

Real-time Site Monitoring

Horst Dietrich

7d aktiengesellschaft
product design + research
oberstrasse 14b, 20144 hamburg, germany
www.7d.net
horst.dietrich@7d.net

Abstrakt

Im Kontext der Personalisierungstechnologie von 7d wird eine Software zum Monitoring der Verhaltensdaten von Websitebesuchern vorgestellt. Im Unterschied zu vielen Analysetools ist der **b-monitor** in der Lage, in real-time qualitative Aussagen über Verhaltensmuster graphisch abzubilden. Das Navigationsverhalten wird als Graphenstruktur aggregiert und vom **b-monitor** in verschiedenen Ansichten für den Anwender visualisiert.

1. b-monitor Visualisierung von Verhaltensdaten

Wesentliches Merkmal der **7d** Personalisierungstechnologie ist die Integration von drei komplementären Ansätzen. Im **7d b-bot** wird die Verknüpfung von inhalts-, verhaltens- und regelbasierten Verfahren realisiert, damit sind die wesentlichen Personalisierungsansätze in einem Technologiekonzept vereint. Das verhaltensbasierte Verfahren ermöglicht die Vorhersage möglicher Bewegungen der Sitebesucher durch eine Site. Vor diesem Hintergrund entstand die Notwendigkeit, das Verhalten der Sitebesucher sichtbar zu machen.

Der **b-monitor** ermöglicht dem Site-Anbieter, Information über die Bewegungsmuster seiner Sitebesucher in Echtzeit zu erlangen. Er besteht aus zwei Modulen: der **Statistic-Engine** und dem eigentlichen **b-monitor**. Um die Analyse und die Darstellung der Verhaltensdaten in Echtzeit zu ermöglichen, werden diese auf getrennten Systemen ausgeführt.

Das Benutzerverhalten wird für die Auswertung in der **Statistic-Engine** aggregiert und analysiert. Diese analysiert im Unterschied zu gängigen Analysetools neben quantitativen Daten (z.B.: Anzahl der Seitenaufrufe) insbesondere qualitative Daten (Bewegungsmuster der Sitebesucher).

Folgende Parameter werden analysiert:

1. Die Wege der Sitebesucher durch eine Website entsprechend ihrem Einstiegspunkt.
2. Die Wege, die Sitebesucher zu einem Dokument hinführen und die Wege, die von diesem wegführen.
3. Die relativen Häufigkeiten der Seitenaufrufe nach den inhaltlichen Kategorien der Website.
4. Die Performance der einzelnen 7d b-bot Module (Anzahl Empfehlungen, Erfolgsfeedback etc.).

Der **b-monitor** dient der eigentlichen visuellen Darstellung. Durch speziell entwickelte Algorithmen wird eine prägnante Darstellung der Daten erreicht. Die Visualisierungstechniken im **b-monitor** erlauben es dem Anwender, intuitiv das Verhalten seiner Sitebesucher zu verstehen und daraus entsprechende Konsequenzen für Sitebetrieb und Optimierung des **7d b-bot** zu ziehen.

2. Datenerhebung

Es wurde ein Verfahren entwickelt, das die Navigationsmuster der Sitebesucher in einer Graphenstruktur abbildet. Die Statistic-Engine zeichnet die Wege der Sitebesucher durch eine Website auf. Dazu werden jeweils die Dokumenten-ID, Session-ID und eine mögliche User-ID in einer Datenbank abgelegt. Die Statistic Engine identifiziert die Wege eines einzelnen Sitebesuchers anhand einer Session-ID und vergleicht dessen Wege mit bereits gegangenen Wegen. Dabei wird an den Knoten der Graphen gezählt, wie oft ein Dokument aufgerufen wurde – an den Kanten, wie oft ein Weg gegangen wurde.

Um die Abbildung der Wege möglichst performant zu gestalten und das Informationsaufkommen zu reduzieren, werden Teile der Baumstruktur, die besonders häufig gegangen werden zu Subsets verschiedener Ordnung zusammengefasst. Dieses Verfahren erlaubt es, mit relativ geringem Datenvolumen die komplexen Wege der Sitebesucher abzubilden.

Ansätze zur Komplexitätsreduzierung von Graphen wurden u.a. von *Eui-Hong (Sam) Han et al. (1997)*, *Karypis und Kumar (1999)* und *Coppersmith et al. (1999)* beschrieben.

Da sich ein einzelnes Dokument in Abhängigkeit vom Navigationsverhalten der Sitebesucher an verschiedenen Positionen im Baum befinden kann, bietet der **b-monitor** zwei Darstellungen des Navigationsbaumes. Einmal wird das einzelne Dokument an seinen jeweiligen Positionen in den lokalen Teilbäumen dargestellt. Oder ein ausgewähltes Dokument wird im Kontext der globalen Wege dargestellt, die zu diesem Dokument hinführen oder von diesem wegführen.

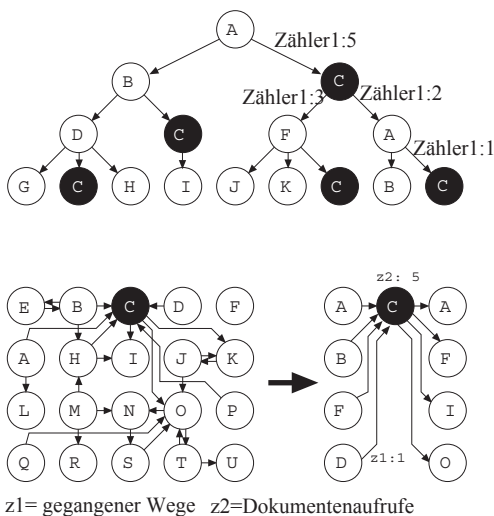


Abbildung 1 Aufzeichnung des Navigationsverhaltens als lokale (oben) und globale Graphen.

Aufzeichnung der Wege lokal

Diese Form der Aufzeichnung zeigt das einzelne Dokument an den jeweilige Positionen im Navigationsbaum. Da der Sitebesucher über verschiedene Einstiegs- punkte auf eine Site gelangen kann, werden auch unterschiedliche Bäume abgebildet. Das bedeutet, das Dokument kann auch in verschiedenen Bäumen an verschiedener Position auftreten.

Aufzeichnung der Wege global

Abgesehen von vom jeweiligen Einstiegs- punkt werden die Dokumente in einer globalen Netzstruktur erfasst. Diese Struktur besteht aus allen mit einem Dokument verknüpften Wegen, unabhängig ihrer Richtung. Für jedes einzelne Dokument können die hin- oder weg- führenden Wege in gerichtete Graphen transformiert werden.

Aufzeichnung der Seitenaufrufe

Zu jeder Dokumenten-ID wird die Anzahl der globalen Seitenaufrufe und die Dokumentenkategorie aufgezeichnet. Dies erlaubt es, Dokumente nach den Themenbereichen zu sortieren und entsprechend abzubilden.

3. GUI b-monitor

Bei der Konzeption des **b-monitor** bestand unser Anspruch darin, dem Anwender vielfältige Information in übersichtlicher Form darzustellen. Ihm sollte ein „visuelles Data Mining“ auf einem durchschnittlichen Desktopmonitor möglich sein.

Abbildung der Wege

Wir entschieden uns, die Wege der Sitebesucher als Graph darzustellen, da diese die Navigationsmuster am anschaulichsten modellieren können. Weil sich bei großen Websites hyperkomplexe Graphen entwickeln können, wurde ein Verfahren gesucht, das diese auf einer endlichen Fläche darstellen kann. Außerdem bestand das Problem der Darstellung verschiedener Einstiegs- punkte in eine Site. Ausgeschlossen war das Clustern von Graphen, weil damit ein nicht akzeptabler Rechenaufwand verbunden wäre.

Bei der Recherche zur visuellen Darstellung großer Graphen fiel auf, dass es eine Unmenge theoretischer Arbeiten sowie einige experimentelle Systeme gibt. Beispiele für experimentelle Systeme sind: das Projekt „daVinci“ der Universität Bremen (Fröhlich, 1995) oder „NicheWorks“ von Wills (1997). Eine Übersicht zu experimentellen Studien findet man bei Dodge (1999) und zu theoretischer Arbeiten bei Herman (1999).

Alle diese Ansätze haben den Nachteil, dass sie wenig Relevanz für einen praktischen Einsatz haben. Hervorstechende Defizite dabei sind: lange Berechnungszeiten, komplizierte Navigation in den Graphen, unübersichtliche Darstellung und schlechte Orientierung und hohe Anforderungen an die Präsentationshardware.

Wir entschieden uns für einen Ansatz, bei dem der Anwender immer nur einen Abschnitt der existierenden Graphen präsentiert bekommt. Der Anwender navigiert dabei durch einen Baum oder kann in Teilbereiche des Baumes springen. Außerdem ist die Breite des Baumes auf die Darstellung relevanter Dokumente beschränkt. Das heißt, bei der Auswahl der visuellen Darstellung entschieden wir uns für gerichtete Graphen. Abbildung 2 zeigt, wie sich ein Anwender im **b-monitor** durch einen gerichteten Graph navigiert. Ungerichtete Graphen bilden zwar die Komplexität der Navigationsmuster ab, aber die Wege der Sitebesucher sind schon ab einer geringen Tiefe der Site nicht mehr sinnvoll darstellbar.

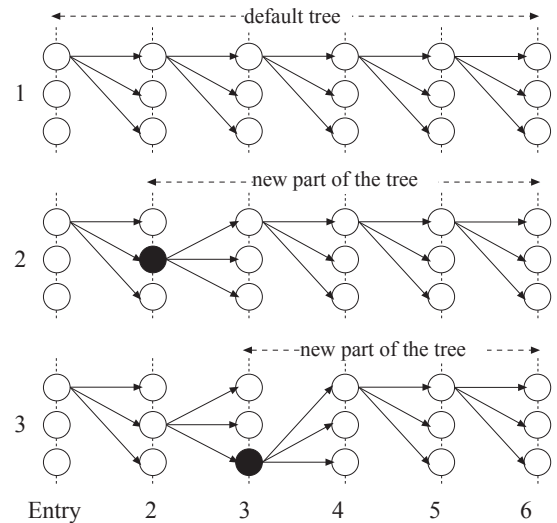


Abbildung 2 Diese Abbildung zeigt schematisch, wie sich im **b-monitor** ein Anwender durch einen Baum mit mehreren Einstiegspunkten navigiert.

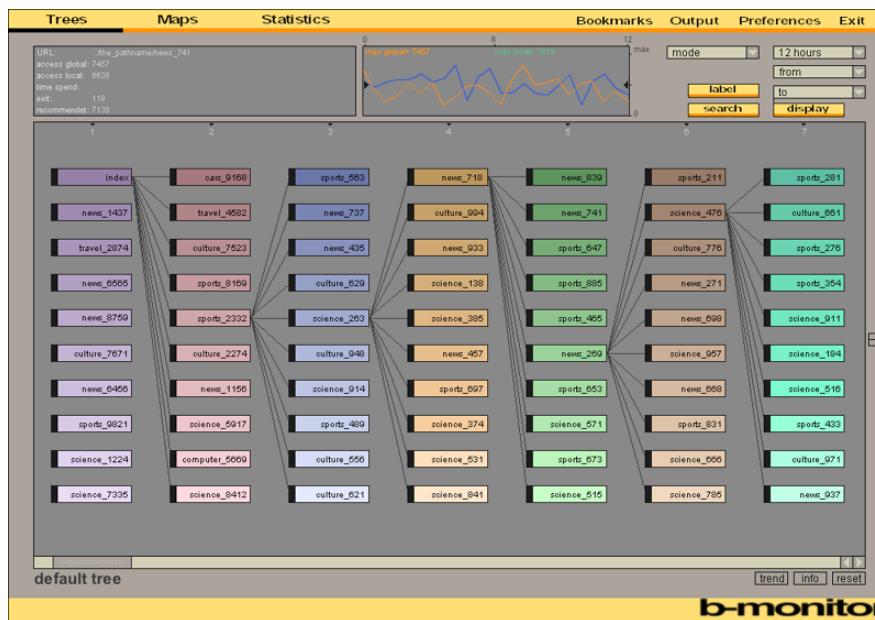


Abbildung 3 Diese Abbildung zeigt die Visualisierung der lokalen Wege im **b-monitor**. Neben der Darstellung eines Dokumentes in der Hierarchie eines Baumes werden die lokalen (an genau dieser Position im Graph) und globalen Seitenaufrufe (über den gesamten Navigationsbaum), die Anzahl der Siteausstiege an dieser Position präsentiert.

diesen Graph bewegen. Bei jedem Click baut sich für jedes Dokument die darunter liegende Hierarchie neu auf. Das hat den Vorteil, dass auch sehr komplexe Navigationsmuster auf einer beschränkten Fläche abgebildet werden.

Das Userinterface für den **b-monitor** wurde so gestaltet, dass der jeweilige Graph in festen Zeilen und Spalten abgebildet wird. In der ersten Spalte werden die Einstiegspunkte in die Site und mit jeder weiteren Spalte die nächst tiefere Hierarchiestufe im Baum dargestellt. Die Abfolge der Zeilen repräsentiert die Rangfolge der Aufrufe nach deren Häufigkeit. Also stehen in der ersten Zeile immer die am häufigsten aufgerufenen Dokumente etc.

Der Anwender kann sich dabei interaktiv durch

Zusätzlich werden quantitative Informationen dargeboten: für jedes Dokument werden die globalen und lokalen Aufrufe und die Anzahl der Siteausstiege präsentiert. Außerdem wird eine Anzeige der globalen und lokalen Häufigkeitsverteilung der Aufrufe angeboten.

Die Abbildung der globalen Wege im **b-monitor** ermöglicht dem Anwender, zu erkennen, auf welchem Weg die Sitebesucher zu einem Dokument gelangt sind und wie sie sich von diesem Dokument weiter bewegt haben. Die Darstellung der globalen Wege präsentiert an dieser Stelle die Gesamtheit der Wege aus allen Teilbäumen.

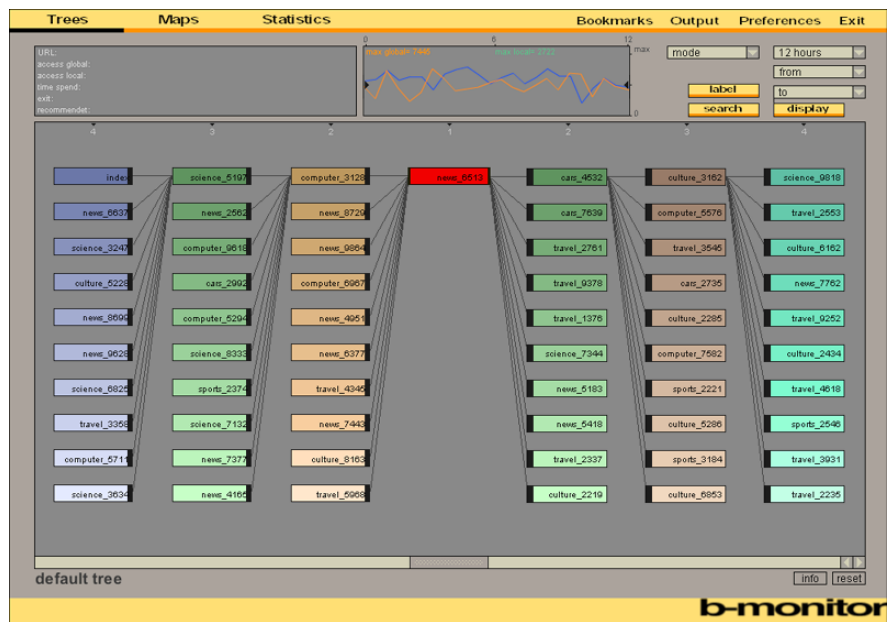


Abbildung 4 Diese Abbildung zeigt die Visualisierung der globalen Wege im **b-monitor**. Es werden die häufigsten Wege zu einem Dokument hin und von einem Dokument weg dargestellt.

Abbildung der Häufigkeitshierarchie

Neben der Abbildung der Wege ergab sich zusätzlich die Forderung an den **b-monitor**, die Seitenaufrufe für Dokumentenkategorien und die jeweiligen Dokumente darzustellen. Dies sollte für eine große Anzahl von Dokumenten möglich sein.



Abbildung 5 Darstellung der relativen Häufigkeiten von Seitenaufrufen und Dokumentenkategorien im **b-monitor**

Für die Darstellung der kategorisierten Aufrufe wurde ein Verfahren eingesetzt, das auf einem Ansatz von *Johnson und Shneiderman (1991)* basiert. Weiterentwickelt wurde dieser Algorithmus von *Brules et.al. (2000)*. Dieser erlaubt, komplexe Hierarchien äußerst kompakt als Treemaps abzubilden.

Bereich aus dem die meisten Dokumente aufgerufen wurden, nimmt dabei die größte Fläche ein. Innerhalb jedes dieser Bereiche werden dann die einzelnen Dokumente wiederum als Fläche unterschiedlicher Größe abgebildet.

Unser Ziel war es, einerseits Themenbereiche und andererseits die Dokumente innerhalb eines Bereiches nach der Häufigkeit der Seitenaufrufe abzubilden. Der

Abbildung der Performancedaten der einzelnen b-bot Module

Im Kontext der gesamten b-bot Technologie dient der **b-monitor** dazu, den Erfolg der jeweiligen Empfehlungen zu kontrollieren. Er bildet die Effizienz ab, mit der die einzelnen Module Empfehlungen generieren und wie diese Empfehlungen von den Sitebesuchern angenommen werden.

4. Performance

Je nach Wunsch und technischer Ausstattung des Anwenders, kann die **Statistic-Engine** in einem Zeittakt von weniger als einer viertel Stunde aktuelle Analysedaten liefern. Die Anzahl der b-monitor Clients, die auf diese Daten zugreifen ist theoretisch nicht begrenzt.

5. Fazit

Mit dem **b-monitor** haben wir ein Tool entwickelt, das bei hohem Bedienkomfort qualitative Aussagen über das Verhalten von Sitebesuchern ermöglicht. Im Unterschied zu herkömmlichen Webstatistiken bietet der **b-monitor** den Vorteil, die Wege der Sitebesucher durch eine Site real-time darzustellen. Die verschiedenen Formen der Visualisierung erleichtern dem Anwender das Auffinden von Zusammenhängen zwischen strukturellen beziehungsweise inhaltlichen Kriterien des Angebotes und dem Verhalten der Sitebesucher.

Ziel dieser Entwicklung war nicht, eine multivariate Datenanalyse anzubieten, sondern zeitnahe, visualisierte Informationen über das Navigationsverhalten der Sitebesucher.

6. Referenzen

- G.di Battista, P.Eades, R.Tamassia and I.G. Tollis: Graph drawing; algorithms for the visualisation of graphs, Prentice Hall, 1999.
- D. Coppersmith, S.J. Hong, and J. Hosking: Partitioning Nominal Attributes in Decision Trees, in Journal of Data Mining and Knowledge Discovery, Volume 3, Number 2, June 1999.
- M. Dodge: Mapping the Net: Some ideas from cartography & geography, presentation at the *ISMA: Network Visualization workshop*, 15-16th April 1999, San Diego (See also the project's current homepage: www.cybergeography.org).
- M. Fröhlich and M. Werner: Demonstration of the interactive graph visualization system daVinci. In: Proceedings of DIMACS Workshop on Graph Drawing '94, R. Tamassia, I. Tollis (eds.), Lecture Notes in Computer Science No.894, Springer Verlag, 1995. (See also the project's current homepage: www.informatik.uni-bremen.de/daVinci/).
- E.H. Han, G. Karypis, V. Kumar, and B. Mobasher: Clustering in a highdimensional space using hypergraph models. In Technical Report, University of Minnesota, Department of Computer Science, 97-019, 1997.
- I. Herman, G. Melançon, M.M. de Ritter and M. Delest: Latour- a tree visualisation system, Information Systems (INS) INS-R9904, April 1999.
- B. Johnson and B. Shneiderman. Treemaps: A space-filling approach to the visualization of hierarchical information structures. In Proceedings of the 2nd International IEEE Visualization Conference, October 1991.
- G. Karypis and V. Kumar: A fast and high quality multilevel scheme for partitioning irregular graphs. SIAM Journal on Scientific Computing, 20(1):359-392, 1999.
- G.J. Wills: Niche Works – interactive visualization of very large graphs. In 5th International Symposium, Graph Drawing '97, Rome, Italy, Lecture Notes in Computer Science 1353, Springer Verlag, 1997
- J.J. van Wijk and H. van de Wetering: Cushion treemaps - visualization of hierarchical information. In Proceedings of IEEE Symposium on Visualization , May 2000.