über das Dokument abgerufen werden. Diese sollen dem Benutzer eine komplette Durchsicht des Dokumentes ersparen, d. h., er kann unter Umständen aufgrund der dargebotenen Informationen entscheiden, ob es für ihn sinnvoll ist das Dokument im Ganzen anzusehen oder nicht.

[0036] Analog zur Konstruktion des Benutzernetzes konstruiert das System nun aus den Suchbegriffen zwei weitere Netzagenten (Netzagent 2 und Netzagent 3), die als Textnetze bezeichnet werden. Diese Netze geben an, welche Bedeutung die Suchbegriffe im Text haben und zwar auf der Basis ihrer Häufigkeit im Text. Diese Häufigkeit wird in zwei Netzstrukturen übersetzt und das Netz wird „aktiviert“, d. h. es wird in eine Dynamik versetzt. Dies wird dadurch realisiert, dass die Einheiten mit bestimmten numerischen Anfangswerten versehen werden, worauf der dynamische Prozess in Gang gesetzt wird. Das Ergebnis des dynamischen Prozessierens des Netzes, nämlich die Endaktivierungswerte der Begriffe, ist dann die Bedeutung der Begriffe im Text. Diese Bedeutungsgenerierung wird sowohl in Form visueller Graphiken als auch als sprachliche Interpretation ausgegeben. Zusätzlich kann eine Analyse ausgegeben werden, inwieweit die strukturierte Vorgabe von dem Ergebnis der Textanalyse abweicht.

[0037] Die beiden Netzagenten unterscheiden sich darin, dass Netzagent 2 gleiche Verbindungswerte von 0.5 von und zu allen Einheiten hat. Die Dynamik des Netzagenten 2 entsteht dann dadurch, dass die Einheiten mit Werten aktiviert werden, die der Häufigkeit der Begriffe im Text entsprechen. Beim Netzagenten 3 werden die Werte der Verbindungen proportional zur Häufigkeit im Text festgelegt, während die Aktivierung durch gleiche numerische Werte geschieht. Die jeweiligen Endaktivierungen der beiden Netzagenten werden als Kurven dargestellt, die sich in der Verlaufsform entsprechen. Da die absoluten Endwerte gewöhnlich differieren, wird der arithmetische Mittelwert gebildet, der als Bedeutung interpretiert wird.

[0038] Da derartige Kurven für ungeübte Benutzer gewöhnlich nur schwer zu verstehen sind, werden die Kurven in eine umgangssprachliche Erläuterung übersetzt.

[0039] Die bisher beschriebenen Programmteile und -operationen bilden den ersten Teil des Gesamtsystems. Der Übersichtlichkeit halber werden die einzelnen Komponenten noch einmal schematisch dar-

gestellt:

Komponentenliste 1:

- a) Schnittstelle für den Benutzer zur Eingabe von eigenen Begriffen sowie für die Ausgaben der Ergebnisse an den Benutzer;
- b) Werkzeug „Suche“ zur Anweisung an eine Suchmaschine für eine spezielle Form der Suche (= Modifikation der Operationen der Suchmaschine);
- c) Werkzeug „Agentenkonstruktionen“;
- d) Netzagent 1 (Benutzernetz);
- e) Analyseagent zur Konstruktion eines Ranking, was die Grundlage für die Konstruktion der Netzagenten ist;
- f) Netzagenten 2 und 3 (Textnetze);
- g) Werkzeug „Graphik“ zur Konstruktion der visuellen Darstellung;
- h) Werkzeug „Übersetzung“ zur Konstruktion eines sprachlichen Textes.

[0040] Da das System zusätzlich zur Suchliste des Benutzers durch einen weiteren Suchagenten (Suchagent 2) noch eine zweite Liste von Begriffen erstellt, die im Text enthalten sind, kann der Benutzer entscheiden, ob er sein ursprüngliches Benutzernetz durch neue Begriffe aus dieser zweiten Liste erweitern will. Dies kann manuell oder auch automatisch erfolgen. Im ersten Fall gibt der Benutzer an, welche zusätzlichen Begriffe aus der zweiten Liste in die bisherigen Netzagenten eingefügt werden sollen. Daraus entstehen nach dem gleichen Verfahren, wie oben beschrieben, zwei weitere Netzagenten 3 und 4. Im zweiten Fall sucht das System nach bestimmten Kriterien, nämlich orientiert an Häufigkeiten wie beim Ranking-Verfahren, zusätzliche Begriffe aus der zweiten Liste heraus und konstruiert zwei weitere Netzagenten 5 und 6. In beiden Fällen bestehen also die neuen Netzagenten aus den bisherigen Netzagenten 2 und 3 sowie den entweder manuell eingegebenen oder automatisch selektierten zusätzlichen Begriffen. Falls gewünscht, kann der Benutzer auch beide Optionen nacheinander ausführen lassen.

[0041] Es erfolgt somit seitens des Benutzers nicht etwa ein Eingriff während des Verfahrens, sondern vielmehr ist das Verfahren insbesondere iterativ anwendbar, wobei selbstverständlich die zuvor generierten Analyseergebnisse als Eingangsgrößen eines weiteren Verfahrensdurchlaufes dienen können.

[0042] Je nach gewählter Option wird wieder eine Aktivierung der Netzagenten vorgenommen und das Ergebnis dem Benutzer als die Bedeutung der gesamten Begriffe ausgegeben. Die Ausgabe an den Benutzer geschieht mit den gleichen Werkzeugen wie bereits beschrieben. Mit diesen Erweiterungsverfahren kann der Benutzer jetzt wieder fortfahren oder die Analyse des Textes abbrechen. Dies Erwei-

terungskonzept basiert auf der Erfahrung, dass zu Beginn einer Suche häufig nur wenige Begriffe verfügbar sind und dass schon Zwischenergebnisse der Suche den Benutzer darauf bringen können, seine Suchbegriffe durch die vom Programm zusätzlich gefundenen zu erweitern. Wenn der Benutzer die Textanalyse abbricht, werden die Ergebnisse gespeichert und das Programm beginnt nach dem gleichen Verfahren mit der Interpretation des zweiten Textes in der Rankingliste.

[0043] Um zu gewährleisten, dass das Verfahren und System bei der Erstellung der zweiten Liste nur Begriffe aufnimmt, die für den Text relevant sind, sind etablierte „Negativlisten“ im Programm enthalten. Diese enthalten Wörter der Umgangssprache, die bei der Konstruktion der zweiten Liste nicht berücksichtigt werden sollen. Diese Negativlisten sind Standard und konnten dem System einfach hinzugefügt werden. Der wesentliche Vorteil besteht nun darin, dass der Benutzer selbst diese Negativlisten erweitern kann.

[0044] Auch die eben dargestellten zusätzlichen Komponenten werden noch einmal aufgelistet.

Komponentenliste 2:

- i) Suchagent 2 zur Erstellung einer erweiterten Liste;
- j) Datei „Negativliste“;
- k) Datei Liste 2;
- l) Werkzeug „Option“ für manuelle und/oder automatische Erweiterung der Netzagenten;
- m–p) Netzagenten 3. 4. 5 und 6;
- q) Datei „Ergebnisse“ des ersten Textes.

[0045] Falls der Benutzer nach einer einmaligen Erweiterung der Netzagenten 1 und 2 mit den gleichen Prozeduren fort fährt, entstehen zusätzliche Komponenten, die jedoch im Einzelnen nicht mehr aufgeführt werden. Hinsichtlich einer schematischen Übersicht zu sämtlichen Komponenten siehe [Fig. 1](#).

[0046] Das prinzipiell Neue und besonders vorteilhafte an diesem Verfahren und System besteht insbesondere in folgenden Aspekten:

- a) Die Bedeutung von Begriffen aus der Sicht des Benutzers einerseits und für einen bestimmten Text andererseits wird jeweils als Ergebnis eines dynamischen Netzprozesses aufgefasst. Die Konstruktion der jeweiligen Netze wird durch eine neuartige Lernregel geleistet. Damit folgt die Grundlogik des Verfahrens im Wesentlichen den Prozessen, die im menschlichen Gehirn ablaufen. Derartige Überlegungen für die Konstruktion von Meta-Suchmaschinen sind bisher noch nicht realisiert worden.
- b) Das Programm kann insbesondere in dem Sinne streng am Benutzer orientiert sein, dass er

nicht nur die anfänglichen Vorgaben eingibt. Insbesondere kann die Basis für den gesamten Ablauf des Programms die Sicht des Benutzers sein und nicht so sehr die Orientierung an einer objektiven Textstruktur – diese wird allerdings auch verwendet. Der Benutzer kann insbesondere die Netzerweiterung wieder durch seine eigene Sicht beeinflussen, was für ihn relevant ist oder nicht. Das Verfahren präsentiert dann auch die erweiterten Ergebnisse in Relation zur Sicht des Benutzers.

Hilfe-Funktionen für die Programmnutzung

[0047] Zur Unterstützung der Programmbenutzer wurden drei grundlegende Hilfe-Systeme entwickelt.

- Innerhalb des Programms wurden statische Hilfetexte auf jedem Programmfenster eingefügt, die den Benutzer durch die einzelnen Programmschritte leiten.
- Als Nachschlagewerk dient eine implementierte „Online-Hilfe“, in der, nach Stichpunkten geordnet, alle wesentlichen Bestandteile des Programms erklärt werden.
- Zum schnellen Einstieg in die Bedienung des Programms dient ein Video-Tutorial, bei dem permanent die Programmoberfläche sichtbar ist und über gesprochene Erläuterungen alle Programmschritte in Lektionen unterteilt durchgearbeitet werden.

[0048] Der Benutzer wird also auditiv und visuell angesprochen, wodurch unterschiedliche Bereiche des Gehirns aktiviert werden und ein größerer Lernerfolg möglich ist. Durch diese drei Systeme wird ein Benutzer optimal und kontinuierlich bei der Verwendung des Programms unterstützt. Grundsätzlich wurde bei der Entwicklung zudem auf eine intuitive Bedienbarkeit geachtet.

Vorteile auf einen Blick

- Für den Anfänger auf einem Gebiet: Das Erweiterungskonzept basiert auf der Erfahrung, dass zu Beginn einer Suche häufig nur wenige Begriffe verfügbar sind und dass schon Zwischenergebnisse der Suche den Benutzer darauf bringen können, seine Suchbegriffe durch die vom Programm zusätzlich gefundenen zu erweitern. Somit erhält der Benutzer Synonyme bzw. zusätzliche Begriffe, die eine Recherche effektivieren.
- Für den Experten: Durch die strukturierte und erweiterte Suche kann der Benutzer nach spezialisierten Informationen suchen.
- Der Benutzer kann selbst die Negativlisten erweitern, um unerwünschte Suchergebnisse (z. B. Werbung, Seminar- oder Vorlesungsankündigungen) zu umgehen.
- Die Zusammenhänge zwischen den Begriffen in einem Text können graphisch betrachtet werden.
- Zeitersparnis durch eine effektive und Benutzer

orientierte Recherche.

Primäre Einsatzgebiete

- Für alle, die den neuesten Stand in der Forschung und Technik verfolgen müssen.
- Das Programm kann als eine neue Form des E-Learning verwendet werden. Je nach Wissensstand können die Informationen über verschiedene Texte zugänglich gemacht werden.

Patentansprüche

1. Such- und Auswerteverfahren zur Auswertung von elektronisch gespeicherten Texten mittels eines Computersystems und neuronaler Netze unter Verwendung von vorgebbaren oder zuvor ermittelten Suchbegriffen, wobei ein jedes Neuron eine Aktivierung aufweist und über eine gewichtete Verbindung mit zumindest einem anderen Neuron verbunden ist, wobei das Produkt aus der Aktivierung des Neurons mit dem Verbindungsgewicht eine Endaktivierung bildet, **dadurch gekennzeichnet**, dass mehrere Suchbegriffe vorgegeben werden und zumindest zwei neuronale Netze generiert werden, wobei jeder Suchbegriff jeweils einem Neuron eines jeden Netzes zugeordnet wird, wobei bei dem ersten neuronalen Netz die Verbindungsgewichte konstant sind und die Aktivierung eines jeden Neurons in Abhängigkeit der Häufigkeit des Suchbegriffes im auszuwertenden Text erfolgt, und wobei bei dem zweiten neuronalen Netz die Verbindungsgewichte in Abhängigkeit der Häufigkeit des Suchbegriffes im auszuwertenden Text sind und die Aktivierung konstant ist, und eine Berechnung der jeweiligen Endaktivierungen beider Netze erfolgt.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass für jeden Suchbegriff der arithmetische Mittelwert aus beiden Endaktivierungen ermittelt wird.

3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, dass eine Sortierung der Suchbegriffe nach absteigender Endaktivierung durchgeführt wird.

4. Verfahren nach einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass es sich bei den Suchbegriffen um mittels einer vorhergehenden Suche unter Verwendung eines ersten Suchbegriffes ermittelten korrespondierende weitere Begriffe handelt.

5. Verfahren nach einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass nach der Auswertung von mehreren Texten eine Rangliste der Texte nach Relevanz gebildet wird, wobei die Relevanz aus der Endaktivierung eines oder beider neuronaler Netze und/oder aus einem Mittelwert der Endaktivierungen beider neuronaler Netze abgeleitet wird.

6. Verfahren nach einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass mittels einer vorgebbaren Negativliste bestimmte Begriffe von einer Suche und/oder Auswertung ausnehmbar sind.

7. Verfahren nach einem der vorherigen Ansprüche, dadurch gekennzeichnet, dass mittels einer ersten Anwendung des Verfahrens korrespondierende Suchbegriffe zur Erweiterung der Suche ermittelt werden, mittels derer unter erneuter Anwendung des Verfahrens iterativ eines oder mehrere weitere Paare von neuronalen Netzen erzeugt und deren Endaktivierungen ermittelt werden.

8. Such- und Auswertesystem zur Auswertung von elektronisch gespeicherten Texten, insbesondere zur Anwendung eines Verfahrens nach einem der vorherigen Ansprüche, aufweisend eine Benutzerschnittstelle zum Erhalt von Suchbegriffen, sowie eine Suchmaschinenschnittstelle zur Veranlassung einer Suche unter Verwendung der Suchbegriffe und zum Erhalt des Resultates der Suche und der aufgefundenen Texte, dadurch gekennzeichnet, dass das System einen Netzgenerator aufweist, mittels dessen zumindest zwei neuronale Netze generiert werden, wobei ein jedes Neuron eine Aktivierung aufweist und über eine gewichtete Verbindung mit zumindest einem anderen Neuron verbunden ist, wobei das Produkt aus der Aktivierung des Neurons mit dem Verbindungsgewicht eine Endaktivierung bildet, wobei jeder Suchbegriff jeweils einem Neuron eines jeden Netzes zugeordnet wird, und wobei eine Textauswerteeinheit vorgesehen ist, mittels derer die absoluten und/oder relativen Häufigkeiten der Suchbegriffe in einem Text ermittelt und an den Netzgenerator übergeben werden, wobei mittels des Netzgenerators bei dem ersten neuronalen Netz die Aktivierung eines jeden Neurons in Abhängigkeit der Häufigkeit des zugeordneten Suchbegriffes bei konstanten Verbindungsgewichten und bei dem zweiten neuronalen Netz die Verbindungsgewichte in Abhängigkeit der Häufigkeit des zugeordneten Suchbegriffes bei konstanter Aktivierung durchgeführt wird und die jeweiligen Endaktivierungen ermittelt werden, und aufweisend eine Ausgabeeinheit, mittels derer die Endaktivierungen der neuronalen Netze ausgegeben werden.

9. System nach Anspruch 8, dadurch gekennzeichnet, dass es eine graphische Benutzerschnittstelle aufweist, mittels derer die Endaktivierungen der beiden neuronalen Netze in Bezug auf die verwendeten Suchbegriffe graphisch dargestellt werden.

10. System nach Anspruch 8 oder 9, dadurch gekennzeichnet, es eine Benutzerschnittstelle aufweist, mittels derer eine Erläuterung in einer natürlichen Sprache über die Relevanz der Suchbegriffe in Bezug auf den ausgewerteten Text ausgegeben wird,

wobei die Erläuterung mittels der Textauswerteeinheit aus den Endaktivierungen der beiden neuronalen Netze in Bezug auf die verwendeten Suchbegriffe generiert und.

11. System nach einem der Ansprüche 8 bis 10, dadurch gekennzeichnet, dass mittels der Ausgabereinheit eine Rangliste einer Mehrzahl von ausgewerteten Texten gebildet und ausgegeben wird, wobei die Rangliste sich nach der Relevanz der Texte in Bezug auf die Suchbegriffe richtet.

12. System nach einem der Ansprüche 8 bis 11, dadurch gekennzeichnet, dass das System Datenspeicherungsmittel zur Abspeicherung und/oder Zwischenspeicherung der Suchbegriffe und/oder der aufgefundenen und/oder ausgewerteten Texte und/oder der neuronalen Netze sowie Netzparameter und/oder von Ranglisten aufweist.

13. System nach einem der Ansprüche 8 bis 12, dadurch gekennzeichnet, dass eine Schnittstelle zum Erhalt von durch eine vorhergehende Suche ermittelten zu den ersten Suchbegriffen korrespondierenden weiteren Begriffen vorgesehen ist, wobei unter Verwendung der weiteren Begriffe das erste und zweite gebildete neuronale Netz erweitert und/oder ein drittes und viertes neuronales Netz mittels dem Netzgenerator gebildet wird.

14. System nach einem der Ansprüche 8 bis 13, dadurch gekennzeichnet, dass ein Filter vorgesehen ist, mittels dessen eine vorgebbare Negativliste von Begriffen von einer Suche und/oder Auswertung ausnehmbar ist.

15. Computerprogrammprodukt umfassend ein auf einem Computer lauffähiges Computerprogramm, welches das Verfahren nach einem der Ansprüche 1 bis 7 durchführt, wenn es auf einem Computer ausgeführt wird.

Es folgt ein Blatt Zeichnungen

Anhängende Zeichnungen

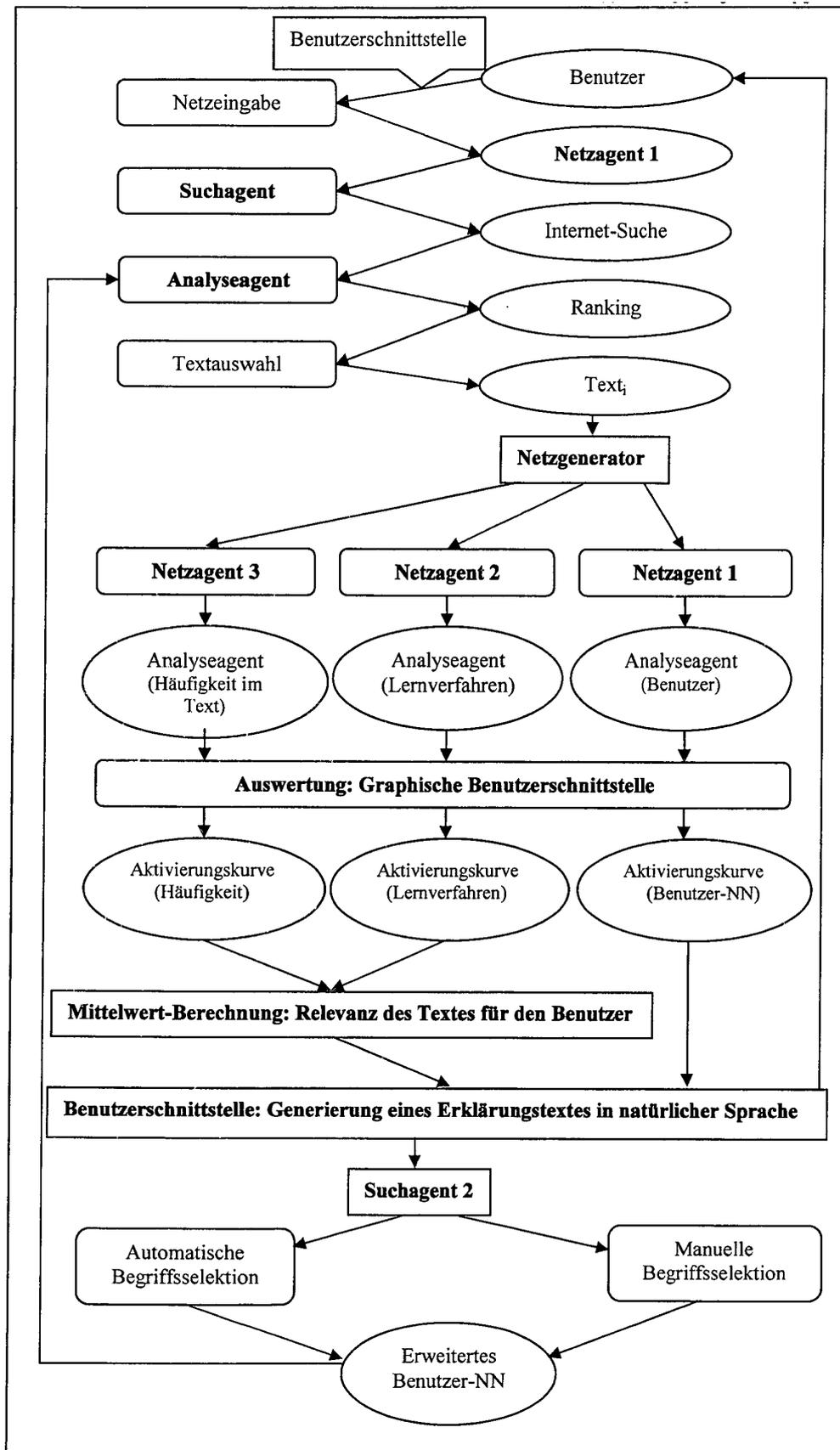


Fig. 1 Übersicht der Komponenten