

E-Learning aus Prozessperspektive

Stefan Iske

DFG-Graduiertenkolleg „Qualitätsverbesserung im E-Learning
durch rückgekoppelte Prozesse“

TU-Darmstadt
Alexanderstraße 6
64283 Darmstadt
s.iske@apaed.tu-darmstadt.de

Erschienen in: Eibl, C./Magenheim, J./Schubert, S./Wessner, M. (Hrsg.): DeLFI 2007: die 5. e-Learning Fachtagung Informatik. Bonn: Gesellschaft für Informatik, S. 21-32.

Abstract: In diesem Artikel wird das Verfahren und das Potenzial der Analyse von E-Learningprozessen mittels explorativ-heuristischer Optimal-Matching Analyse dargestellt. Ziel ist das Identifizieren von Mustern, Regelmäßigkeiten und Strukturen in Navigationsprozessen. E-Learning als Prozess der Interaktion von Lernern mit einer Online-Lernumgebung wird so analysierbar. Auf Grundlage der Levenshtein-Distanz des paarweisen Vergleichs aller Navigationsverläufe können in Verbindung mit Verfahren der Clusteranalyse empirische Navigationsverläufe induktiv typologisiert werden. In methodologischer Perspektive wird das vorgeschlagene Verfahren abgegrenzt von der Analyse aggregierter Logdaten sowie von der auf Markov-Prozessen beruhenden Ereignisanalyse.

1 Einleitung

Bei offenen Online-Lernumgebungen handelt es sich aus medialer Sicht um Hypertexte, mit den Hauptkennzeichen der Fragmentierung und der Verknüpfung [vgl. Ku91]: Ein Gegenstandsbereich wird in Informationseinheiten gegliedert und durch Links untereinander verbunden. Die so entstehende nicht-lineare Netzstruktur ist das Grundmodell von Hypertext [vgl. BC90].

Der Navigationsprozess als Interaktion eines Nutzers mit einer hypertextuellen Online-Lernumgebung entspricht dem linearen Entfalten eines nicht-linearen Hypertextes. Hypertexte entfalten sich angesichts eines Nutzers nicht von allein, sondern sind auf dessen Aktivitäten angewiesen, auf das aktive Auswählen von Verknüpfungen. Genau auf dieses Auswählen und Entfalten zielt die Kennzeichnung als Pull-Medium. Durch die Auswahl konkreter Verknüpfungen aus einer Vielzahl möglicher Verknüpfungen entsteht ein zeitlich-linearer Nutzungspfad. Aus diesem Grund werden Hypertexte auch als multilinear oder multisequenziell bezeichnet [vgl. La97].

Beim Lernen in hypertextuellen Lernumgebungen eignen sich Lernende Wissen an, indem sie ausgehend von ihrem zeitlich-linearen Navigationsprozess ein mentales, nicht-lineares Modell des Gegenstandsbereichs entwickeln [vgl. Me06]. Im Kern handelt es sich bei dieser Autodidaktik um ein Übersetzen von Zeitgestalten in Raumgestalten [vgl. Hö27]. Dabei spielen neben der Zeit- und Ortsunabhängigkeit des E-Learning vor allem Entscheidungen über die Auswahl der Inhalte sowie Entscheidungen über den Lernweg eine besondere Rolle. Aus didaktischer Perspektive wird durch den multisequenziellen Nutzungspfad autodidaktisches Handeln als Strategien und Metaregeln des Entfaltens rekonstruierbar. All dies macht deutlich, dass es sich beim Navigieren in Online-Lernumgebungen um anspruchsvolle und komplexe Tätigkeiten handelt [Is02].

Die Kennzeichnung der Hypertext-Technologie als Pull-Medium mit der Notwendigkeit des Entfaltens ist auch Ausgangspunkt eines relationalen Qualitätsverständnisses des E-Learning [vgl. Eh02]. Dem Lernenden kommt dabei die Rolle eines Koproduzenten zu: Die Online-Lernumgebung liefert beispielsweise den Inhalt und Kommunikationswerkzeuge, der Lernende muss jedoch selbst tätig, selbst aktiv werden. So betont auch Ehlers [Eh02:9] die zentrale Bedeutung der Interaktion für die Konstitution von Qualität: „Qualität entsteht erst dann, wenn der Lernende mit dem Lernarrangement in Interaktion tritt. Erst dann, wenn gelernt wird entsteht auch Lernqualität (Ko-Produktion des Lernerfolges). Ein E-Learning-Lernarrangement hat keine Lernqualität an sich. Es ist lediglich der Rahmen (das Arrangement) mit Hilfe dessen sich der Lernprozess vollzieht.“ Dabei unterscheidet Ehlers [Eh02] verschiedene Qualitätsebenen: die Voraussetzungen (‚Inputqualität‘), den Lernprozess (‚Prozessqualität‘) und das Ergebnis (‚Outcomequalität‘). Insbesondere die in diesem Artikel zur Diskussion stehende Analyse von Navigationsprozessen bezieht sich auf die Relation von Lernendem, Lernumgebung und Lerninhalt. Eine konsequente Qualitätsforschung des E-Learning unter Berücksichtigung der Lernerperspektive muss also neben der ‚Inputqualität‘ und der ‚Outcomequalität‘ vor allem die beschriebenen Relationen der ‚Prozessqualität‘ berücksichtigen.

Auf welcher methodologischen Grundlage kann E-Learning aus Prozessperspektive analysiert werden? Wie kann die Interaktion als das Entfalten einer Online-Lernumgebung Gegenstand der Analyse werden?

2 Analyse aggregierter und sequenzierter Logdaten

In der Regel wird die Interaktion eines Nutzers mit einer Online-Lernumgebung durch serverseitige Logdaten aufgezeichnet. Die Logdaten stellen eine spezifische Art der Dokumentation und Transkription des Navigationsprozesses dar. Durch die Logdaten wird die Abfolge der vom Nutzer aufgerufenen Seiten in eine digitale Textdatei übersetzt, also von im voraus definierten Algorithmen transkribiert (z.B. im Apache-Logfileformat). Es handelt sich dabei um eine formale, automatisierte Protokollierung der Reaktionen des Webservers, die während des Navigationsprozesses automatisch generiert und aufgezeichnet werden, um eine Form passiver Protokollierung. Priemer

[Pr04] nennt als Vorteile dieser Protokollierung u.a. die unbemerkte, detailgenaue, objektive und non-reaktive Aufzeichnung der Handlungen des Nutzers in authentischen Nutzungssituationen. Wie lassen sich diese Logdaten nun auswerten? Grundsätzlich kann die Analyse aggregierter Logdaten von der Analyse sequenzierter Logdaten unterschieden werden.

Die Analyse *aggregierter* Logdaten stellt das am weitesten verbreitete Verfahren zur Analyse der Nutzung von Online-Lernumgebungen dar. Viele Internet-Provider stellen ihren Kunden die Analyse der Logdaten in Form aggregierter Nutzungsdaten zur Verfügung. Aggregiert bedeutet in diesem Zusammenhang, dass die in den Logdaten enthaltenen Informationen zusammengefasst werden und darauf aufbauend durchschnittliche Kennzahlen berechnet werden (z.B. durchschnittliche Nutzungsdauer, durchschnittliche Anzahl der besuchten Seiten u.ä.). Im Vordergrund steht dabei die deskriptiv-statistische Analyse der Logdaten mit der grundlegenden Orientierung am Querschnittsdesign der Datenerhebung und -analyse.

Die systematische Berücksichtigung *sequenzierter* Logdaten findet im Bereich des E-Learning bisher lediglich in Ansätzen statt. Die grundlegende Analyseeinheit der Sequenzanalyse sind verlaufsbezogene Daten. Diese stellen die Navigationsverläufe von Nutzern in einer Online-Lernumgebung dar: der Verlauf der besuchten Internetseiten wird dabei als Sequenz dokumentiert und als Sequenz analysiert. In den sequenzierten Daten sind also Informationen über den zeitlichen Verlauf enthalten, die bei der Aggregation verloren gehen. Allgemeines Kennzeichen von Verlaufsdaten ist die Orientierung am Längsschnittsdesign der Datenerhebung und -analyse, d.h. die Daten werden *wiederholt* in definierten zeitlichen Intervallen bei *gleichen* Individuen erhoben. Während Sequenzdaten Analysen des Verlaufs bzw. der Entwicklung von Navigationsprozessen ermöglichen, sind diese Analysen auf der Grundlage aggregierter Daten grundsätzlich nicht möglich [vgl. IS05].

Sozialwissenschaftliche Forschungen zur Analyse der Internetnutzung basieren bisher auf der Analyse aggregierter Logdaten: Bei der Analyse stehen deskriptive und inferenzstatistische Verfahren im Vordergrund. Bei diesem Forschungsdesign sind jedoch die *Prozesse* der Nutzung selbst – z.B. der konkret zeitliche Verlauf der Navigation - nicht Gegenstand der Analyse. Aus erziehungswissenschaftlicher Perspektive ist diese Fokussierung auf aggregierte Daten unbefriedigend, da die Kenntnis von Aneignungsprozessen Lernender zum einen Aufklärung über pädagogisch-didaktisches Handeln erwarten lässt und zum anderen die Möglichkeit der empirisch basierten Optimierung der Online-Lernumgebung eröffnet.

Wie kann nun der E-Learningprozess anhand sequenzierter Logdaten analysiert werden? Zur Beantwortung dieser Frage werden im folgenden Markov-Ketten und das Optimal-Matching Verfahren skizziert.

2.1 Analyse von Navigationssequenzen mittels Markov-Ketten

Den zentrale Anwendungsbereich von Markov-Ketten (bzw. -Prozesse) bildet die Analyse der Abfolge von Zuständen durch die Berechnung von Übergangswahrscheinlichkeiten: Mit welcher Wahrscheinlichkeit folgt auf dem Zustand A der Zustand B? Auf den Zustand B der Zustand C? Grundlegende Elemente von Markov-Ketten sind ein *Zustandsraum* (als nichtleere, endliche Menge) und eine *stochastische Matrix*, die die Wahrscheinlichkeit enthält, von einem spezifischen Zustand in einem Schritt in einen Folgezustand überzugehen. Ein frühes Beispiel für die Verwendung von Markov-Ketten in der Analyse von Lehr-Lernprozessen stellen Flanders Interaktionsanalyse-Kategorien dar [vgl. Fl70]: An der darauf aufbauenden stochastischen Matrix ist ablesbar, auf welches Lehrerhandeln mit welcher Wahrscheinlichkeit ein bestimmtes Schülerverhalten folgt. Gegenstand der Analyse sind also dyadische Beziehungen von Lehrerhandeln und Schülerreaktion.

Als Markov-Prozesse erster Ordnung werden genau solche Prozesse definiert, bei denen das Auftreten *folgender* Zustände lediglich vom *momentanen* Zustand abhängt und nicht von *vorangehenden* Zuständen beeinflusst wird. Analyseeinheit ist der isolierte Übergang von je zwei Zuständen. Damit wird eine ‚Gedächtnislosigkeit‘ des Prozesses postuliert: die Eintrittswahrscheinlichkeit eines Zustandswechsels, d.h. der Übergangswahrscheinlichkeit eines Markov-Prozesses wird *nicht* von dessen ‚Vorgeschichte‘ beeinflusst und kann demzufolge *unabhängig* von den vorangehenden Zuständen prognostiziert werden. Mit anderen Worten: zusätzliche Informationen über die ‚Vergangenheit‘ des Prozesses in Form vorangehender Zustände verbessern *nicht* die Prognose der folgenden Zustände.

Diese Markov-Prozesse erster Ordnung werden durch das Konzept von Markov-Prozessen zweiter Ordnung erweitert, die auch als Semi-Markov-Ketten bezeichnet werden: die Erweiterung besteht darin, dass bei Markov-Ketten zweiter Ordnung nicht ausschließlich der momentane Zustand zur Prognose des folgenden verwendet wird, sondern eine *begrenzte Anzahl* vorangehender Zustände. Das Postulat der ‚Gedächtnislosigkeit‘ des Markov-Prozesses wird damit erweitert zur Berücksichtigung von ‚Vergangenheit‘. Mit dieser Erweiterung wird Prozessen Rechnung getragen, die nicht als ‚gedächtnislos‘ im Hinblick auf den *Prozessverlauf* oder die *Prozesszeit* betrachtet werden können. Gemeinsam sind beiden Konzepten jedoch die stochastische Grundlage und die Analyse isolierter Übergänge, auch wenn diese durch Markov-Ketten zweiter Ordnung um eine begrenzte Anzahl vorangehender Zustände erweitert werden.

Zusammenfassend kann festgestellt werden, dass das Potenzial der Markov-Ketten in der Analyse von Determinanten von Übergängen sowie in der empirischen Überprüfung entsprechender Modelle und Hypothese besteht. Markov-Prozesse bilden damit die Grundlage der Ereignisdatenanalyse („event history analysis“, vgl. BR02).

2.2 Analyse von Navigationssequenzen mittels Optimal-Matching

Für explorativ-heuristische Analyse von Prozessen des E-Learning sind Markov-Ketten von begrenztem Nutzen. Prozesse des E-Learning können gerade nicht als „gedächtnislos“ interpretiert und analysiert werden, sondern sind abhängig von der Verweildauer in den betreffenden Zuständen sowie von ihrer ‚Vorgeschichte‘. Ein größeres analytisches Potenzial verspricht daher die Verwendung der Sequenzanalyse mittels Optimal-Matching, da diese gerade mit der Analyse vollständiger Sequenzen ein „Prozessgedächtnis“ beinhaltet. Gerade in diesem ‚Prozessgedächtnis‘ zeigen sich ja spezifische Navigationsstrategien und Metaregeln. Im Gegensatz zu dem hypothesengeleitet-konfirmatorischen Vorgehen der Ereignisdatenanalyse stellt die Sequenzdatenanalyse mittels Optimal-Matching ein deskriptives, explorativ-heuristisches Vorgehen dar.

Allgemein versteht man unter einer „Sequenz“ eine Abfolge oder Reihenfolge von Elementen. Als Prototyp einer Sequenz in den Naturwissenschaften – insbesondere in der Molekularbiologie - gilt die DNA als Träger des menschlichen Erbgutes. Als Prototyp einer Sequenz in der Soziologie kann der Lebenslauf bezeichnet werden (vgl. Er01; SW01; Wi01). Den Prototyp einer Sequenz im Kontext von Online-Lernumgebungen stellt der Prozess der Navigation in einer hypertextuellen Lernumgebung dar. Diese Navigationssequenz basiert auf der zeitlichen Abfolge besuchter Seiten als Elemente der Navigationssequenz. Aufbauend auf den Logdaten als „elektronischen Prozessdaten“ (vgl. BM00) können die Navigationsprozesse von Nutzern in einer Lernumgebung als Sequenz rekonstruiert werden. Diese Navigationssequenzen bilden dann den Ausgangspunkt der Sequenzanalyse mittels Optimal-Matching.

Ein spezifischer Algorithmus zur Sequenzanalyse ist das Optimal-Matching Verfahren¹, dessen Verwendung im Bereich der Sozialwissenschaften ein relativ junges methodisches Vorgehen darstellt. Übergeordnetes Ziel des Optimal-Matching Verfahrens ist die auf einem Algorithmus beruhenden Analyse einer großen Anzahl komplexer und oftmals sehr langer Sequenzen mit dem Ziel, *Muster, Strukturen und Regelmäßigkeiten* zu identifizieren. Das Standard- und Referenzwerk der Sequenzanalyse im naturwissenschaftlichen Bereich ist das von David Sankoff und Joseph Kruskal herausgegebene „Time Warps, String Edits, and Macromolecules“ [KS99] aus dem Jahr 1983. Als beispielhafte Anwendungsgebiete der Sequenzanalyse nennen sie u.a. die Analyse der Homologie von Makromolekülen, die Sprecher- oder auch Spracherkennung und den Bereich der technischen Datenübertragung, aus dem auch die grundlegenden Forschungen von Levenshtein (Le66) stammen. Der Transfer dieser Methode auf den sozialwissenschaftlichen Bereich und speziell in den Bereich der

¹ Der Begriff „Optimal-Matching“ wird als Sammelbegriff für Verfahren verwendet, die auf Grundlage der Levenshtein-Distanz und der Operationen *Austauschen* („substitution“), *Einfügen* („insertion“) sowie *Löschen* („deletion“) unter Verwendung iterativer Prozeduren (Algorithmen) die Distanz von Sequenzen bestimmen. Der Begriff der „Sequenzanalyse“ wird verwendet für die übergeordnete Methodologie, der Begriff „Optimal-Matching“ für einen konkreten Algorithmus zur deren Umsetzung.

Soziologie geht auf Abbott (vgl. AF86] zurück.² In der pädagogisch-didaktischen Forschung und der Analyse von E-Learningprozessen findet die Sequenzanalyse mittels Optimal-Matching bislang keine Verwendung.

Den Ausgangspunkt der Sequenzanalyse mittels Optimal-Matching bildet die Frage nach der Bestimmung der Ähnlichkeit von Sequenzen: Wie kann festgestellt werden, ob und wie stark sich Sequenzen ähneln? Mit der Hamming- und der Levenshtein-Distanz werden im Folgenden zwei grundlegend unterschiedliche Antworten skizziert.

Das Konzept der Hamming-Distanz (Ha50) stammt aus dem Bereich der elektronischen Datenübertragung und stellt dort ein Verfahren zur Kontrolle von Übertragungsfehlern dar. Die gesendete und die empfangene Datensequenz werden Position für Position hinsichtlich identischer Elemente verglichen. Das Ergebnis dieses Vergleichs ist eine Maßzahl, die als Hamming-Distanz bezeichnet wird. Der Grad der Ähnlichkeit zweier Sequenzen steigt mit der Anzahl identischer Elemente in der gleichen Position. Bei einer maximalen Ähnlichkeit besteht eine Übereinstimmung der Elemente in jeder Position der zu vergleichenden Sequenzen (d.h. die Sequenzen sind identisch); bei einer maximalen Unähnlichkeit besteht keine Übereinstimmung von Elementen an keiner Position. Die Bestimmung der Hamming-Distanz stößt jedoch dort an Grenzen, wo sie über die Kontrolle von Übertragungsfehlern hinaus in Bereichen angewandt wird, in denen nicht von einer inhärenten Korrespondenz als Entsprechung der Positionen ausgegangen werden kann oder Sequenzen unterschiedlicher Länge miteinander verglichen werden. Für eine Analyse von Prozessen des E-Learning sind daher differenziertere Verfahren des Sequenzvergleichs erforderlich.

Den Kern der Optimal-Matching Analyse bildet der paarweise Vergleich aller Sequenzen eines Datensatzes. Die Distanz zweier Sequenzen als Grad der Unähnlichkeit wird dabei bestimmt in Abhängigkeit der Anzahl der Transformationsschritte³ die erforderlich sind, um eine *Ausgangssequenz* in eine *Zielsequenz* zu überführen und somit eine *Übereinstimmung* („alignment“⁴) herzustellen. Je weniger Operationen benötigt werden, um eine Übereinstimmung herzustellen, umso ähnlicher sind sich die Sequenzen. Die Transformationen beruhen dabei auf den grundlegenden Operationen des *Löschens* („deletion“), *Einfügens* („insertion“) und *Austauschens* („substitution“) von Elementen. Dieses Verfahren wird als „Optimal-Matching“ bezeichnet und beruht auf zwei Prozessen [vgl. KS99]: Auf der Bestimmung aller möglichen Transformationsoperationen, um eine Quellsequenz in die Zielsequenz zu überführen („alignment analysis“); sowie auf der Ermittlung der geringsten Anzahl der dazu notwendigen Operationen („optimum analysis“). Die geringste Anzahl der erforderlichen Operationen

² Eine softwaretechnische Umsetzung findet der OM-Algorithmus in dem Programm „Transition Data Analysis“ (TDA) von Götz Rohwer und Ulrich Pötter; <<http://www.stat.ruhr-uni-bochum.de/tman.html>>.

³ In diesem Artikel werden die Begriffe „Transformation“ und „Operation“ synonym verwendet. Diese (Bearbeitungs)Operationen werden auch als „Edit-Operations“ bezeichnet. Die Levenshtein-Distanz wird daher auch als „Edit-Distance“ bezeichnet.

⁴ Aus diesem Grund wird die Methode des „Optimal Matching“ auch als „Sequence Aligment Method“ bezeichnet.

zur Herstellung des „alignments“ dient dann als Maßzahl für den Grad der Unähnlichkeit zwischen Sequenzen und wird als Levenshtein-Distanz bezeichnet (Le66).

Im Folgenden wird die Bestimmung der Levenshtein-Distanz beispielhaft am paarweisen Vergleich dreier Sequenzen verdeutlicht. In *Abbildung 1* ist der paarweise Vergleich einer Sequenz 1 mit einer Sequenz 2 dargestellt: Zunächst wird in der Ausgangssequenz das erste Element „A“ gelöscht („deletion“). Danach wird ein Element „A“ eingefügt („insertion“). Es sind also minimal zwei Operationen notwendig, um eine Übereinstimmung der Ausgangs- mit der Zielsequenz herzustellen. In *Abbildung 2* ist der paarweise Vergleich der Sequenz 1 mit einer Sequenz 3 dargestellt: An die erste Position der Ausgangssequenz wird das Element „G“ eingefügt („insertion“). Damit verschieben sich alle folgenden Positionen. Es ist also minimal eine Operation notwendig, um eine Übereinstimmung („alignment“) mit der Zielsequenz herzustellen. Im Gegensatz zum Hamming-Algorithmus, der für diesen Fall eine maximale Unähnlichkeit feststellt, ist der Optimal-Matching Algorithmus aufgrund der grundlegenden Operationen in der Lage, Regelmäßigkeiten innerhalb der zu vergleichenden Sequenzen zu identifizieren: in diesem Beispiel sind die Ausgangs- und die Zielsequenz gegeneinander verschoben. In *Abbildung 3* ist der paarweise Vergleich der Sequenz 2 mit einer Sequenz 3 dargestellt: In der Ausgangssequenz wird zunächst ein Element „G“ eingefügt („insertion“). Danach wird das Element „B“ gelöscht („deletion“) und ein Element „B“ eingefügt („insertion“). Es sind also minimal drei Operationen notwendig, um eine Übereinstimmung mit der Zielsequenz herzustellen.

Das Ergebnis des paarweisen Vergleichs aller Sequenzen des Datensatzes wird in Form einer Distanz-Matrix dokumentiert, in der die Levenshtein-Distanz für jeden paarweisen Sequenzvergleich eingetragen wird.⁵ Die Levenshtein-Distanzmatrix als Ergebnis der Optimal-Matching Analyse bildet dann den Ausgangspunkt für sowohl explorativ-heuristische als auch konfirmatorische Forschungsstrategien [vgl. KS99].

Die Analyse von E-Learningprozessen mittels Optimal-Matching entspricht einer explorativ-heuristischen Forschungsstrategie. In einem ersten Schritt werden Sequenzen unter dem Gesichtspunkt der Distanz verglichen, um in einem zweiten Schritt mit Hilfe von Methoden der Clusteranalyse zu homogenen Gruppen ähnlicher Sequenzen zusammengefasst zu werden. Mit Erzberger [Er01] kann die explorativ-heuristische Sequenzanalyse als fallorientierte Analysestrategie gekennzeichnet werden, bei der Sequenzen als Gesamtverläufe bzw. Verlaufsgeschichten in ihrer Vielfalt und Komplexität zum Gegenstand der Forschung werden. Die „Zusammenschau aller Verläufe läßt dann Ordnung entstehen“ [Er01:36], d.h. in der Gesamtschau einer hinreichend großen Anzahl von Sequenzen werden Muster und Regelmäßigkeiten überhaupt erst erkennbar. Die Optimal-Matching Analyse als explorativ-heuristisches Verfahren ermöglicht es, „typische Muster, die sich aus der Empirie ergeben, theoretisch

⁵ An dieser Stelle zeigt sich die Rechenintensität des Optimal-Matching Verfahrens, besonders bei einer großen Anzahl sowie langen Sequenzen: Ein Vergleich von 100 Sequenzen beruht auf 4950 paarweisen Sequenzvergleichen ($100 * 99 / 2 = 9900 / 2 = 4950$), die in die Levenshtein-Distanzmatrix eingetragen werden.

aber nicht 'vorgedacht' wurden“ [Ai00:15] zu identifizieren. Muster in empirischen Sequenzen, die theoretisch nicht 'vorgedacht' wurden, können durch eine konfirmatorisch-deduktive Analyse nicht identifiziert werden.

In Abbildung 4 wird beispielhaft das Ergebnis der Sequenzanalyse mittels Optimal-Matching mit daran anschließender Clusteranalyse dargestellt: die unterschiedlichen Quadrate stehen für unterschiedliche Informationseinheiten; die Abfolge der Informationseinheiten von links nach rechts stellt die Sequenz der besuchten Informationseinheiten dar (als Beschreibung des Navigationsprozesses). Die horizontalen Linien trennen dabei einzelne Cluster ähnlicher Sequenzen. Klammer 1 kennzeichnet ein Cluster ähnlicher Navigationssequenzen. Das für dieses Cluster typische Navigationsmuster ist mit der Ziffer 2 gekennzeichnet. Ziffer 3 markiert eine Sequenz, die sich von dem clustertypischen Navigationsmuster lediglich durch die Informationseinheit an erster Position unterscheidet. Fügt man vor der ersten Position ein Element ein (weiße Informationseinheit), erhält man als Ergebnis das clustertypische Navigationsmuster. Beide Navigationsmuster sind also auf Grundlage des Optimal-Matching Algorithmus als ähnlich einzustufen, da für eine Übereinstimmung lediglich eine Einfügen-Operation notwendig ist. Ziffer 4 markiert Navigationsmuster, die sich vom clustertypischen Navigationsmuster lediglich durch einen Einschub eines Elementes an der zweiten Position der Sequenz unterscheiden (sowie durch ein zusätzliches Element am Ende der Sequenz). Beide Navigationsmuster sind also auf Grundlage des Optimal-Matching Algorithmus als ähnlich einzustufen, da für eine Übereinstimmung lediglich zwei Löschen-Operationen notwendig sind. Ziffer 5 markiert ein Navigationsmuster, das sich vom clustertypischen Navigationsmuster lediglich durch ein zusätzliches Element am Ende der Sequenz unterscheidet. Beide Navigationsmuster sind auf Grundlage des Optimal-Matching Algorithmus als ähnlich einzustufen, da für eine Übereinstimmung dieser Sequenzen lediglich eine Löschen-Operation notwendig ist.

3. Ergebnisse und Ausblick

Grundlegend ist festzuhalten, dass auf Grundlage des Optimal-Matching Verfahrens eine große Anzahl komplexer und oftmals sehr langer Navigationssequenzen analysiert werden können. In den Navigationssequenzen enthaltene *Muster, Strukturen und Regelmäßigkeiten* werden identifizierbar. Damit unterscheidet sich die vorgeschlagene Methodologie grundlegend von der qualitativen Analyse einzelner Navigationssequenzen. Der Navigationsverlauf als Sequenz wird zum Ausgangspunkt und zur Analyseeinheit: Die beschriebene Methodologie aus Optimal-Matching Analyse in Kombination mit Verfahren der Clusteranalyse ermöglicht es, Navigationsprozesse in Online-Lernumgebungen anhand sequenzierter Logdaten zu analysieren. Auf dieser Grundlage können auf induktive Weise ähnliche Navigationsprozesse identifiziert und zu Gruppen zusammengefasst werden. Damit kann die vorherrschende Dominanz der Analyse der Resultate von E-Learningprozessen aufgebrochen werden (vgl. ‚Outcomequalität‘) und gleichzeitig einem relationalen Qualitätsbegriff sowie dem grundlegenden Charakter eines Pull-Mediums Rechnung getragen werden. Die Berücksichtigung von Prozessen des E-Learning ist insbesondere dann notwendig, wenn mit Online-Lernumgebungen die Vermittlung prozeduralen Wissens und tätigkeitsorientierter Kompetenzen angestrebt wird.

Dabei zielt die Analyse von E-Learningprozessen mittels Optimal-Matching auf den Kern einer Didaktik als Handlungswissenschaft, die die konkret-empirische Abbildung von Raumgestalten in Zeitgestalten (Didaktik) bzw. die Abbildung von Zeitgestalten in Raumgestalten (Autodidaktik) analysiert und hinsichtlich Adäquatheit und alternativer Möglichkeiten reflektiert (vgl. Me03; Me06). Gerade Hypertext als grundlegende Technologie von Online-Lernumgebungen stellt ja einen radikalen medialen Strukturwandel dar, in dem bisherige Prozesse der Abbildung grundlegend zur Disposition stehen.

Detailliert wurde die Methodologie und das Potenzial der Sequenzanalyse mittels Optimal-Matching für die Analyse von E-Learningprozessen an anderer Stelle ausgearbeitet, am Beispiel der Analyse der Nutzung einer metadatenbasierten, hypertextuellen Online-Lernumgebung [vgl. Is07]. Analysiert wurden dabei insgesamt ca. 1600 Sequenzen mit insgesamt ca. 4700 Informationseinheiten. Dabei wurden spezifische Muster, Regelmäßigkeiten und Strukturen in Navigationssequenzen identifiziert, die sowohl in formaler als auch in inhaltlicher Hinsicht als Navigationsstrategien interpretiert werden konnten: Z.B. lineare Navigationsmustern der „Erkundung“ und der „Auseinandersetzung“, sowie nicht-lineare Muster der direkten und gezielte Navigation: Diese können als „erklärungsfokussierte“, „beispielfokussierte“, „aufgabenfokussierte“ und „testfokussierte“ Navigationsstrategien gekennzeichnet werden. Neben linearen und nicht-linearen Navigationsmustern wird auch die Fokussierung des Navigationsprozesses auf spezifische dominierende Typen von Informationseinheiten erkennbar. Darüber hinaus werden auch Navigationsmuster identifizierbar, in denen spezifische Informationseinheiten gerade nicht enthalten sind.

Die dargestellte Methodologie der Analyse von E-Learningprozessen leistet einen zentralen Beitrag für die Qualitätsentwicklung von E-Learning unter Prozessperspektive („Prozessqualität“), d.h. hinsichtlich der Relation von Lernendem, Lernumgebung, Lerninhalt. Der Lernende und seine Handlungen werden zum zentralen Gegenstand der Analyse. Damit kommen Fragen der Interaktivität in den Fokus und besonders Fragen des Potenzials von Prozessen der Rückkopplung und des Feedback zur Unterstützung von Online-Lernprozessen. Die Kenntnis der konkreten Navigationsprozesse ist dabei die Voraussetzung für differenzierte Rückmeldungen – personal sowie digital – und bildet die Grundlage für Strategien der Mikro-Adaptation [vgl. Le92]. Gerade neuere Entwicklungen wie *Web 2.0*, *Social Software* und *Social Navigation* versprechen neuartige und vielfältige Möglichkeiten der Art, des Umfangs und des Zeitpunktes einer lernförderlichen Rückkopplung.

Bei *Social Software* und *Social Navigation* tritt der kooperative Aspekt des E-Learning in den Vordergrund und geht damit weit über die 1 : 1 Situation eines isolierten Lerners vor einem Computer hinaus: Lernen wird zunehmend als sozialer Prozess verstanden, als Lernen *in* einer Gruppe und Lernen *von* einer Gruppe. Insbesondere stellt die Kenntnis konkreter Navigationsprozesse und deren Analyse die Voraussetzung für die Entwicklung differenzierter pädagogisch-didaktischer Empfehlungssysteme als spezifische Form lernförderlicher Rückkopplung dar. Analog zu *amazon.de* interpretiert ein Empfehlungssystem die Handlungen des Nutzers und gibt auf Grundlage dieser Interpretation Empfehlungen, die für den Nutzer hilfreich sind, d.h. ihn in seinem E-Learningprozess hilfreich unterstützen. Gegenwärtige Empfehlungssysteme auf der Grundlage von Assoziationsanalysen sind vor allem aus dem Bereich des E-Commerce bekannt, z.B. als *Warenkorbanalyse*. Dabei steht die Analyse von Nutzungsinformationen in Hinblick auf gemeinsame Interesse der Nutzer im Vordergrund. Auf Grundlage dieser *Warenkorb-analyse* werden Kaufempfehlungen für Nutzer abgeleitet, wie dies z.B. bei *amazon.de* als dynamisches Empfehlungssystem implementiert ist: „Kunden, die diesen Artikel gekauft haben, kauften auch:“, „Kunden, die diesen Artikel angesehen haben, haben auch angesehen:“, „Unser Vorschlag: Kaufen Sie jetzt diesen Artikel zusammen mit“. Allerdings unterscheiden sich die Aktivitäten eines Käufers von denen eines Lernenden, der Kaufprozess unterscheidet sich vom Lernprozess, das Ergebnis eines Lernprozesses unterscheidet sich vom Ergebnis eines Kaufprozesses. Darüber hinaus bleibt z.B. die Frage offen, ob ein pädagogisch-didaktisches Empfehlungssystem ähnliche Informationen vorschlägt - wie dies z.B. bei *amazon.de* der Fall ist - oder aber im Sinne einer absichtsvollen Irritation abweichende bzw. konträre Informationen. Grundlage eines solchen pädagogischen Empfehlungssystems ist jedoch in jedem Fall die Analyse und Interpretation der Handlungen des Nutzers, des E-Learningprozesses.

Literaturverzeichnis

- [AF86] Abbott, Andrew; Forrest, John: Optimal Matching Methods for Historical Sequences. *Journal of Interdisciplinary History*, 1986, 16, 3; S. 471-494.

- [Ai00] Aisenbrey, Silke: Optimal Matching Analyse: Anwendungen in den Sozialwissenschaften. Leske und Budrich: Opladen, 2000.
- [BM00] Bergmann, Jörg; Meier, Christoph: Elektronische Prozessdaten und ihre Analyse. In (Flick, U.; Kardorff, E. von; Steinke, I., Hrsg.): Qualitative Forschung: Ein Handbuch. Reinbek: Rowohlt, 2000; S. 429-437.
- [BC90] Berners-Lee, Tim; Cailliau, Robert: WorldWideWeb: Proposal for a HyperText Project. <<http://www.w3.org/Proposal.html>>, 1990, (25.02.2007).
- [BR02] Blossfeld, Hans-Peter; Rohwer, Götz: Techniques of event history modeling - new approaches to causal analysis. Erlbaum: Mahwah, NJ., 2002, 2. edition.
- [Eh02] Ehlers, Ulf-Daniel: Qualität beim eLearning. Der Lernende als Grundkategorie der Qualitätssicherung. In (Deutsche Gesellschaft für Erziehungswissenschaft, Hrsg.): MedienPädagogik – Onlinezeitschrift für Theorie und Praxis der Medienbildung. Ausgabe 01/2002. <<http://www.medienpaed.com/02-1/ehlers1.pdf>>, (25.02.2007).
- [Er01] Erzberger, Christian: Sequenzmusteranalyse als fallorientierte Analysestrategie. In (Sackmann, R.; Wingens, M., Hrsg.): Strukturen des Lebenslaufs: Übergang - Sequenz - Verlauf. Weinheim: Juventa, 2001; S. 135-162.
- [Fl70] Flanders, Ned A.: Analyzing teaching behaviour. Addison-Wesley Publishing:. Reading, Mass., 1970.
- [Ha50] Hamming, R. W.: Error-Detecting and Error-Correcting. Bell System Technical Journal, 1950, 2; S. 147-160.
- [Hö27] Hönigswald, Richard: Über die Grundlagen der Pädagogik: ein Beitrag zur Frage des pädagogischen Universitäts-Unterrichts. Reinhardt: München, 1927; 2., überarbeitete Auflage.
- [HBM03] Höök, Kristina; Benyon, David; Munro, Alan J.: Designing Information Spaces: The Social Navigation Approach. Springer: London, 2003.
- [Is02] Iske, Stefan: Vernetztes Wissen: Hypertext-Strategien im Internet. Bertelsmann: Bielefeld, 2002.
- [Is07] Iske, Stefan: Navigationsanalyse. Methodologie der Analyse von Prozessen der Online-Navigation mittels Optimal-Matching. Dissertation, Universität Duisburg-Essen, 2007, in Druck.
- [IS05] Iske, Stefan; Swertz, Christian: Methodologische Fragen der Verwendung von Bild-, Ton- und Textdaten zur Navigationsanalyse. In (Deutsche Gesellschaft für Erziehungswissenschaft, Hrsg.): MedienPädagogik – Onlinezeitschrift für Theorie und Praxis der Medienbildung. Ausgabe 01/2005. <www.medienpaed.com/04-1/iske_swertz04-1.pdf>, (25.02.2007).
- [KS99] Sankoff, D.; Kruskal, J. (Hrsg.): Time warps, string edits, and macromolecules: the theory and practice of sequence comparison. Stanford: CSLI., 2. edition.
- [Ku91] Kuhlen, Rainer: Hypertext. Ein nicht-lineares Medium zwischen Buch und Wissensbank. Springer: Berlin, 1991.
- [La97] Landow, George P.: Hypertext 2.0: Hypertext - the convergence of contemporary critical theory and technology. Johns Hopkins Univ. Press: Baltimore, 1997.
- [Le92] Leutner, Detlev: Adaptive Lehrsysteme - instruktionspsychologische Grundlagen und experimentelle Analysen. Weinheim: Psychologie-Verlags-Union, 1992.
- [Le66] Levenshtein, Vladimir. I.: Binary codes capable of correcting deletions, insertions, and reversals. Cybernetics and Control Theory, 1966, 10, 8; S. 707-710.
- [Me03] Meder, Norbert Grundlagen zu einer Theorie der Didaktik. In (Schlüter, Anne, Hrsg.): Aktuelles und Querliegendes zur Didaktik und Curriculumentwicklung. Bielefeld: Janus Presse, 2003; S. 34-47.
- [Me06] Meder, Norbert: Web-Didaktik: eine neue Didaktik webbasierten vernetzten Lernens. Bertelsmann: Bielefeld, 2006.

- [Pr04] Priemer, Burkhard: Logfile-Analysen: Möglichkeiten und Grenzen ihrer Nutzung bei Untersuchungen zur Mensch-Maschine-Interaktion. In (Deutsche Gesellschaft für Erziehungswissenschaft, Hrsg.): MedienPädagogik – Onlinezeitschrift für Theorie und Praxis der Medienbildung, 2004, . <<http://www.medienpaed.com/04-1/priemer1.pdf>>, (25.02.2007).
- [SWH01] Sackmann, Reinhold; Wiggins, Matthias; Heinz, Walter R. (Hrsg.): Strukturen des Lebenslaufs: Übergang - Sequenz - Verlauf. Weinheim: Juventa-Verlag, 2001.
- [Wi01] Windzio, Michael: Übergänge und Sequenzen. Der Einfluss von Arbeitslosigkeit auf den weiteren Erwerbsverlauf. In (Sackmann, R.; Wiggins, M., Hrsg.): Strukturen des Lebenslaufs: Übergang - Sequenz - Verlauf. Weinheim: Juventa, 2001; S. 163-198.

Abbildungen

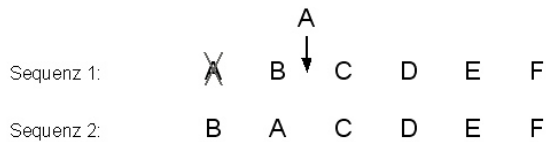


Abbildung 1: Vergleich Sequenz 1 und 2

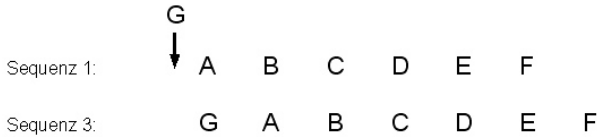


Abbildung 2: Vergleich Sequenz 1 und 3

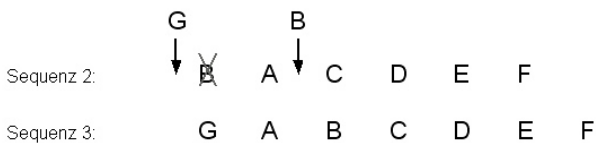


Abbildung 3: Vergleich Sequenz 2 und 3

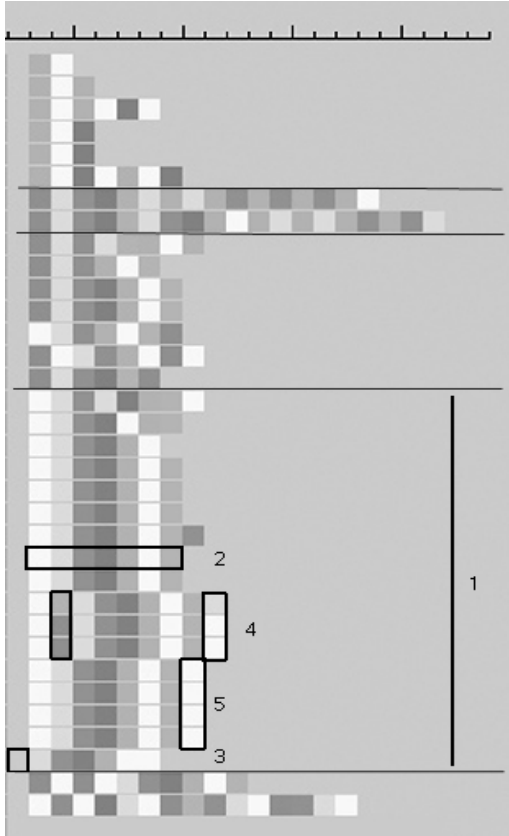


Abbildung 4: Gruppierung ähnlicher Navigationsverläufe mittels Optimal-Matching