

# Techniken für Preisvergleiche im World Wide Web

Stefan Kuhlins

Universität Mannheim  
Lehrstuhl für Wirtschaftsinformatik III  
L 5, 6  
68131 Mannheim  
stefan@kuhlins.de

**Zusammenfassung** Im Zuge der stürmischen Entwicklung des Electronic Commerce haben sich unzählige Online-Shops im World Wide Web angesiedelt, um dort ihre Produkte anzubieten. Zum günstigen Erwerb eines Produktes ist es generell sinnvoll, Preisvergleiche anzustellen. Dazu können Programme – beispielsweise Web-Robots und intelligente Softwareagenten – eingesetzt werden, die Preisinformationen zu Produkten von mehreren Online-Shops einholen und dann eine nach Preisen sortierte Übersicht erstellen.

In diesem Beitrag werden nach einer einführenden Motivation für Preisvergleiche im WWW, einige Techniken zur Realisierung vorgestellt. Es wird gezeigt, dass Preisvergleiche nur eine spezielle Anwendung dieser Techniken darstellen, die sich für zahlreiche weitere Anwendungen, welche Informationen mehrerer Web-Sites zusammenfassend auswerten, eignen. Des Weiteren soll ein Überblick über derzeit aktive Preisvergleichsangebote gegeben werden.

## 1 Einleitung

Für homogene Produkte („*Commodity Products*“) wie beispielsweise Musik-CDs, die sich nur durch den Preis unterscheiden, bietet Electronic Commerce prinzipiell ideale Möglichkeiten die Markttransparenz durch automatische Preisvergleiche zu erhöhen. Denn die Daten, die Online-Shops für ihre Kunden im WWW bereitstellen, sind maschinenlesbar und damit einer automatischen Bearbeitung zugänglich. Allerdings *verstehen* Maschinen den Inhalt nicht.

In der Regel werden Preisvergleiche im WWW dadurch realisiert, dass Preisinformationen mit Hilfe von so genannten *Wrappern* mühsam aus HTML-Seiten

---

\* Gemeinsamer Workshop „Sprachen, Architekturen und neue objektorientierte Softwaretechniken (Generatoren, Agenten, Komponenten, Muster)“ der Fachgruppen 2.1.4 und 2.1.9, 7.–9. Mai 2001, Physikzentrum Bad Honnef, <http://www.informatik.uni-kiel.de/~wg/Honnef2001/Program.html>

von Online-Shops extrahiert werden. Im Allgemeinen benötigt man dazu für jeden Online-Shop einen spezifischen Wrapper. Wenn sich das Layout der Web-Site ändert, ist der zugehörige Wrapper anzupassen.

Die Qualität der so extrahierten Preisinformationen gibt oft Anlass zur Kritik. Deshalb gehen immer mehr Online-Shops dazu über, ihre Preisinformationen zusammen mit Angaben über den angebotenen Service (Lieferbarkeit, Transportkosten usw.) zum Abruf bereitzustellen. Da sich bisher jedoch kein Standardformat dafür etabliert hat, ist dies erst ein kleiner Schritt in die richtige Richtung, der es bekannten Geschäftspartnern erlaubt, ihre Daten auszutauschen. Für eine vollständige Automatisierung des Vorgangs werden standardisierte Schnittstellen benötigt, so dass die relevanten Daten und Funktionen in einheitlicher Weise vorliegen.

Dieser Beitrag ist folgendermaßen strukturiert: Kapitel 2 beschäftigt sich mit der Motivation für Preisvergleiche im WWW. Im dritten Kapitel wird der erste Preisvergleichsagent vorgestellt. Anschließend wird die zugrunde liegende Technik in Kapitel 4 erläutert. Ein praktischer Test heute aktiver Preisvergleichsagenten ist Gegenstand des fünften Kapitels. Abschließend folgen Zusammenfassung und Ausblick.

## 2 Motivation für Preisvergleiche im WWW

In einem Special Report von *Ernst & Young* [5], der auf einer im Oktober/November 2000 durchgeführten Online-Umfrage, an der 7.222 Konsumenten aus zwölf Ländern teilnahmen, basiert, heißt es: *“Commodity products are still the top sellers.”* Beispielsweise haben von den befragten Deutschen 70 % Bücher, 60 % Computerprodukte und 53 % Software- sowie Musik-CDs mindestens einmal im WWW eingekauft.

Als Hauptgrund für einen Online-Einkauf wurde das Sparen von Zeit und Geld angegeben, weil für die Kunden der Weg ins Geschäft entfällt und weil die Unternehmen aufgrund elektronischer Bestellungen Abwicklungskosten sparen, was bei den Kunden wiederum die Erwartung nach günstigeren Konditionen im Vergleich zum physischen Einkauf im Geschäft weckt. Als wichtigste Ursachen für den Abbruch von Einkaufstransaktionen nannten die Befragten, dass die Lieferkosten zu hoch seien (40 %), dass der Preis zu hoch sei (35 %) oder dass sie den Preis nur prüfen wollten (33 %).

*Commodity Products* unterscheiden sich per definitionem nur im Preis. Daher können Preisvergleiche für sie verhältnismäßig einfach realisiert werden; für Handy-Tarife beispielsweise ist dies schwieriger. Da *Commodity Products* die beliebtesten Produkte für den Einkauf im WWW darstellen, besteht potenziell eine große Nachfrage nach Preisvergleichen für sie. Die Preissensibilität der Online-Kunden verstärkt diese Nachfrage zusätzlich.

Der hohe Stellenwert der Lieferkosten beim Abbruch von Einkaufstransaktionen zeigt, dass der Produktpreis allein nicht entscheidend ist. Es sollten möglichst alle mit einem Kauf verbundenen Kosten berücksichtigt werden.

Mit einem Browser können unzählige Online-Shops im WWW bequem vom heimischen PC aus besucht werden. Für ein begehrtes Produkt können so in verhältnismäßig kurzer Zeit viele Angebote eingeholt werden. Verglichen mit der Situation vor Einführung des Electronic Commerce sinken dadurch die Kosten für die Suche nach einem günstigen Anbieter. Aber auch das Finden möglichst vieler Online-Shops, das Laden ihrer Seiten und die Suche nach dem gewünschten Produkt sind mit einem erheblichen Zeitaufwand und entsprechenden Kosten verbunden. Diese stumpfsinnige Tätigkeit sollte daher automatisiert werden.

### 3 Der erste Preisvergleichsagent

Der erste Softwareagent für automatisierte Preisvergleiche war *BargainFinder*, der am 30. Juni 1995 der Öffentlichkeit vorgestellt wurde [7]. *BargainFinder* konnte für Musik-CDs, die durch Interpret und Titel identifiziert werden, bei rund einem Dutzend Anbietern – darunter *CD Universe*, *CDNow!*, *CD Land* und *CDworld* – die Preise ermitteln.

Einige Händler waren mit den extrahierten Daten unzufrieden, weil der Service, den sie ihren Kunden boten, nicht gewürdigt wurde. Im reinen Preisvergleich schnitten sie beispielsweise schlechter ab, weil Kosten für die Lagerhaltung zur Erzielung kurzer Lieferzeiten zu höheren Preisen führen. Aus diesem Grund wurde *BargainFinder* von einigen Händlern (u. a. *CDNow!* und *CD Land*) boykottiert. Andere Händler dagegen baten darum, berücksichtigt zu werden; und auch *CD Land* hob die Blockade später wieder auf.

Technisch konnte *BargainFinder* gezielt boykottiert werden, weil die Anfragen immer von demselben Server (mit derselben IP-Adresse) ausgingen. Bei einer Client-basierten Realisierung würden die Anfragen demgegenüber vom Kunden-PC ausgehen und wären von den üblichen Kundenanfragen nicht zu unterscheiden. Aus technischer Sicht haben Online-Shops daher keine Möglichkeit, automatisierte Preisvergleiche zu verhindern, ohne gleichzeitig ihre potenzielle Kundschaft zu treffen.

### 4 Die zugrunde liegende Technik

Die Informationen, die Online-Shops für ihre Kunden bereitstellen, liegen in maschinenlesbarer Form vor, damit Browser sie anzeigen können. Deshalb ist es prinzipiell möglich, Preise automatisch zu extrahieren. Allerdings werden Online-Shops für die Benutzung durch Menschen entworfen. Insbesondere HTML-Seiten bieten zwar ein nett anzusehendes Layout, stellen aber kaum maschinenverwertbare Informationen darüber zur Verfügung, was die dargestellten Daten bedeuten. Mit welchen Problemen Softwareagenten wie *BargainFinder* zu kämpfen haben, soll das folgende Beispiel, Preisermittlung für die Digitalkamera *Sony DSC-P1* bei *Imaging One* [6], verdeutlichen.

Zunächst muss innerhalb eines Online-Shops die passende Seite zu einem Produkt gefunden werden. Kleine Shops offerieren ihre Produkte oft nur auf

einer einzigen Seite. Größere Shops benutzen hingegen in der Regel einen hierarchisch aufgebauten Produktkatalog. Für einen schnelleren Zugang gibt es meistens Suchformulare. Diese stellen auch einen guten Einstiegspunkt für Softwareagenten dar.

Artikelbeschreibung	Bestell-Nr.	Preis in DM
<b>Sony Digitalkamera DSC-P1</b> <a href="#">ab Lager lieferbar</a> Kamera inkl. InfoLithium-Akku NP-FS11, Netz-/Ladegerät, 8 MByte Memory Stick, Videoanschlusskabel, USB-Anschlusskabel, Trageschleife, Bildbearbeitungssoftware MGI PhotoSuite 8.1 für Windows (3.1/95/98/98SE/2000/NT), Bildbearbeitungssoftware MGI PhotoSuite SE 1.1 für Macintosh, Moviesoftware VideoWave SE + für Windows (95/98/98SE/2000/NT), USB-Treiber für Windows (98/2000) und Macintosh – Auflösung 2.048 x 1.536 Bildpunkte – CCD-Sensor mit 3.340.000 Pixeln – 3-fach Zoomobjektiv 39 bis 117 mm (entspr. Kleinbild) – eingebauter 1,5"-LCD-Monitor – Videoausgang (PAL/NTSC umschaltbar) – USB-Schnittstelle – Memory Stick Wechselspeicher – 12 Monate Herstellergarantie <a href="#">ausführliche Technische Daten</a>	101 150	1.829,-
<b>Sony Akku NP-FS11</b> <a href="#">ab Lager lieferbar</a> für DSC-F55 und DSC-F505(V); 1.000 mAh,	100 773	159,-

**Abbildung 1.** Ausschnitt der Browseransicht bei *Imaging One* [6]

Wenn die Seite mit den Produktinformationen gefunden wurde, sind die Preisinformationen zu extrahieren. Bei *Imaging One* handelt es sich um eine HTML-Seite, die aus mehreren geschachtelten Tabellen besteht (siehe Abb. 1). Kamerabezeichnung, Lieferbarkeit und der Preis sind für das menschliche Auge übersichtlich angeordnet und leicht zu erfassen. Der zugehörige HTML-Code mutet indessen chaotisch an (siehe Abb. 2). Die gesamte HTML-Datei umfasst 625 Zeilen, deshalb wird in Abb. 2 nur der für die Preisermittlung relevante Ausschnitt abgebildet.

Das dargestellte Codefragment enthält einen Fehler: Ein schließendes Font-Tag ist überflüssig. Für die komplette Datei meldet *HTML Tidy* [8] 125 Fehler und Warnungen. HTML-Browser sind sehr tolerant, was die Darstellung von fehlerhaften HTML-Seiten angeht. Infolgedessen müssen Softwareagenten ebenfalls fehlertolerant programmiert werden, wodurch sich der Implementierungsaufwand erhöht.

Zum Extrahieren von Informationen aus HTML-Dateien kann man reguläre Suchausdrücke verwenden. Für das Beispiel in Abb. 2 liefert der reguläre Ausdruck `<a name="\([0-9]+\)"></a>.*<br>\n *\.*)</strong>` in \1 (erster geklammerter Ausdruck) die Bestellnummer und in \2 die Artikelbezeichnung. Mit ähnlichen Ausdrücken können Preis und Lieferbarkeit ermittelt werden. Der



Verarbeitung liefern, werden *Wrapper* genannt. Mit Hilfe der Wrapper-Technik (siehe Abb. 3) können Daten aus unterschiedlichen Ressourcen abgefragt und durch einen *Mediator* integriert werden [11].

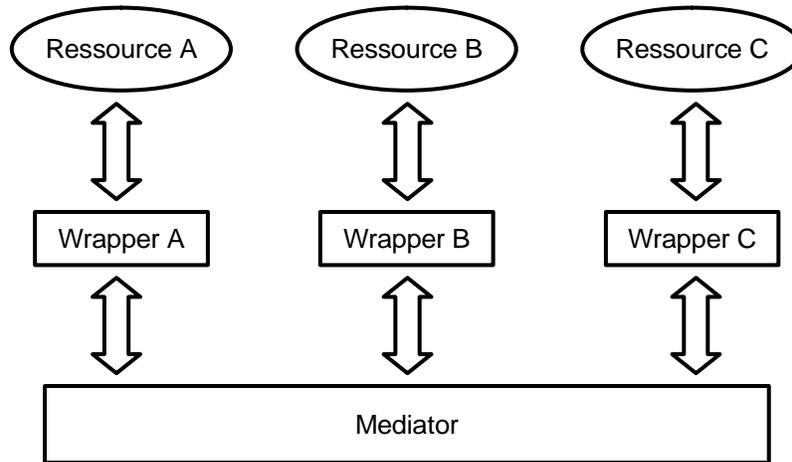


Abbildung 3. Wrapper-Technik

Die Wrapper für *BargainFinder* waren handcodiert. Das größte Problem handcodierter Wrapper ist der enorme Aufwand zur Entwicklung und Wartung. Da jeder Online-Shop sein eigenes Layout besitzt, braucht man für jeden Shop einen speziellen Wrapper. Ändert sich das Layout eines Shops, ist auch der zugehörige Wrapper anzupassen. Der Einsatz von Werkzeugen zur Wrapper-Generierung (beispielsweise W4F [9]) beschleunigt zwar die Entwicklung von Wrappern, ändert aber am grundsätzlichen Problem nichts.

Mit *ShopBot* wurde demzufolge versucht, ein System zu entwickeln, das automatisch Wrapper zum Extrahieren von Preisen generiert [4]. Dazu verfügte *ShopBot* über einen Lernalgorithmus, der typische Eigenschaften von Online-Shops, wie das Vorhandensein von Formularen zur Suche nach Produkten und die zeilenorientierte Präsentation von Produkten, ausnutzte. Obwohl erste Tests ansprechende Ergebnisse lieferten, wurde eine vollständige Automatisierung nicht erreicht, was u. a. durch folgende Fußnote zum Ausdruck gebracht wurde [4]: “Prices are extracted using special hand-coded techniques.”

*BargainFinder* und *ShopBot* sind die einzigen Softwareagenten für Preisvergleiche, deren Arbeitsweise in öffentlich zugänglichen Artikeln dokumentiert wurde. Die Arbeitsweise heutiger aktiver Systeme wird aus Wettbewerbsgründen nicht offen gelegt. Eine Liste mit deutschen und internationalen Anbietern für Preisvergleiche befindet sich unter [1].

## 5 Heute aktive Preisvergleichsagenten

Ein am 1. März 2001 in den USA durchgeführter praktischer Test zeigt, dass sich mit Hilfe von Preisvergleichen viel Geld sparen lässt. Für die Digitalkamera *DSC-P1* gibt der Hersteller *Sony* eine unverbindliche Preisempfehlung von \$ 799,95 an. Zu diesem Preis wird das Produkt von einigen Händlern auch angeboten. Preisvergleiche mit *DealTime*, *mySimon*, *BizRate*, *GoTo Shopping*, *PriceGrabber* und *PriceSCAN* bringen dagegen Preise von unter \$ 640 hervor, was einer Ersparnis von rund \$ 160 oder 20 % entspricht (siehe Tab. 1). Die zweite Spalte der Tabelle 1 enthält die Anzahl der Angebote („Treffer“), die beim Preisvergleich aufgelistet werden. In der dritten Spalte steht der günstigste Preis. Sofern man die Qualität der Anbieter von Preisvergleichen nur nach dem von ihnen aufgespürten günstigsten Preis bewertet, unterscheiden sie sich mit einer maximalen Preisdifferenz von \$ 2 kaum nennenswert.

	Anzahl	\$-Preis	Lieferbar	Lieferkosten	Rating
DealTime.com	39	637,99	X	– (selten)	** Gómez
mySimon.com	41	637,99	–	–	** Gómez
BizRate.com	19	638,00	X	–	8,2
GoTo Shopping	22	639,90	X	\$ 0,25	–
PriceGrabber.com	44	639,99	X	ca. \$ 38,40	Not Rated
PriceSCAN.com	36	639,99	See Site	See Site	–

**Tabelle 1.** Preisvergleich in den USA

Für die Auswahl eines Händlers, bei dem ein Produkt erworben werden soll, sind neben dem Preis weitere Faktoren von Bedeutung. Bei einigen Preisvergleichen werden deshalb die Lieferbarkeit (vierte Spalte in Tab. 1), Lieferkosten und Händler-Ratings, die gewisse Rückschlüsse auf die Qualität von Händlern ermöglichen sollen, angegeben. Während die Lieferbarkeit immerhin bei vier von sechs Preisvergleichen aufgeführt wird, fehlen Angaben über die Lieferkosten weitestgehend. Vermutlich ist dies darauf zurückzuführen, dass der Aufwand zum Extrahieren dieser Informationen den der einfachen Preissuche noch erheblich übersteigt. Die Kunden müssen diese Daten deshalb selbst recherchieren.

Für Deutschland ergeben sich ähnliche Ergebnisse. So brachte ein Preisvergleich mit *DealTime Deutschland* [3] einen Preis von 1.574 DM bei insgesamt 36 Treffern hervor (siehe Abb. 4). Wieder beträgt die Ersparnis gegenüber Sonys unverbindlicher Preisempfehlung von 1.999 DM rund 20 %. Angaben über Lieferbarkeit und Lieferkosten fehlen allerdings ebenso wie ein Händler-Rating.

Interessanterweise kostet die Kamera beim günstigsten Händler in Deutschland umgerechnet um \$ 745, also über \$ 100 mehr als in den USA. Diese Preisdifferenz von rund 15 % ist aber wohl nicht groß genug, damit sich in diesem Fall ein Einkauf bei einem amerikanischen Händler rentiert. Trotzdem zeigt das Bei-

spiel, das Preisvergleiche über Ländergrenzen hinweg sinnvoll sein können, wenn dabei Wechselkurse, Transportkosten, Zollgebühren usw. eingerechnet werden.

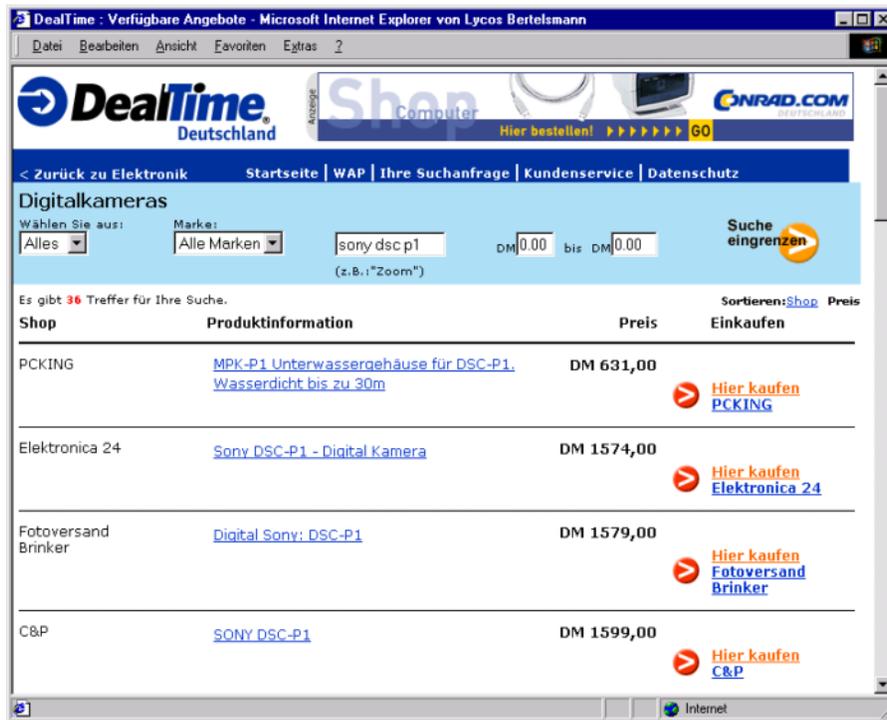


Abbildung 4. DealTime Deutschland [3]

## 6 Zusammenfassung und Ausblick

Die genannten kommerziellen Systeme für Preisvergleiche sind aus Wettbewerbsgründen kaum öffentlich dokumentiert. Es ist aber davon auszugehen, dass viele auf Web-Robots und Wrappern basieren, also ähnlich wie *BargainFinder* arbeiten. Wenn man sich vor Augen hält, dass die meisten Online-Shops ihre Produktdaten in Datenbanken speichern und Eingabeformulare zur Suche nach bestimmten Produkten anbieten, wobei als Antwort auf derartige Abfragen HTML-Seiten mit den Datenbankinhalten automatisch generiert werden, dann dienen Wrapper, die diese Produktdaten aus den generierten HTML-Seiten extrahieren, einer Art Reverse-Engineering von Datenbankschnittstellen in HTML. Insgesamt handelt es sich also um einen äußerst ineffizienten Prozess.

Es gibt Online-Shops, die mit Anbietern von Preisvergleichen kooperieren, indem sie ihre Produkt- und Preisdaten auszugsweise (in komprimierter Form)

zum Download zur Verfügung stellen, wobei der von ihnen gebotene Kundenservice Berücksichtigung finden kann. Dazu werden in der Regel jedoch individuelle Formate verwendet, so dass auch hier geeignete Wrapper zur Verarbeitung benötigt werden. Allerdings ist die Erstellung solcher Wrapper einfacher und der Wartungsaufwand entfällt weitestgehend, weil eine Änderung des Layouts der Web-Site sich nicht auf das Format der Datei mit den Produktdaten auswirkt. Mit steigendem Produktangebot nimmt auch die Dateigröße zu, infolgedessen ist diese Art der Datenbereitstellung nur für kleine Online-Shops geeignet.

Wenn Online-Shops kooperieren, eröffnen sich aber auch bessere Möglichkeiten als ein Dateidownload. Zur Ermittlung von Produkt- und Preisdaten wäre es am besten, wenn Online-Shops eine einheitliche Schnittstelle für Abfragen zur Verfügung stellen würden, wobei die Informationen zum jeweils gebotenen Kundenservice berücksichtigt würden. Dann könnten Softwareagenten für ein bestimmtes Produkt alle für einen Kauf relevanten Daten analog zu einfachen Datenbankabfragen ermitteln. In Verbindung mit einem Verzeichnis, in dem sämtliche Online-Shops aufgeführt werden, könnten so umfassende Preisübersichten erstellt werden.

Für den direkten Weg zu den Produkt- und Preisdaten wäre es allerdings unumgänglich, dass Online-Shops geeignete Schnittstellen für entsprechende Abfragemöglichkeiten installieren. Wenn eines Tages die meisten Kunden Preisvergleiche im WWW nutzen, könnten die Shops dadurch faktisch gezwungen sein, mitzumachen, weil die Umsätze nicht gelisteter Shops rapide zurückgehen. Um einem ruinösen Preiswettbewerb zu entrinnen, werden die Shops in Bezug auf *Commodity Products* versuchen, sich über den Kundenservice zu differenzieren, was letztendlich auch im Interesse der Kunden ist. Die Kunden können dann selbst entscheiden, wie viel ihnen beispielsweise eine kürzere Lieferzeit wert ist.

Bisher decken die Anbieter von Preisvergleichen jeweils nur einen Teilbereich des relevanten Marktes ab. Analog zu Meta-Suchmaschinen, die mehrere Suchmaschinen abfragen und deren Ergebnisse zusammenfassen, wäre daher ein Meta-Preisvergleich denkbar. Ob die einzelnen Anbieter damit einverstanden sind, dürfte von ihrem Finanzierungsmodell abhängen. Außerdem wäre zu klären, ob ein Meta-Preisvergleich rechtlich überhaupt zulässig ist, sofern einzelne Anbieter, die abgefragt werden, damit nicht einverstanden sind.

Einige Anbieter von Preisvergleichen unterstützen die Abfrage mittels mobiler Geräte (z. B. Handys) unabhängig vom Aufenthaltsort. Die Verhandlungsposition der Kunden im „Geschäft um die Ecke“ wird damit erheblich gestärkt. Ein erster Vertreter dieser Art mobiler Preisvergleiche ist *Pocket BargainFinder* [2].

Zusammenfassend kann festgehalten werden, dass Preisvergleiche im WWW ein interessantes Anwendungsgebiet für Softwareagenten darstellen und dass die zugrunde liegenden Techniken für zahlreiche Anwendungen, bei denen Informationen aus mehreren Quellen integriert werden, nutzbar sind.

## Literatur

1. Anbieter von Preisvergleichen im WWW,  
<http://www.wifo.uni-mannheim.de/~kuhlins/preisvergleich/>
2. Brody, Adam B. und Edward J. Gottsmann: Pocket BargainFinder – A Handheld Device for Augmented Commerce, in: Gellersen, Hans-W. (Hrsg.): Handheld and Ubiquitous Computing, Springer-Verlag, Berlin 1999, S. 44–51,  
<http://www.accenture.com/xdoc/en/services/cstar/publications/PocketBargainFinder-HUC99.PDF>
3. DealTime Deutschland: <http://www.dealtime.de/>
4. Doorenbos, Robert B., Oren Etzioni und Daniel S. Weld: A scalable comparison-shopping agent for the World-Wide Web, in: Proceedings of the International Conference on Autonomous Agents, Marina del Rey, CA, 1997, S. 39–48,  
<ftp://ftp.cs.washington.edu/pub/etzioni/softbots/agents97.ps>
5. Ernst & Young: Global Online Retailing 2001, Special Report,  
[http://www.ey.com/global/vault.nsf/US/2001\\_Retail\\_Study/\\$file/GOR.pdf](http://www.ey.com/global/vault.nsf/US/2001_Retail_Study/$file/GOR.pdf)
6. Imaging One: Digitalkamera *Sony DSC-P1*,  
<http://www.imaging-one.de/Kameras/SonyDSC-P1.htm> (abgerufen am 1.3.2001)
7. Krulwich, Bruce: Bruce Krulwich's ideas page,  
<http://www.geocities.com/ResearchTriangle/9430/>
8. Raggett, Dave: HTML Tidy, 30. April 2000,  
<http://www.w3.org/People/Raggett/tidy/>
9. Sahuguet, Arnaud und Fabien Azavant: W4F (WysiWyg Web Wrapper Factory),  
<http://db.cis.upenn.edu/W4F/>
10. W3C: Document Object Model (DOM), <http://www.w3.org/DOM/>
11. Wiederhold, Gio: Mediators in the Architecture of Future Information Systems, IEEE Computer, 25(3):38–49, 1992,  
<http://www-db.stanford.edu/pub/gio/1991/afis.ps>