

**Manuel Kirchner**

# Strategien zur Erkennung von Betrug in der Onlinewerbung

**Diplomarbeit**

## **Bibliografische Information der Deutschen Nationalbibliothek:**

Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de/> abrufbar.

Dieses Werk sowie alle darin enthaltenen einzelnen Beiträge und Abbildungen sind urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsschutz zugelassen ist, bedarf der vorherigen Zustimmung des Verlanges. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen, Auswertungen durch Datenbanken und für die Einspeicherung und Verarbeitung in elektronische Systeme. Alle Rechte, auch die des auszugsweisen Nachdrucks, der fotomechanischen Wiedergabe (einschließlich Mikrokopie) sowie der Auswertung durch Datenbanken oder ähnliche Einrichtungen, vorbehalten.

Copyright © 2005 Diplomica Verlag GmbH  
ISBN: 9783832490713

**Manuel Kirchner**

# **Strategien zur Erkennung von Betrug in der Onlinewerbung**



---

Manuel Kirchner

# Strategien zur Erkennung von Betrug in der Onlinewerbung

Diplomarbeit

Fachhochschule Frankfurt am Main - University of Applied Sciences

Fachbereich Informatik

Abgabe September 2005



***Diplom.de***

Diplomica GmbH \_\_\_\_\_

Hermannstal 119k \_\_\_\_\_

22119 Hamburg \_\_\_\_\_

Fon: 040 / 655 99 20 \_\_\_\_\_

Fax: 040 / 655 99 222 \_\_\_\_\_

agentur@diplom.de \_\_\_\_\_

www.diplom.de \_\_\_\_\_

ID 9071

Kirchner, Manuel: Strategien zur Erkennung von Betrug in der Onlinewerbung

Hamburg: Diplomica GmbH, 2005

Zugl.: Fachhochschule Frankfurt am Main - University of Applied Sciences, Diplomarbeit, 2005

---

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtes.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Die Informationen in diesem Werk wurden mit Sorgfalt erarbeitet. Dennoch können Fehler nicht vollständig ausgeschlossen werden, und die Diplomarbeiten Agentur, die Autoren oder Übersetzer übernehmen keine juristische Verantwortung oder irgendeine Haftung für evtl. verbliebene fehlerhafte Angaben und deren Folgen.

Diplomica GmbH

<http://www.diplom.de>, Hamburg 2005

Printed in Germany

<b>1. Einleitung</b>	<b>1</b>
1.1. Problemstellung	3
1.2. Ziel und Gang der Arbeit	3
<b>2. Grundlagen</b>	<b>5</b>
2.1. Onlinewerbung	5
2.2. Onlinewerbung in Zahlen (Deutschland)	5
2.3. Konzepte der Onlinewerbung	7
2.3.1. Abrechnung nach Seitenimpressionen	7
2.3.2. Pay-Per-Click	7
2.3.2.1. Keyword Cluster	8
2.3.2.2. Berechnung der Rangfolge	8
2.3.2.3. Kosten pro Klick	9
2.3.2.4. Restriktionen	10
2.3.3. Das Google AdSense Programm	10
2.3.4. Affiliate Programme	11
2.4. Klickbetrug	11
2.4.1. Motivation für Klickbetrug	12
2.4.1.1. Motivationen in Konkurrenz Situationen	12
2.4.1.2. Motivation im AdSense Umfeld	13
2.4.2. Bedeutung für die Suchmaschinenbetreiber	13
2.4.3. Humane und automatische Klickbetrüger	15
2.5. Wer kann Klickbetrug erkennen ?	16
<b>3. Strategien</b>	<b>18</b>
3.1. Lokalität von Klickbetrüger und Bots	18
3.2. Bewegungsräume	19
3.2.1. Website des Werbetreibenden	19
3.2.2. Suchergebnisseiten	19
3.2.3. Internet	20

<b>3.3. Das Verfahren im Überblick</b>	<b>21</b>
<b>3.3.1. Rohmaterial Aggregation und Konvertierung</b>	<b>22</b>
<b>3.3.2. Datenbereinigung (Cleaning) / Separation</b>	<b>24</b>
<b>3.3.3. User / Session Identifikation</b>	<b>24</b>
3.3.3.1. Grundlagen	25
3.3.3.2. User Identifikation	26
3.3.3.3. Session Identifikation	28
3.3.3.4. Probleme durch inkonsistente Daten	29
3.3.3.5. Pfad Vervollständigung	29
<b>3.3.4. Mustererkennung</b>	<b>30</b>
<b>3.3.5. Kennzahlen pro Session</b>	<b>30</b>
<b>3.3.6. Auswertung</b>	<b>30</b>
<b>3.3.7. Kennzahlenpool</b>	<b>31</b>
<b>3.4. Erkennungsstrategien</b>	<b>31</b>
<b>3.4.1. Nachweis anhand von betrügerischen Mustern</b>	<b>31</b>
<b>3.4.2. Nachweis anhand von Abweichungen in Kennzahlbereichen</b>	<b>32</b>
3.4.2.1. Strategien zur Datenhaltung	32
3.4.2.2. Vergleichsstrategien	34
<b>4. Kennzahlen und Metriken</b>	<b>35</b>
<b>4.1. Basisdaten</b>	<b>35</b>
<b>4.2. Kategorisierung</b>	<b>36</b>
<b>4.3. Übersicht</b>	<b>37</b>
<b>4.4. Kennzahlen und Metriken im Detail</b>	<b>39</b>
<b>4.4.1. Content</b>	<b>40</b>
<b>4.4.2. Page Impressions (PI)</b>	<b>42</b>
<b>4.4.3. Statistische Daten</b>	<b>46</b>
<b>4.4.4. Graphentheoretische Kennzahlen</b>	<b>47</b>
<b>4.4.5. Gewichtung von Kennzahlen</b>	<b>56</b>



<b>5. Algorithmen gegen Klickbetrug</b>	<b>57</b>
<b>5.1. Graphentheoretische Kennwerte</b>	<b>57</b>
5.1.1. <i>Distanzmatrix Algorithmus</i>	58
5.1.2. <i>Compactness (Cp)</i>	60
5.1.3. <i>Stratum (S)</i>	61
<b>5.2. Betrug durch Data-Mining erkennen</b>	<b>63</b>
5.2.1. <i>Definitionen</i>	63
5.2.2. <i>Verfahren</i>	64
5.2.2.1. Clusteranalysen	64
5.2.2.2. Klassifikation	65
5.2.2.3. Assoziationsregeln	66
5.2.3. <i>Ermittlung häufig besuchter Wege</i>	69
5.2.3.1. Maximal-Forward-Reference (MFR) Algorithmus	70
5.2.3.2. Large-Reference-Sequences (LRS)	73
5.2.3.3. Maximal-Reference-Sequences	77
5.2.3.4. Bewertung	77
<b>6. Zusammenfassung und Ausblick</b>	<b>79</b>
<b>Literaturverzeichnis</b>	<b>81</b>
<b>Abbildungsverzeichnis</b>	<b>83</b>

# 1. Einleitung

Durch die immer stärkere Durchdringung des Internets über alle Altersschichten hinweg ist Onlinewerbung ein wichtiges Medium für die Vermarktung von Werbung geworden. Im World Wide Web lassen sich äusserst zielgruppenspezifische Anzeigen schalten und deren Verbreitung detailliert kontrollieren. Durch Beobachtung des Besucherverhaltens werden dem Verbraucher massgeschneiderte Anzeigen präsentiert. Anzeigen lassen sich genauestens reglementieren, budgetieren und kontrollieren. Sogar die direkte Erfolgskontrolle von Kampagnen im Vergleich zum Absatz ist möglich (Konversionsrate). In dieser Form ist das bei keiner anderen Werbeplattform möglich. Im Internet sind verschiedene Abrechnungssysteme für Onlinewerbung vorhanden. Grob unterschieden wird die Abrechnung nach Seitenimpressionen (Pay-Per-View) und die Abrechnung pro Klick (Pay-Per-Click, PPC, BezahlLink). In dieser Arbeit liegt der Schwerpunkt der Untersuchung auf Betrug im Zusammenhang mit Pay-Per-Click Abrechnungssystemen.

Pay-Per-Click Systeme bieten eine vermeintlich gerechte Art der Abrechnung. Bezahlt werden nur tatsächliche Klicks. Mit der Einführung von AdWords durch den Suchmaschinenbetreiber Google wurde das System perfektioniert und für jedermann leicht einsetzbar. Die monopolartige Stellung der Google Suchmaschine bietet die ideale Werbeplattform. AdWords sind kontextbezogene Werbeanzeigen die zu einer Suchanfrage thematisch passende Anzeigen liefern. Die bestechende Idee dahinter ist, dass der Suchmaschinennutzer immer die passende Werbung zu seiner Suchanfrage (und damit auch seinem Interessengebiet) präsentiert bekommt. Google geht mit seinen AdWords aber auch optisch neue Wege. Die aus reinem Text bestehenden AdWords Anzeigen heben sich angenehm von den vormals dominierenden blinkenden und bunten Bannerwerbungen ab. Für Laien sind sie kaum von normalen Suchergebnissen zu unterscheiden. Der nächste Schritt war die Einführung von AdSense. Bei AdSense ist es Betreibern von Webseiten möglich, mit dem Hosting von AdWords auf ihren Seiten Geld zu verdienen. Der Webseitenbetreiber bindet dabei dynamisch AdWords auf seinen Seiten ein und partizipiert so am AdWords Umsatz von Google. Google analysiert die Seiten und liefert zum Kontext passen-