



Burkhard Priemer

2.6.2004

Logfile-Analysen: Möglichkeiten und Grenzen ihrer Nutzung bei Untersuchungen zur Mensch-Maschine-Interaktion

Die Aufzeichnung der Computernutzung in automatisch generierten elektronischen Protokollen – so genannten Logfiles – entwickelt sich zu einer zunehmend genutzten Erfassungsmethode bei Untersuchungen der Interaktion zwischen Mensch und Computer. Dieser Beitrag bietet einen einführenden Überblick über Verfahren der Logfile-Aufzeichnung und der Analyse der Daten. Neben einer Erläuterung der Begrifflichkeiten werden sowohl Vor- und Nachteile von Logfile-Auswertungen als auch grundlegende Methoden, mathematische Beschreibungen des Nutzerverhaltens und Ansätze der Typisierung von Nutzeraktivitäten dargestellt.

Nutzerverhalten in beliebigen Hypertexten kann mithilfe von Logfiles gespeichert und analysiert werden. Dabei wird von passiver Protokollierung gesprochen, da vom Nutzer keinerlei Aktivitäten verlangt werden. Das Verfahren sowie die aufgezeichneten Protokolle werden häufig vereinfachend mit den Begriffen «Logfiles», «Data-» oder «Web-Mining» bzw. «User Tracking» bezeichnet. In der Literatur erfolgt bislang keine einheitliche Verwendung der Begriffe. Dies rührt her von der ebenso unterschiedlichen Nutzung dieser Protokolle in Wirtschaft und Universität, im Rahmen von Webadministration, Forschung, Überwachung und weiteren Anwendungen. Da sich im akademischen Umfeld der Begriff «Logfile» für die Aufzeichnung jeglicher Computernutzung durchzusetzen scheint, wird dieser in dem vorliegenden Beitrag ebenfalls verwendet.

Ziele und Instrumente der Logfile-Analysen

Ziel vieler Arbeiten der interdisziplinären Hypermedia-Forschung ist die Untersuchung der Navigation von Probanden in elektronischen Lern-

umgebungen. Zu diesem Zweck werden Protokolle der Computernutzung angelegt, analysiert, bewertet und gewonnene Variablen in Relation zu anderen Parametern gesetzt. Im Folgenden werden zwei Blickwinkel erläutert, die dabei von besonderem Interesse sind: «Die Lernweganalyse kann zwei Interessen verfolgen, die unterschiedliche Kriterien für ihr Vorgehen anlegen müssen: Es kann einerseits um grundsätzliche Erkenntnisse über Lernwege und ihre Determinanten gehen. Dann wird man auf eine Typisierung (oder zumindest: Parametrisierung) von Lernwegen nicht verzichten können. Eine andere Fragestellung liegt vor, wenn die Qualitätsprüfung und -sicherung eines bestimmten Mediums (im Sinne formativer Evaluation) im Vordergrund steht. Hier ist der Informationsverlust durch das Aggregieren von Lernwegen zu Typen ggfs. Unangemessen hoch» (Kerres, 1998, 225).

Bezugsobjekt der Datenaufzeichnung

Grundsätzlich können folgende Betrachtungsweisen des Aufrufverhaltens von Knoten in Hypertexten unterschieden werden (nach Rossmanith, 2001, 132):

- ein Nutzer einer Anwendung (Software, Lern-Modul, Webserver), 1:1-Betrachtung
- ein Nutzer auf verschiedenen Anwendungen, 1:n-Betrachtung
- viele Nutzer einer Anwendung, n:1-Betrachtung
- viele Nutzer auf verschiedenen Anwendungen, n:n-Betrachtung.

In vielen wissenschaftlichen Untersuchungen steht das Verhalten mehrerer Probanden in einem vorgegebenen Hypertext im Vordergrund (*n:1*-Betrachtung). Mithilfe dieser Studien sollen in der Regel Lernumgebungen evaluiert bzw. die Lernstrategien von Probanden identifiziert und bewertet werden. Dies ist nicht ganz unproblematisch, da sich beide Aspekte bedingen: «This is an example of a chicken-or-egg problem of user-centered design: We cannot discover how users can best work with systems until the systems are built, yet we should build systems based on knowledge of users and how they work» (Marchionini, 1995, 75).

Tools der Datenaufzeichnung

Die Zielsetzung der Logfile-Analyse bestimmt die Wahl des geeigneten Instrumentes. Prinzipiell lassen sich nach Leard und Hadwin (2001, 7) drei

Typen von Aufzeichnungsverfahren unterscheiden:

Customized Logging Tools: auf eine bestimmte Anwendung zugeschnittene Client-seitige Software, welche die Nutzung nur dieser Anwendung aufzeichnet und auswertet. (Ein einfaches Beispiel ist die Verlaufsaufzeichnung bzw. Historyliste von Standardbrowsern.)

Web Server Logging Tools: Web Server zeichnen Server-seitig die Nutzung von Internetseiten auf. Gespeichert werden in der Regel die IP-Adresse, von der die Anforderung kam, wann der Aufruf erfolgte, welche Datei aufgerufen wurde, ob der Aufruf erfolgreich war, von welcher Webseite aus der Aufruf gestartet wurde und welchen Browser der Benutzer verwendet.

Keylogging Tools: eine Client-seitige von speziellen Anwendungen unabhängige zusätzliche Software, die alle Aktionen eines Nutzers aufzeichnet.

Inzwischen gibt es eine Vielzahl an Produkten, die Logfiles aufzeichnen und analysieren. Diese Entwicklung wird langfristig zumindest z. T. das bislang übliche Verfahren ablösen, dass je nach Forschungsfrage spezielle Tools für die Datensammlung und Analyse entwickelt und programmiert werden müssen. Allerdings haben sich einheitliche Standards noch nicht durchgesetzt, wenngleich dies verschiedentlich gefordert wurde (siehe z. B. Hadwin & Leard, 2001). Einige Programme bieten bereits die Möglichkeit, sowohl die Erfassung als auch die Auswertung von Logfiles weitgehend den Untersuchungszielen der Forscher anzupassen (siehe z. B. Noller, Naumann & Richter, 2001; Fuller, Simonson, Tiwari & Rebelsky, 2002). Neben reinen Häufigkeitsauswertungen und Zeitmessungen (Anzahl besuchter Seiten, Zeit auf Knoten usw.) und der Berechnung vorgegebener Parameter (z. B. zur mathematischen Beschreibung der Struktur von Pfaden) bieten verschiedene Programme weitere Analysehilfsmittel, die auch die Überprüfung der Logfiles auf bestimmte Muster oder Regelmässigkeiten zulassen. Die ausgewerteten Daten der Logfiles können in der Regel graphisch und durch Text, im «Replay» (Wiedergabe des Ablaufs der besuchten Knoten und der Aktivitäten der Nutzer) und als Rohdaten dargestellt werden. Weiterhin lassen sich Daten ebenfalls in Files konvertieren, die von «üblicher» Standardsoftware wie Tabellenkalkulations- und Textverarbeitungsprogrammen oder von Webbrowsern gelesen werden können.

Vor- und Nachteile von Logfile-Analysen

Bevor auf Methoden und Verfahren der Logfile-Analyse näher eingegangen wird, sollen schlaglichtartig Vor- und Nachteile aufgezeigt werden. Aus dieser Darstellung wird ersichtlich, welche Möglichkeiten aber auch Einschränkungen beachtet werden müssen. Letztere gelten allerdings z. T. ebenfalls für andere «nicht-elektronische» Verfahren der Erfassung von Nutzerverhalten, beispielsweise für die Verwendung von Fragebögen.

Vorteile von Logfile-Analysen

- Die Aufzeichnung der Computernutzung erfolgt unbemerkt, unsichtbar und fast ohne Beeinflussung des Nutzers (vgl. z. B. Kleining, 1999). (Eventuell kann die Kenntnis über die Aufzeichnung der Computernutzung das Verhalten der Probanden beeinflussen.) Dies ist ein wesentlicher Unterschied zu vielen traditionellen Erhebungsmethoden.
- Die Erfassung ist objektiv, eindeutig, vollständig und detailgenau. Aufgezeichnet werden können z. B. besuchte Seiten, Zeit auf Seiten, genutzte Links, eingegebene und gelöschte Zeichen, kopierte Texte und Abbildungen, Mausbewegungen usw.
- Das Nutzerverhalten kann ohne zusätzlichen Zeitaufwand, weder für den Forscher noch für den Probanden, aufgezeichnet werden.
- Logfiles können durch das Lernmedium selbst «live» erfasst werden. Zusätzliche Werkzeuge oder Instrumente sind nicht erforderlich.
- Die Aktivitäten des Nutzers können unter authentischen Verhältnissen und Bedingungen erfasst werden. Dies ist insbesondere für Felduntersuchungen von Bedeutung.
- Mit Logfile-Aufzeichnungen können Daten einer grossen Anzahl von Probanden mit relativ geringem Aufwand erhoben und ausgewertet werden. Die Erfassung der Daten erfolgt automatisch, und die Auswertungsprogramme können schnell eine grosse Anzahl an Daten verarbeiten (siehe z. B. Degenhardt, 2001).
- In der Regel ist die elektronische Aufzeichnung technisch leicht realisierbar und zuverlässig.

Nachteile von Logfile-Analysen

- Logfiles sind in der Regel sehr umfangreich und enthalten in den meisten Fällen für einzelne Untersuchungsfragen überflüssige Daten. Aus diesem Grund müssen Logfiles ggf. je nach Fragestellung vor einer Analyse durch zusätzliche spezielle Programme bereinigt werden.

- Parallel ablaufende Prozesse (z. B. das gleichzeitige Arbeiten mit mehreren Fenstern) werden in Logfiles in der Regel streng nach der zeitlichen Reihenfolge der einzelnen Aktivitäten eindimensional hintereinander festgehalten. Allein aus der Betrachtung der Logfiles ist es u. U. im Nachhinein schwierig, solche parallelen Prozesse zu identifizieren (siehe Hadwin & Leard, 2001).
- Verschiedene Aktionen der Nutzer (z. B. Nachdenken, handschriftliche Notizen, Warten, Verlassen des Arbeitsplatzes, Kommentare, Suchen und Lesen auf einer Seite) können nicht aufgezeichnet werden (Körber, 2000, 15). Es ist deshalb nicht möglich, die komplette Mensch-Maschine-Interaktion allein durch Reaktionen des Systems zu beschreiben.
- Logfiles unterscheiden in der Regel nicht zwischen vom Nutzer oder vom System erzeugten Aktivitäten. Automatismen, die nicht aktiv vom Nutzer ausgelöst werden, sondern vom Programm erzeugt werden (z. B. Pop-Up Windows), werden genauso aufgezeichnet wie Nutzer-gesteuerte Aktionen (Hadwin & Leard, 2001).
- Werden Zeiten auf verschiedenen Knoten eines Hypertextes zur Analyse herangezogen, so bleibt fraglich, ob der Nutzer geistig oder physisch anwesend ist, ob die Granularität (Umfang und Grösse) der betrachteten Knoten vergleichbar ist oder ob nicht weitere Faktoren (z. B. Lesegeschwindigkeit) die Aktivitäten des Nutzers beeinflussen (Zajonc & Priemer, 2001).
- Ein Rückschluss von den Logfile-Daten auf zu Grunde liegende kognitive Prozesse ist schwierig (siehe z. B. Unz, 2000; Priemer, 2004). Die Motivation der Nutzer für bestimmte Aktivitäten, beispielsweise die Wahl bestimmter Knoten, kann mit Logfiles nicht erfasst werden. Einzelne Aktionen von Nutzern, die mit Logfiles aufgezeichnet werden, lassen sich darüber hinaus oftmals nur im Zusammenhang mit anderen Aktionen deuten und ergeben für sich genommen nicht unbedingt Sinn (Hadwin & Leard, 2001).

Mit Kenntnis der Schwierigkeiten lassen sich einige der aufgezeigten Probleme z. T. durch geeignete Massnahmen (wie z. B. ergänzende Erfassungen oder Stichprobenuntersuchungen wie der Methode des «lauten Denkens», schriftliche Protokollierungen bzw. Fragebögen, Interviews oder Video-Aufzeichnungen) einschränken. Aus diesem Grund sollte der Einsatz von Logfiles als Datenerhebungsmethode nicht grundsätzlich als Ersatz «herkömmlicher» Verfahren, sondern als deren Ergänzung bzw.

Erweiterung gesehen werden.

Verfahren von Logfile-Analysen

Die Auswertung der Logfiles erfolgt in der Regel in fünf Schritten (siehe Abb. 1 aus Wilde, Hippner & Merzenich, 2002; vgl. auch Degenhardt, 2001):

- *Datenauswahl*: Heranziehen der relevanten Logfiles
- *Datenaufbereitung*: Bereinigen der Daten hinsichtlich der betrachteten Fragestellung und Identifizierung von Nutzern und Sitzungen
- *Datenintegration*: Einbinden zusätzlicher Informationen oder Daten über die Nutzer
- *Mustersuche*: Überprüfung der Daten auf Muster und Regelmässigkeiten mit verschiedenen mathematischen Verfahren
- *Interpretation*: Bewertung und Interpretation der Ergebnisse

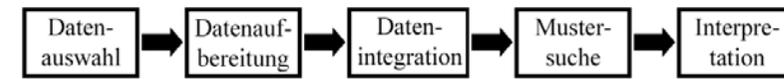


Abb. 1. Ablauf der Logfile-Analyse (aus Wilde, Hippner & Merzenich, 2002).

Grundsätzlich können drei prinzipielle Verfahren der Mustersuche unterschieden werden (vgl. Noller, Naumann & Richter, 2001; Hadwin & Leard, 2001; Niegemann, 2001), die üblicherweise in der angegebenen Reihenfolge durchlaufen werden: Auswertungen von Einzelereignissen (Nutzung einzelner Seiten, Zeit auf einzelnen Knoten), Auswertungen von Mustern (Folgen von Seiten; Linearität von Seitenfolgen) und Auswertungen der vollständigen Navigation (Beschreibung gesamter Nutzungsprozesse). Zur Beschreibung werden verschiedene mathematische Verfahren angewendet, die im Folgenden kurz skizziert sind. Dabei geht es in diesem Beitrag darum, zwei häufig genutzte Analysetechniken vorzustellen und prinzipielle Ansätze zu skizzieren, und nicht darum einen vollständigen Überblick zu bieten. Schwerpunkt der Darstellung sind dabei graphentheoretische Verfahren, da sich diese in verschiedenen Untersuchungen bewährt haben.

Häufigkeits- und Zeiterfassungen

Erfasst und gezählt werden z. B. Häufigkeiten von Knotenbesuchen im

Hypertext (Anzahl aller Seiteneindrücke «PageImpressions», Anzahl aller unterschiedlichen Seiteneindrücke), vom Nutzer verbrachte Zeit im Hypertext (Zeit pro Knoten, Gesamtbesuchsdauer) sowie Elaborationstiefe (maximale Eindringtiefe in hierarchisch strukturierten Hypertexten) (vgl. Barab, Bowdish & Lawless, 1997). Reine Häufigkeitsstatistiken werden hauptsächlich zur Evaluation von Websites genutzt ($n:1$ -Betrachtungen). Zur Interpretation von Nutzerverhalten sind diese statistischen Daten in der Regel nicht ausreichend, da kaum Zusammenhänge zu weiteren relevanten Daten zu erkennen sind (siehe z. B. Astleitner, 1997).

Graphentheorie

Ein Graph ist definiert als ein Tupel bestehend aus einer endlichen nicht leeren Menge von Knoten und einer Menge von Kanten (zwei-elementige Teilmenge der Menge der Knoten). Prinzipiell kann ein Navigationspfad durch einen Hypertext als gerichteter Graph aufgefasst werden. Die Knoten des Graphen entsprechen den Seiten im Hypertext, die Kanten den Links zwischen den Seiten. Neben der Darstellung als Graph lässt sich ein Pfad auch als Matrix (Adjazenzmatrix) mathematisch beschreiben. Ist n die Anzahl der nummerierten Knoten K_1 bis K_n des Hypertextes, dann ist die den Pfad beschreibende ($n \times n$)-Matrix A folgendermassen definiert: Das Element a_{ij} ist gleich der Anzahl der Kanten, die K_i mit K_j verbinden.

Es sei beispielsweise ein Navigationspfad gegeben, der aus folgenden Knotenbesuchen in der angegebene Reihenfolge besteht (vgl. Abb. 2): $K_1, K_2, K_3, K_1, K_4, K_2, K_1, K_4$.

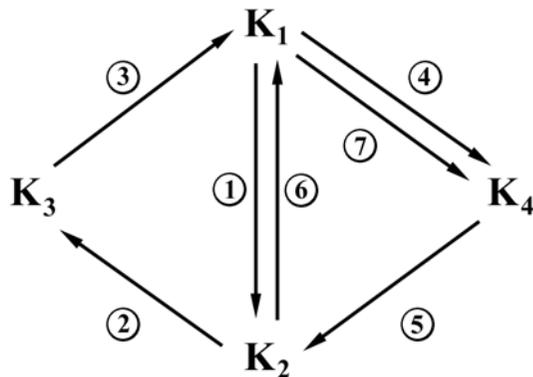


Abb. 2. Beispiel eines Navigationspfads bestehend aus einer Folge besuchter Knoten.

Dann ist A eine 4×4 Matrix der Gestalt

$$(a_{ij}) = A = \begin{pmatrix} 0 & 1 & 0 & 2 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Auf diese Weise lässt sich nicht nur der Pfad eines Nutzers beschreiben, sondern auch der gesamte Hypertext (siehe Botafogo, Rivlin & Shneiderman, 1992).

Die Graphentheorie liefert eine Reihe von Massen zur Beschreibung der Navigation bzw. des Hypertextes. Diese werden im Folgenden der Vollständigkeit halber benannt, aber nicht näher ausgeführt: Anzahl der Knoten und Kanten, Distanz von Knoten, Dichte (siehe z. B. Unz, 2000, 91), Kompaktheit, Stratum (Botafogo, Rivlin & Shneiderman, 1992), Eindringtiefe, Varianz der Elaboration (Niegemann, 2001), Zentralität von Knoten (Modjeska & Marsh, 1997; zu so genannten «Landmarks» siehe Valdez, Chignell & Glenn, 1988), Knotengewicht, Glaubensgrad, Grösse, Informationsgehalt, Verständlichkeit (Astleitner, 1997, 104) usw.

Winne, Gupta und Nesbit (1994) versuchen beispielsweise, mit graphentheoretischen Massen (genannt «Bridge», «Cut Point») Teilgraphen in durchlaufenen Netzen zu finden, um Phasen der Navigation voneinander trennen zu können. Weiterhin nutzen sie Distanzmasse zur Charakterisierung der Ähnlichkeit von Knoten und «Multiplicity» zum Vergleich von Knotenfolgen. Der Ansatz, damit kognitive Prozesse abbilden und beschreiben zu können, erscheint m. E. aber fragwürdig. Denn zum einen ist die Analyse der Pfade stark auf den speziellen verwendeten Hypertext zugeschnitten (Muster könnten Artefakte darstellen). Zum anderen ist die Identifizierung der kognitiven Prozesse durch graphentheoretische Masse wenig eindeutig.

Neben der Struktur des Graphen und der Knoten lassen sich auch die Kanten bzw. die Links näher beschreiben und klassifizieren (siehe Astleitner, 1997, 105).

Von besonderem Interesse sind sechs einfach zu identifizierende und interpretierende Indizes zur Beschreibung von Mustern in Navigations-

pfaden, die von Canter, Rivers und Storrs (1985, 96) vorgeschlagen wurden (Abb. 3):

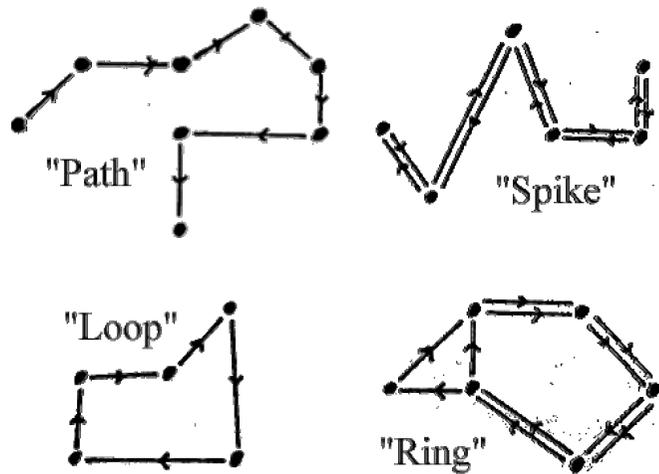


Abb. 3. Navigationsmuster nach Canter, Rivers und Storrs (1985).

Anzahl an Pfaden (Path): Ein Pfad ist eine Folge (Kette) von Kanten, in der kein Knoten zweimal auftritt (PQ : «Path Quantity»).

Anzahl an Ringen (Ring): Ein Ring ist eine Folge von Kanten, die zu einem Startknoten zurückführt und in sich selbst weitere Ringe enthalten kann (RQ : «Ring Quantity»).

Anzahl an Schleifen (Loop): Eine Schleife ist ein Ring, der keine Teilringe enthält (LQ : «Loop Quantity»).

Anzahl an Stacheln oder Ähren (Spikes): Ein Stachel liegt vor, wenn eine Folge von Kanten ab einem bestimmten Knoten in entgegengesetzter Reihenfolge zurückverfolgt wird (SQ : «Spike Quantity»).

Explorationsgrad: Quotient aus unterschiedlichen besuchten Knoten (NV : «Nodes visited») und Gesamtzahl aller im Hypertext vorhandener Knoten (NT : «Nodes total»).

Redundanzgrad: Quotient aus unterschiedlichen besuchten Knoten (NV) und Gesamtzahl aller aufgerufenen Knoten mit Wiederholungen (NS : «Nodes seen»).

Diese Basisparameter haben sich aufgrund ihrer Einfachheit in einer Reihe von Untersuchungen bewährt. Deren Vorteil ist, dass sie sich leicht aus den Logfiles gewinnen lassen und keine besonderen Voraussetzungen an die Experimentalsituation (z. B. den verwendeten Hypertext) stellen. Weiterhin gibt es Validierungen dieser Masse.

Wenzel (2001, 44) schlägt darüber hinaus vor, die ersten vier dieser Masszahlen durch ihr jeweiliges Maximum zu teilen, damit interindividuelle Vergleiche einfacher möglich werden. Weiterhin definiert er zusätzliche Parameter zur Beschreibung von Effizienz und Effektivität (Wenzel, 2001, 180). Dazu werden zunächst die für eine Aufgabenstellung relevanten Knoten festgelegt (RN : «relevant Nodes»). Diese Einstufung der Relevanz von Knoten unterliegt allerdings subjektiven Einflüssen des Beurteilers. Relevante Knoten werden auch als «Authorities» und Seiten, die auf diese «Authorities» verweisen, selbst diese Informationen aber nicht enthalten, als «Hubs» bezeichnet. «Good hubs point to many good authorities, and good authorities are pointed at by many good hubs» (Lempel & Moran, 2000). Wird ein relevanter Knoten mindestens 7 Sekunden besucht – Wenzel diskutiert diese gesetzte Schranke kritisch –, so zählt dieser Besuch als Informationsaufnahme (I). Bei Wenzel (2001) werden mehrfache Informationsaufnahmen eines Knotens nur einmal gezählt. Dann gilt:

Effektivität: Quotient aus Informationsaufnahmen (I) und Anzahl besuchter relevanter Knoten (RN)

Effizienz: Quotient aus Informationsaufnahmen (I) und Gesamtzahl aufgerufener verschiedener Knoten (NV)

Es gibt darüber hinaus eine Reihe weiterer Möglichkeiten, mit graphentheoretischen Methoden Muster in Pfaden z. B. nach bildlichen Kriterien festzulegen. Zum Beispiel unterscheidet Horney (1993) fünf Muster: lineares Durchwandern, Seiten-Trips, Sterne, erweiterte Sterne und chaotisches Navigieren. Ähnlich definieren Tauscher und Greenberg (1997) Nabe-Speichen-Muster: Von einem zentralen Knoten (in einem kreisförmig organisierten Bild die Nabe) gehen einzelne Pfade ab, die den nacheinander verfolgten Optionen entsprechen (radiale Speichen). Nach eher hierarchischen Kriterien klassifizieren Stanton und Baber (1992) Navigation durch die Attribute «top-down» (hierarchisch höher liegende Seiten werden zuerst besucht), «sequenziell» (Seiten werden nach einer

vorgeschlagenen Reihenfolge besucht) und «elaborativ» (Zick-Zack-Bewegungen zwischen Seiten auf verschiedenen hierarchischen Stufen). Insgesamt bieten Verfahren der Graphentheorie weitreichende Möglichkeiten, Hypertexte und Nutzerverhalten quantitativ zu erfassen und zu beschreiben. Sie stellen damit einen erfolgversprechenden Zugang zur Analyse von Logfiles dar (Niegemann, 2001; Leard & Hadwin, 2001), da mit diesen Methoden die geringsten Einschränkungen verbunden sind, wenn Wege durch einen Hypertext quantifiziert werden sollen. Graphentheoretische Verfahren sind deshalb Grundlage vieler Logfile-Analysen.

Wahrscheinlichkeitstheorie

Anhand verschiedener Modelle der Wahrscheinlichkeitstheorie werden Linknutzungen bzw. Übergänge zwischen Knoten mit Wahrscheinlichkeiten beschrieben. Beispielsweise kann untersucht werden, inwiefern bestimmte Knoten eines Hypertextes von zentraler Bedeutung sind, wenn sie besonders oft besucht werden. Mithilfe der Wahrscheinlichkeitstheorie lässt sich bestimmen, ob es sich dabei um zufällige oder systematische (durch die Nutzer gesteuerte) Effekte handelt.

Prinzipiell steigt die Wahrscheinlichkeit des Aufrufs eines bestimmten Knotens K_j mit der Anzahl an Links, die auf ihn verweisen. Diese Wahrscheinlichkeit ist aber wiederum abhängig davon, wie gross die Wahrscheinlichkeit ist, einen Knoten K_2 aufzurufen, der auf den ursprünglichen Knoten K_1 verweist. Insgesamt nimmt also die Wahrscheinlichkeit des Aufrufs von Knoten K_j ab, je weiter dieser vom Ausgangsknoten des Nutzers entfernt ist.

Die Wahrscheinlichkeit P_j für einen Klickpfad j der Länge g ergibt sich aus den Wahrscheinlichkeiten p_f der g Übergänge:

$$P_j = \prod_f p_f$$

($f=1, \dots, g$). Der Erwartungswert $E(K_i)$ für die Häufigkeit eines Aufrufs des Knotens K_i lässt sich aus der Häufigkeit $H(K_i)_j$ des Auftretens eines Knotens in einem Klickpfad j und der Wahrscheinlichkeit P_j dieses Klickpfads bestimmen:

$$E(K_i) = \sum_j H(K_i)_j \cdot P_j$$

(Es sei $i=1, \dots, n$ die Anzahl der Knoten und $j=1, \dots, m$ die Anzahl der Pfade

einer bestimmten Länge.)

Aus dem Vergleich zwischen erwarteten und tatsächlichen Knotenaufrufen kann geschlossen werden, ob bestimmte Knoten «überzufällig» häufig besucht werden.

Derartigen Ansätze sind nicht trivial, da bedingte Wahrscheinlichkeiten betrachtet werden müssen (Niegemann, 2001; Wenzel, 2001, 37). Hinzu kommt, dass Nutzer unterschiedlich lange Pfade aufweisen und u. U. so wenig Schritte in einem Hypertext absolvieren, dass für einige Knoten die Aufrufwahrscheinlichkeit praktisch null wird. Insgesamt ist eine solche mathematische Beschreibung nicht ohne grösseren Aufwand zu verwirklichen, insbesondere dann, wenn umfangreichere Navigationspfade betrachtet werden.

Ein weiteres Verfahren der «probabilistischen Linkvorhersage» und Pfadanalyse nutzt Markow-Ketten (Sarukkai, 2000; Levene & Loizou, 2001, Lempel & Moran, 2000; Qiu, 1993). Dabei werden Knoten als Zustände und Links als Übergänge interpretiert. Die fundamentale Eigenschaft dieses Modells ist die Abhängigkeit vom jeweils vorausgegangenem Zustand. Befindet sich ein Nutzer beispielsweise auf einem Knoten, so kann dieser Zustand durch einen Vektor s beschrieben werden. Der Wechsel zu einem neuen Knoten wird dann durch eine Matrix A ausgedrückt, die den Ausgangszustand s in den neuen Zustand s' überführt:

$$s' = s A.$$

Gibt es n Knoten, so ist s ein n -dimensionaler Vektor und A eine $(n \times n)$ -Matrix. Die Einträge der Matrix können z. B. mithilfe der Maximum-Likelihood Methode bestimmt werden. Dabei wird berücksichtigt, wie oft ein Nutzer von einem Knoten aus (Zustand 1) einen Link zu einem anderen Knoten (Zustand 2) gewählt hat.

Auf diese Weise ist es möglich, zu «Vorhersagen» von Webserverbesuchen zu gelangen, adaptive Systeme zu konstruieren, «Touren» durch Hypertexte zu generieren und nutzerspezifische Sammlungen aus Authorities und Hubs zu konstruieren. Problematisch bleiben aber weiterhin die für dieses Verfahren notwendigen grossen Datenmengen und der erhebliche Rechenaufwand für umfangreichere Hypertexte.

Identifizierung von Strategien mithilfe von Logfile-Analysen

In der Hypermedia-Forschung gibt es eine Vielzahl von Ansätzen, Navigation von Nutzern und deren durchlaufene Muster und angewendete

Strategien zu beschreiben und zu kategorisieren. Dies liegt nicht zuletzt an der Tatsache, dass die Begriffe Navigation und Strategie zum einen nicht einheitlich verwendet werden und zum anderen sehr viele unterschiedliche Klassifikationsverfahren denkbar sind. In der Regel liegen diesen Verfahren qualitative Merkmale zu Grunde, wie z. B. zielorientiertes versus unsystematisches Vorgehen, ohne dass eine Methode angegeben wird, wie derartige Typen quantitativ erkannt werden können. Deshalb bleibt ein grosser Interpretationsspielraum hinsichtlich der Genauigkeit dieser Typeneinteilungen. Des Weiteren wird die Navigation erheblich davon beeinflusst, mit welcher Zielsetzung gearbeitet wird: handelt es sich beispielsweise um reine Informationssuche oder um Lernen mit Hypertexten. Typeneinteilungen sind immer in Verbindung mit Aufgabenstellungen zu sehen.

Strategie soll in diesem Beitrag gemäss Ballstaedt, Mandl, Schnotz und Tergan (1981, 250) als «zielgerichtete Aktivitäten des Individuums» verstanden werden, «mit der Absicht, das Verstehen, Behalten und Erinnern zu verbessern.» Dabei wird mit Astleitner (1997, 97) davon ausgegangen, dass Strategien selbst trotz ihrer «zusätzlichen» Belastung des Arbeitsgedächtnisses nicht lernhinderlich sind.

Im Folgenden werden verschiedene Klassifikationen kurz vorgestellt. Es wird dabei nicht angestrebt, einen vollständigen Überblick zu bieten (siehe hierzu Astleitner, 1997; Unz, 2000). Vielmehr soll aus der Darstellung ersichtlich werden, in welchen unterschiedlichen Dimensionen Nutzeraktivitäten prinzipiell beschrieben werden können. Zu Gunsten einer besseren Übersicht wird grob zwischen der prinzipiellen Struktur, verschiedenen Phasen, Verhaltensmustern und Typen der Navigation unterschieden, obwohl die Grenzen zwischen diesen Aspekten fließend sind.

Struktur der Navigation

Um Muster und Regelmässigkeiten in Navigationsprozessen erkennen zu können, muss zunächst festgelegt werden, wie Navigationspfade analysiert werden können und aus welchen Teilprozessen diese bestehen.

Bates (1990, zitiert nach Unz, 2000, 87) beschreibt Suchaktivitäten durch die elementaren Basiseinheiten *Moves* (z. B. die Nutzung eines Links), *Taktiken* bestehend aus einer Reihe von *Moves* (Formulierung einer Suchanfrage), *Stratagem* als komplexe Menge an Aktionen, die Taktiken nach sich ziehen (Suche nach Zitaten), und *Strategien* als Gesamtplan (Informationssuche). Bei Marchionini (1995, 72) wird zusätzlich der

Aspekt berücksichtigt, inwieweit Nutzer Navigationsentscheidungen bewusst treffen. Ausgehend von *Information Seeking Patterns* (Muster, die Nutzer oftmals unbewusst ausführen, z. T. als Zusammenfassung von Strategien oder Taktiken) wird die Navigationsanalyse verfeinert in *Strategies* (von Nutzern bewusst ausgewählte Ansätze, Probleme oder Aufgaben zu lösen), *Tactics* (Routinen oder Entscheidungen innerhalb von Strategien) und *Moves* (einzelne Schritte).

Beiden Ansätzen liegt eine hierarchische Anordnung zu Grunde: *Patterns* bestehen aus *Strategies*, *Strategies* aus *Tactics*, *Tactics* aus *Moves*. Schwierigkeiten ergeben sich jedoch, in Folgen von *Moves* diese verschiedenen Klassen zu erkennen, gegeneinander abzugrenzen und zu unterscheiden.

Phasen der Navigation

Statt Nutzerverhalten in unterschiedlich detaillierte Teilprozesse zu untergliedern, kann der gesamte Navigationsvorgang auch mit Blick auf aufeinander folgende Teilphasen analysiert werden. Guthrie (1988) benennt beispielsweise fünf Komponenten des Informationszugriffs: Zielbildung, Knotenwahl, Informationsextraktion, Integration und Recycling. Ganz ähnlich unterscheidet Marchionini (1995) in «Accept Information Problem», «Define and Understand the Problem», «Choose a Search System», «Formulate a Query», «Execute Search», «Examine Results» und «Extract Information». Darüber hinaus nennen Shneiderman, Byrd und Croft (1997) speziell für die Suche: Formulierung der Suchanfrage, Start der Suche, Beurteilung der Ergebnisse und Verfeinerung und Modifikation der Suche. Körber (2000, 11) ergänzt dazu eine Phase des Browsens auf Seiten, die nicht zur Suchmaschine gehören.

Das Erkennen der entsprechenden Phasen sowie deren gegenseitige Abgrenzung untereinander anhand der Logfile-Daten stellt eine wesentliche Schwierigkeit dar. All diese Einteilungen spiegeln weiterhin idealisiert beschriebene Phasen wider. Der Suche nach Informationen kann jedoch nicht immer solch ein chronologischer Verlauf derartiger Phasen und Muster unterstellt werden. Vielmehr laufen diese Prozesse in der Regel nicht-linear, komplex miteinander verwoben und z. T. parallel nebeneinander ab.

Muster des Knotenaufrufs

Eng an die verschiedenen Phasen des Arbeitens und Lernens in Informationsnetzen sind Muster von Knotenaufrufen gekoppelt, die im Folgenden

beschrieben werden. Diese Muster präzisieren das Nutzerverhalten durch eine genauere Klassifikation von Folgen von *Moves* und kommen damit den Anforderungen einer quantitativen Logfile-Analyse näher.

Choo, Detlor und Turnbull (2000) nennen sechs Verhaltensmuster: *Starting* beschreibt den Beginn der Suche, z. B. das Aufrufen von Startseiten wie Suchmaschinen oder Portale. Die Navigation rund um eine relevante Quelle wird als *Chaining* bezeichnet, z. B. das Folgen von Links von Startseiten aus. Will sich ein Nutzer einen Überblick über gefundene relevante Quellen verschaffen, so spricht man vom *Browsing*, z. B. werden dabei Inhaltsverzeichnisse usw. überflogen. *Differentiating* umfasst darauf aufbauend die Auswahl der überflogenen Dokumente für eine nähere Betrachtung, es werden z. B. Seiten gespeichert («Bookmarks» bzw. «Favoriten») oder Ausdrucke erstellt. Will sich ein Anwender auf dem Laufenden halten, indem er z. B. aktuelle Quellen wiederholt besucht oder an Newsgroups teilnimmt, so wird dies als *Monitoring* bezeichnet. *Extracting* umschreibt schliesslich das systematische Bearbeiten eines Dokuments, z. B. die Informationsentnahme aus bestimmten Quellen. Ähnlich dazu unterscheidet Marchionini (1995) in Scanning, Observing, Navigating und Monitoring und Waterworth (1992) in Starten, Verketteten, Differenzieren, Beobachten, Extrahieren und Evaluieren.

Reed, Oughton, Ayersman, Ervin und Giessler (2000) beschreiben einen Ansatz, mit dessen Hilfe die Linearität der Navigation gemessen werden kann. Allerdings sind die zu Grunde liegenden Definitionen leider sehr eng an den verwendeten (hierarchisch organisierten) Hypertext geknüpft, sodass eine Verallgemeinerung – wie sie für solche Untersuchungen wünschenswert wäre – nur schwer möglich ist. Insbesondere für die Beschreibung der Navigation in stark nicht-linearen Hypertexten (wie z. B. dem Internet) ist der vorgeschlagene Weg wenig brauchbar.

Im Gegensatz zu den angeführten Phasen der Navigation sind die Muster der Knotenaufrufe weniger allgemein, z. T. mathematisch erfassbar, bieten eine detaillierte Beschreibung und lassen damit weniger Interpretationsspielraum. Dennoch können auch diese Muster u. U. schwierig gegeneinander abzugrenzen sein.

Typen der Navigation

Ziel vieler Logfile-Analysen ist, Typen von Nutzern und Strategien zu unterscheiden, diese zu beschreiben und in Zusammenhang zu weiteren Variablen wie z. B. Lernstilen oder Lernerfolgen zu stellen.

Typisierungen versuchen, eine möglichst eindeutige und reproduzierbare Zuordnung von Eigenschaften von Personen zu Typen zu finden. Hierfür werden verschiedene theoretische und empirische Verfahren verwendet. Ferner sollten die gefundenen Typen leicht zu interpretieren sein. Bei Typisierungen handelt es sich aber in der Regel immer um Extrembeispiele, welche die wenigsten Probanden vollständig erfüllen. Vielmehr zeigen Nutzer in den meisten Fällen verschieden starke Ausprägungen der Merkmale der erkannten Typen. Entsprechend sind auch die im Folgenden beschriebenen Typen einzuschätzen.

Auf empirischen Weg haben Lawless und Kulikowich (1998) mithilfe einer Clusteranalyse drei Typen generiert: *Knowledge Seekers*, die basierend auf dem Inhalt Quellen wählen, die ihr Verständnis des behandelten Stoffes fördern, *Feature Explorers*, die besonders viel Zeit mit multimedialen Elementen wie Videos, Applets usw. verbringen, und *Apathetic Hypertext Users*, die nicht zu einer der beiden Gruppen gehören und oberflächlich und scheinbar ohne logische Struktur vorgehen. Diese Typisierung ist in ihrer ersten Form bereits vor zehn Jahren aufgestellt worden und spiegelt zu einem gewissen Grad Zeiten wider, als Arbeiten mit Hypertexten und die Anwendung von «Features» noch neu waren. Es wäre interessant zu untersuchen, ob dieser «Neuigkeitseffekt» immer noch so stark ausgeprägt ist und die Typen weiterhin so deutlich auftreten. Leider lässt sich der Ansatz der beiden Autoren auch nicht verallgemeinern, da die in die Clusteranalyse eingehenden Variablen auf den speziell verwendeten Hypertext zugeschnitten sind. Dieses Problem sollte dadurch umgangen werden, dass verallgemeinerbare und auf jeden Hypertext anwendbare Variablen für eine Clusteranalyse verwendet werden. Dies können z. B. die Indizes von Canter, Rivers und Storrs (1985, 96) sein.

Marchionini (1995, 76) unterscheidet folgende «Analytical Search Strategies» mit Verweis auf verschiedene Quellen: *Building Blocks* (Harter, 1986) ist das Durchführen einer Sucheingabe und darauf basierend das Sammeln von relevanten Einzeldokumenten, den so genannten Blocks. *Successive Fractions* (Meadow & Cochrane, 1981) ist das sukzessive Einschränken der Datenmenge «von oben» (viele Informationen) «nach unten» (wenige ausgewählte Dokumente). Navarro-Prieto, Scaife und Rogers (1999) bezeichnen dies auch als *Top-down* Vorgehen. *Pearl Growing* (Markey & Cochrane, 1981) umschreibt den Beginn der Navigation mit einem relevanten Dokument (Pearl) und der Suche nach ähnlich guten Quellen. *Interactive Scanning* (Hawkins & Wagers, 1982)

beginnt mit einer groben Suche und dem Scannen der erhaltenen Dokumente. Damit erfolgt dann schrittweise ein Verfeinern der Suche durch Wiederholen dieses Prozesses. Ergänzend sei an dieser Stelle das *Bottom Up* Vorgehen (Navarro-Prieto, Scaife & Rogers, 1999) erwähnt, das «in umgekehrter Richtung» zum *Top Down* Vorgehen von Detaildokumenten ausgehend zu einem Gesamtüberblick führt.

Diese Aufzählung deutet an, dass es eine grosse Anzahl unterschiedlicher Typeneinteilungen gibt. Dies folgt nicht zuletzt auch daraus, dass angewendete Strategien von einer Vielzahl von Faktoren abhängen und darüber hinaus während der Bearbeitung einer Aufgabenstellung wechseln können. Astleitner (1997, 97) bemängelt, dass zur Personenspezifität und Situationsangemessenheit bei den Strategien wenig differenziert wird: «Obwohl Handlungsempfehlungen gegeben werden, fehlen theoretische Grundlagen und empirische Überprüfungen des Vorkommens oder der Lernwirksamkeit der vorgeschlagenen Strategien. [...] Weitgehend ungeklärt ist, welcher Informationsnetzbenuer wann welche Strategie einsetzen soll, um optimales Lernen gewährleisten zu können.» Um derartige Empfehlungen geben zu können, müssen Untersuchungen sowohl die Eigenschaften der Personen als auch die (Lern-) Situation berücksichtigen und benennen.

Navigationsstrategien nach Canter, Rivers und Storrs

Canter, Rivers und Storrs (1985) unterscheiden in *Browsing*, *Searching*, *Wandering*, *Scanning* und *Exploring*. Diese Typen werden im Folgenden detailliert erläutert, da sie grundlegend sind und sich auf beliebige Hypertexte übertragen lassen. Canter et al. geben in ihrer Typisierung der Strategien im Gegensatz zu den meisten anderen Autoren ein Verfahren an, wie die beschriebenen Strategien quantitativ erfasst werden können. Allerdings räumen Canter et al. selbst Schwächen in ihrer vorgeschlagenen Quantifizierung ein. Ein Umstand, der Anlass gab, ein Verfahren zu entwickeln, um diese Typen empirisch zu quantifizieren (siehe Priemer, 2004). Dabei wird auf die Indizes zurückgegriffen, die im Abschnitt 3 über graphentheoretische Verfahren definiert werden. Canter et al. (1985, 100) definieren folgende Typen (vgl. Abb. 4):

Scanning: A mixture of deep spikes and short loops as users seek to cover a large area but without great depth.

Browsing: Many long loops and a few large rings, where users are happy to

go wherever the data takes them until their interest is caught.

Searching: Ever-increasing spikes with a few loops for users motivated to find a particular target.

Exploring: Many different paths, suggesting users who are seeking the extent and nature of the field.

Wandering: Many medium-sized rings as the user ambles along and inevitably revisits nodes in an unstructured journey.

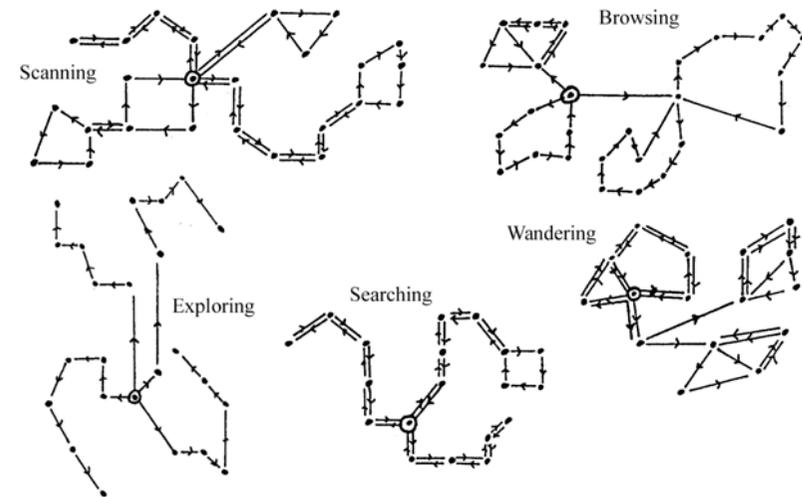


Abb. 4. Graphische Darstellung verschiedener Navigationsstrategien
(aus Canter, Rivers & Storrs, 1985).

Der Zusammenhang zwischen den verbal beschriebenen Strategien und dem Ausdrücken dieser Typen durch graphentheoretische Masse (Tab. 1) ist bei Canter et al. aus theoretischen Überlegungen heraus entstanden. Unklar ist, welche praktische Relevanz diese Typen haben, ob derartige Muster bei Probanden wirklich identifiziert werden können und ob die vorgenommene Operationalisierung der Typen durch die Indizes ausreichend und zutreffend ist.

Tab. 1. Zuordnung graphentheoretischer Masse zu den Navigationsstrategien nach Canter, Rivers und Storrs (1985).

	<i>NV/NT</i>	<i>NV/NS</i>	<i>SQ</i>	<i>LQ</i>	<i>RQ</i>	<i>PQ</i>
<i>Scanner</i>	high		high	medium		
<i>Browser</i>		medium		medium	medium	
<i>Searcher</i>		low	high	medium		
<i>Explorer</i>	high					high
<i>Wanderer</i>	medium	low			high	

(zur Bedeutung der Abkürzungen siehe Graphentheorie)

Ausblick

In der Hypermedia-Forschung sind Aufzeichnungen, Analysen und Auswertungen elektronischer Protokolle der Computernutzung (Logfiles) zum Standardwerkzeug geworden. Diese dienen der Erfassung, Beschreibung und Klassifizierung von Nutzeraktivitäten und werden darüber hinaus vielfach in Zusammenhang mit weiteren Daten der Nutzer gestellt. Man verspricht sich davon Hinweise zur Entwicklung optimierter Lernumgebungen bzw. die Nutzung verbesserter Verfahren der Evaluation von Nutzerverhalten in Hypertext.

Verschiedene Untersuchungen mit Logfile-Analysen haben gezeigt, dass das Navigationsverhalten von Lernenden in Hypertexten sehr unterschiedlich sein kann. Dies lässt sich in der Regel nicht auf einfache Weise durch Vorkenntnisse in der Domäne oder im Umgang mit Computern erklären. Viel mehr scheinen individuelle Präferenzen zu bestehen, die das Navigationsverhalten bestimmen. Diese können nicht als konstant angenommen werden, sondern unterliegen lang- und kurzfristigen Änderungen bzw. hängen von der Aufgabenstellung und der Lernumgebung ab. Der tatsächlich beschrittene Weg durch einen Hypertext dürfte darüber hinaus von einer Reihe spontaner und unbewusst vollzogener Aktivitäten («Klicken eines Links») beeinflusst sein, sodass es insgesamt (z. Zt. noch) schwer ist, von Navigationsprozessen auf kognitive Prozesse zu schließen oder umgekehrt. Es wäre deshalb Gewinn bringend, die Analyse von Logfiles dahingehend zu verfeinern, dass Verknüpfungen zwischen dem Durchlaufen von Mustern in Hypertexten und Zielen bzw. Absichten der User in Form von plausiblen Interpretationen möglich werden. Hierzu gehört zum einen, relevante Variablen zur Beschreibung der Navigation zu finden, diese durch «klassische» Erhebungsinstrumente abzusichern und ggf. nachvollziehbare Typisierungen damit aufzustellen. Zum anderen

sollte die auf einem Knoten verbrachte Zeit stärker berücksichtigt werden. Man kann beispielsweise davon ausgehen, dass eine bestimmte Aufenthaltsdauer auf einer Seite notwendig ist, damit eine Informationsaufnahme überhaupt erfolgen kann. Um hier zu gut begründeten Aussagen zu kommen, sind weitere Untersuchungen notwendig.

Die Möglichkeit der kompletten Aufzeichnung einer Mediennutzung ohne Beeinflussung des Anwenders stellt einen wesentlichen Vorteil der Logfile-Aufzeichnung gegenüber «klassischen» Methoden dar. Logfiles bieten damit einen Zugang auf statistische Daten der Nutzung von Lernumgebungen, der zuvor nicht bestand. Mit einer Verbesserung sowohl der technischen Aufzeichnungs- und Analysemittel (Standardisierung) als auch der Interpretation der gewonnenen Daten steht der interdisziplinären Lehr-Lernforschung mit neuen Medien ein sehr wertvolles Tool zur Verfügung.

Literatur

- Astleitner, H. (1997). *Lernen in Informationsnetzen*. Frankfurt am Main, Berlin, Bern, Brüssel, New York, Oxford, Wien: Lang.
- Ballstaedt, S.-P., Mandl, H., Schnotz, W. & Tergan, S.-O. (1981). *Texte verstehen. Texte gestalten*. München: Urban & Schwarzenberg.
- Barab, S. A., Bowdish, B. E. & Lawless, K. A. (1997). Hypermedia Navigation: Profiles of Hypermedia Users. *Educational Technology Research and Development*, 45 (3), S. 23–41.
- Bates, M. J. (1990). Where Should the Person Stop and the Information Search Start? *Information Processing and Management*, 26(5), 575-591.
- Botafogo, R. A., Rivlin, E. & Shneiderman, B. (1992). Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics. *ACM Transactions on Information Systems*, 10/2, S. 142–180.
- Canter, D., Rivers, R. & Storrs, G. (1985). Characterizing User Navigation Through Complex Data Structures. *Behaviour and Information Technology*, 4(2), S. 93–102.
- Choo, C. W., Detlor, B. & Turnbull, D. (2000). Information Seeking on the Web: An Integrated Model of Browsing and Searching. *First Monday. Peer-Reviewed Journal on the Internet*. <<http://www.firstmonday.dk>>.
- Degenhardt, M. (2001). Möglichkeiten empirischer Erfassung der Computernutzung von Schüler/innen im Unterricht. <www.medienpaed.com>.
- Fuller, G., Simonson, J., Tiwari, A. & Rebelsky, S. A. (2002). Clio's Assistants. A Tool Suite for Exploring Student Web Usage. In

- Association for the Advancement of Computing in Education AACE (Hrsg.), *Proceedings of ED MEDIA 2002*. Norfolk, VA/USA.
- Guthrie, J. T. (1988). Locating Information in Documents. Examination of a Cognitive Model. *Reading Research Quarterly*, 23, S. 179–199.
- Hadwin, A. F. & Leard, T. (2001). Navigation Profiles: Self-Regulating Learning Examined Through 5 Analytical Representations of Logfile Data. In A. F. Hadwin (Hrsg.), *Logfile Navigation Profiles and Analysis: Methods for Tracking and Examining Hypermedia Navigation*. Seattle, WA: Paper prepared for the Annual meeting 2001 of the AERA.
- Harter, S. P. (1986). *Online Information Retrieval: Concepts, Principles and Techniques*. Orlando, FL: Academic Press.
- Hawkins, D. T. & Wagers, R. (1982). Online Bibliographic Search Strategy Development. *Online*, 6(3), S. 12–19.
- Horney, M. A. (1993). Case Studies of Navigational Patterns in Constructive Hypertext. *Computers Educational*, 20(3), S. 257–270.
- Kerres, M. (1998). *Multimediale und telemediale Lernumgebungen. Konzeption und Entwicklung*. München, Wien: Oldenbourg.
- Kleining, G. (1999). *Qualitative Sozialforschung. Teil II: Der Forschungsprozess*. Hagen: FernUniversität.
- Körper, S. (2000). *Suchmuster erfahrener und unerfahrener Suchmaschinenutzer im deutschsprachigen World Wide Web. Ein Experiment*. Unveröffentlichte Magisterarbeit an der Westfälischen Wilhelms-Universität Münster.
- Lawless, K. A. & Kulikowich, J. M. (1998). Domain Knowledge, Interest and Hypertext Navigation: A Study of Individual Differences. *Journal of Educational Multimedia and Hypermedia*, 7(1), S. 51–69.
- Leard, T. & Hadwin, A. F. (2001). Logfile Analysis: A Review of the Literature. In A. F. Hadwin (Hrsg.), *Logfile Navigation Profiles and Analysis: Methods for Tracking and Examining Hypermedia Navigation*. Seattle, WA: Paper prepared for the Annual meeting 2001 of the AERA.
- Lempel, R. & Moran, S. (2000). The Stochastic Approach for Link-Structure Analysis (SALSA) and the TKC Effect. In R. More (Hrsg.), *9th International World Wide Web Conference*.
- Levene, M. & Loizou, G. (2001). Web Interaction and the Navigation Problem in Hypertext. In A. Kent, J. G. Williams & C. M. Hall (Hrsg.), *Encyclopedia of Microcomputers*. New York: Marcel Dekker.
- Marchionini, G. (1995). *Information Seeking in Electronic Environments*. Cambridge: Cambridge University Press.
- Markey, K. & Cochrane, P. (1981). *Online Training and Practice Manual for ERIC Database Searchers*. Syracuse, NY: ERIC Clearinghouse on Information Resources.
- Meadow, C. T. & Cochrane, P. A. (1981). *Basics of Online Searching*. New York: Wiley.
- Modjeska, D. & Marsh, A. (1997). *Structure and Memorability of Web Sites. CSRI Technical Report*. Toronto: University of Toronto.
- Navarro-Prieto, R., Scaife, M. & Rogers, Y. (1999). *Cognitive Strategies in Web Searching*. <<http://zing.ncsl.nist.gov/hfweb/proceedings/navarro-prieto>>.
- Niegemann, H. M. (2001). Analyzing Navigation Patterns of Learning in Hypermedia Learning Environments. In A. F. Hadwin (Hrsg.), *Logfile Navigation Profiles and Analysis: Methods for Tracking and Examining Hypermedia Navigation*. Seattle, WA: Paper prepared for the Annual meeting 2001 of the AERA.
- Noller, S., Naumann, J. & Richter, T. (2001). LOGPAT – ein webbasiertes Tool zur Analyse von Navigationsverläufen im Hypertext. In *Deutsche Gesellschaft für Online-Forschung (DGOF)*. <http://www.psych.uni-goettingen.de/congress/gor-2001/contrib/contrib/abstracts_oral.html>.
- Priemer, B. (2004). Physiklernen mit dem Internet. Das World Wide Web als Informationsquelle für Schüler zur Erarbeitung des Themas «Die Entstehung der Gezeiten». Dissertation: in Druck.
- Qiu, L. (1993). Markov Models of Search State Patterns in a Hypertext Information Retrieval System. *Journal of the American Society for Information Science*, 44(7), S. 413–427.
- Reed, W. M., Oughton, J. M., Ayersman, D. J., Ervin, J. R. & Giessler, S. F. (2000). Computer Experience, Learning Style, and Hypermedia Navigation. *Computers in Human Behavior*, 16, S. 609–628.
- Rossmannith, T. (2001). *Informationsverhalten und Involvement im Internet. Eine Labor- und Feldstudie zu den Determinanten der Informationsnachfrage im World Wide Web*. Karlsruhe: Dissertation.
- Sarrukai, R. R. (2000). Link Prediction and Path Analysis Using Markov Chains. *Computer Networks*, 33(1-6), S. 377–386.
- Shneiderman, B., Byrd, D., & Croft, W. B. (1997). Clarifying Search. A User-Interface Framework for Text Searches. In *D-Lib Magazine*. <<http://www.dlib.org/dlib/january97/01.shneiderman.html>>.

- Stanton, N. & Baber, C. (1992). An Investigation of Styles and Strategies in Self-Directed Learning. *Journal of Educational Multimedia and Hypermedia*, 1 (2), S. 147–167.
- Tauscher, L. & Greenberg, S. (1997). How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems. *International Journal of Human-Computer Studies*, 47, S. 97–137.
- Unz, D. (2000). Lernen mit Hypertext – Informationssuche und Navigation. In *Internationale Hochschulschriften* (Bd. 326). Münster, New York, München, Berlin: Waxmann.
- Valdez, F., Chignell, M. & Glenn, B. (1988). Browsing Models for Hypermedia Databases. *Proceedings of Human Factors Society 32nd Annual Meeting*. S. 318–322.
- Waterworth, J. A. (1992). *Multimedia Interactions with Computers*. New York: Ellis Horwood.
- Wenzel, O. (2001). *Webdesign, Informationssuche und Flow*. Lohmar, Köln: Joseph Eul Verlag.
- Wilde, K. D., Hippner, H. & Merzenich, M. (2002). Web Usage Mining. *WiSt – Wirtschaftswissenschaftliches Studium*, 31 (2).
- Winne, P. H., Gupta, L. & Nesbit, J. (1994). Exploring Individual Differences in Studying Strategies Using Graph Theoretic Statistics. *The Alberta Journal of Educational Research*, XL, 2, S. 177–193.
- Zajonc, R. & Priemer, B. (2001). Time – Logfiles' Neglected Data. In A. F. Hadwin (Hrsg.), *Logfile Navigation Profiles and Analysis: Methods for Tracking and Examining Hypermedia Navigation*. Seattle, WA: Paper prepared for the Annual meeting 2001 of the AERA.