

# Der shopinfo.xml-Standard: Datenintegration zwischen Shops und Shopping-Portalen

Torsten Wolter, Jens Schumacher, Markus Matschi, Stefan Kuhlins

Hochschule Heilbronn  
Studiengang Electronic Business  
Max-Planck-Str. 39  
74081 Heilbronn  
kuhlins@hs-heilbronn.de

**Abstract:** Shopping-Portale bieten Nachfragern Orientierungshilfen beim Online-Einkauf. Besonders nützlich sind Produktsuchmaschinen, die zu einem vorgegebenen Produkt aufzeigen, welche Online-Shops es zu welchen Konditionen anbieten. Die Produktdaten stellen die Shops den Portalen dabei zur Verfügung. Um eine möglichst große Reichweite zu erzielen, registrieren sich die Shops bei vielen Portalen. Die Registrierung und der anschließende Austausch von Produktdaten werden jeweils individuell geregelt. Im Ergebnis bereiten Shopbetreiber ihre Daten daher in vielen verschiedenen Formaten auf. Der im Folgenden vorgestellte so genannte shopinfo.xml-Standard soll den Datenaustausch vereinheitlichen und somit entscheidend vereinfachen. Erstmals wird hier exemplarisch der Zugriff auf die Daten mit PHP 5 beschrieben.

## 1 Der shopinfo.xml-Standard

### 1.1 Prinzip

Um eine möglichst große Reichweite zu erzielen, registrieren sich Online-Shops bei vielen Portalen. Die Registrierung und der anschließende Austausch von Produkt- und Shopdaten werden jeweils individuell geregelt. Im Ergebnis bereiten Shopbetreiber ihre Produktdaten daher in vielen verschiedenen Formaten auf und hinterlassen ihre Shopdaten bei jeder Registrierung erneut bei Shopping-Portalen (siehe Abbildung 1a). Dadurch steigt mit der Zeit die Gefahr der Inkonsistenz der redundant hinterlegten Daten.

Der shopinfo.xml-Standard ist Teil eines an der Universität Mannheim gestarteten Forschungsprojekts zur Automatisierung von Angebotsrecherchen im Electronic Commerce durch Standardisierung des Datenaustauschs zwischen Online-Shops und Shopping-Portalen. Kern des Standards ist die so genannte „Shopdatei“. In ihr werden die Stammdaten eines Online-Shops gespeichert. Es handelt sich dabei um eine XML-Datei namens shopinfo.xml, die üblicherweise im Hauptverzeichnis einer Shop-Website abgelegt wird.

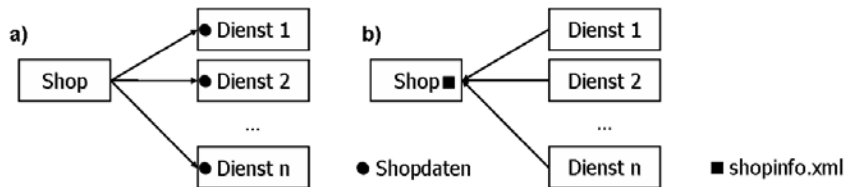


Abbildung 1: Registrierung von Online-Shops a) ohne und b) mit shopinfo.xml-Datei

Der Aufbau von Shopdateien ist durch ein XML-Schema definiert, was die vollautomatische Auswertung der Inhalte mit frei verfügbaren XML-Parsern ermöglicht [HKS02]. Zum einfachen Erstellen von Shopdateien können Shopbetreiber ein HTML-Formular unter <http://elektronischer-markt.de/> nutzen. Hier werden die benötigten Daten zunächst erfasst und am Ende wird eine Shopdatei generiert, die auf der Website des Shops abzulegen ist. Auf diese Weise stehen die Daten beim Shop, der sie bei Bedarf aktualisieren kann (siehe Abbildung 1b). Die Shopdatei beschreibt auch, in welcher Form der Zugriff auf die Produktdaten eines Shops erfolgt. Dazu existieren im Wesentlichen zwei Möglichkeiten, die kombinierbar sind: eine regelmäßig aktualisierte Produktdatei, die zum Download bereitsteht, und eine Schnittstelle, die spezielle Produkthanfragen in Echtzeit beantwortet.

Produktdateien lassen sich so generieren, wie es für einen Shop am einfachsten ist. Welches Format benutzt wird und wie die Inhalte zu interpretieren sind, beschreibt die Shopdatei. Typischerweise wird die Produktdatei täglich aktualisiert. Welche Produkte in der Produktdatei stehen, bestimmt der Shop.

Eine Echtzeitabfrageschnittstelle basiert auf dem gleichen Mechanismus wie die Bearbeitung von HTML-Formularen. Somit können dazu CGI-Skripte, Java Servlets, Active Server Pages, PHP u. Ä. eingesetzt werden. Suchanfragen nach Produktdaten treffen wie üblich mittels HTTP als GET- oder POST-Request ein. Die dabei verwendeten Parameter deklariert die Shopdatei. Als Antwort ist ein XML-Dokument zu generieren, für das ein XML-Schema vorgegeben ist. Es muss also lediglich eine XML-Schablone mit den gesuchten Produktdaten gefüllt werden.

Die eingesetzte Technik wird als REST (*representational state transfer*) bezeichnet. „*In the REST model, XML exchanges are treated in a very web-like way, using HTTP methods (GET, PUT, POST, DELETE) as verbs, XML documents as messages, and URIs to identify the services. ... REST is simple, straightforward, and gets the job done with minimal effort.*“ [HM04]

## 1.2 Spezifikation

Die Shopdateien shopinfo.xml basieren auf einem XML-Schema, welches unter <http://elektronischer-markt.de/schema/shop.xsd> zu finden ist. Listing 1 zeigt ein sehr einfaches Beispiel für eine Shopdatei, die lediglich einen Zugriff auf die Produktdaten mittels Echtzeitschnittstelle (siehe Element <OnlineRequest>) enthält.

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<osp:Shop xmlns:osp="http://elektronischer-markt.de/schema" ... >
  <Common> ... </Common>
  <Name>Test Shop</Name>
  <Url>http://testshop.de/</Url>
  <Requests>
    <OnlineRequest>
      <Processor>http://testshop.de/servlets/search</Processor>
      <ParamProduct>product</ParamProduct>
    </OnlineRequest>
  </Requests>
  <Categories>
    <Item>
      <Name>Bücher</Name>
      <Mapping>Books</Mapping>
    </Item>
  </Categories>
</osp:Shop>

```

Listing 1: Einfaches Beispiel einer shopinfo.xml

Das Element OnlineRequest deklariert den Ort der Echtzeitabfrage und die zur Verfügung stehenden Parameter zur Produktsuche.

## 2 Szenarien der Datenintegration

### 2.1 Integration der Shopdaten

Die Integration der Shopdaten in die Systeme der Shopping-Portale erfolgt über plattformsspezifische XML-Schnittstellen. Um die Daten integrieren zu können, bedarf es i. d. R. nur weniger Schritte – ganz gleich welche Plattform genutzt wird. Nachfolgend wird die Integration exemplarisch mit der Programmiersprache PHP 5 dargestellt.

**Abrufen der shopinfo.xml eines Shops.** Meldet sich ein Shop bei einem Shopping-Portal an, hinterlegt der Shopbetreiber dort die URL zu seiner shopinfo.xml. (Alternativ kann auch das Portal nach shopinfo.xml-Dateien suchen.) In der Regel ist die Datei im Hauptverzeichnis des Webservers abgelegt, also z. B. `http://testshop.de/shopinfo.xml`. Mittels der Erweiterung SimpleXML [Ph06] der Programmiersprache PHP 5 gestaltet sich der Abruf der Datei bspw. einfach so:

```

$url = 'http://testshop.de/shopinfo.xml';
$shop = simplexml_load_file($url);

```

Listing 2: Abruf der shopinfo.xml mittels SimpleXML in PHP 5

**Zugriff und Verwendung der Daten im eigenen System.** Der Zugriff auf die Daten der shopinfo.xml erfolgt im Beispiel über das erzeugte shop-Objekt, z. B.:

```

$shopname = (string) $shop->Name;
$purl = (string) $shop->Requests->OnlineRequest->Processor;

```

Listing 3: Zugriff auf die Informationen der shopinfo.xml in PHP 5

## 2.2 Integration der Produktdaten

Die Integration der Produktdaten lässt sich über eine Produktdatei oder über eine Echtzeitanfrage realisieren.

**Integration der Produktdaten über eine Produktdatei.** Die URL der Produktdatei wird analog zur URL der Echtzeitschnittstelle ermittelt. Der Aufbau der Produktdatei ist im Format-Element der shopinfo.xml hinterlegt.

```
<Requests>
  <OfflineRequest>
    <Format>
      <Tabular>
        <CSV>
          <Url>http://testshop.de/Preisliste.csv</Url>
          <Header columns="3">'Artikelnummer';'Preis';'Deeplink'</Header>
          <SpecialCharacters delimiter=";" escaped=""quoted=""/>
        </CSV>
        <Mappings>
          <Mapping column="1" columnName="Artikelnummer" type="privateid"/>
          <Mapping column="2" columnName="Preis" type="price"/>
          <Mapping column="3" columnName="Deeplink" type="url"/>
        </Mappings>
      </Tabular>
    </Format>
  </OfflineRequest>
</Requests>
```

Listing 4: Deklaration des Produktdateizugriffs – Auszug aus der shopinfo.xml

Für Produktdateien ist das textbasierte CSV-Format (*comma separated values*) bei Online-Shops weit verbreitet, welches eine Tabelle beschreibt, die pro Textzeile einen Datensatz enthält und innerhalb einer Zeile die einzelnen Spalten (Werte) für Artikelnummern, Preise usw. durch Kommas trennt. Es gibt zahlreiche Varianten für Spaltentrennzeichen, Textbegrenzungszeichen und Escape-Zeichen, mit dem bspw. die Bedeutung des Spaltentrennzeichens innerhalb eines Textfeldes aufgehoben wird. Die Reihenfolge der Spalten ist nicht vorgeschrieben, muss aber innerhalb einer CSV-Datei in jeder Zeile gleich sein.

```
'Artikelnummer';'Preis';'Deeplink'
'23';'19.90';'http://testshop.de/catalog/23_ie.html'
'42';'22.95';'http://testshop.de/catalog/42_ie.html'
```

Listing 5: Produktdatei passend zur shopinfo.xml

Die gleichen Produktdaten könnten z. B. auch wie folgt aussehen, wobei die shopinfo.xml entsprechend anzupassen wäre:

```
Bruttopreis|Url|Id
19,90|http://testshop.de/catalog/23_ie.html|23
22,95|http://testshop.de/catalog/42_ie.html|42
```

Listing 6: Geändertes Format für die gleiche Produktdatei

In der shopinfo.xml wird der Aufbau der Produktdatei flexibel beschrieben, sodass ein voll automatisierter Zugriff möglich ist – unabhängig davon, wie die Produktdatei erstellt wurde und wie sie aufgebaut ist. So kann jeder Shop die Technik zur Erstellung seiner Produktdatei einsetzen, die ihm zur Verfügung steht, und wird nicht gezwungen, ein bestimmtes Format zu erzeugen.

Im shopinfo.xml-Standard sind einheitliche Bezeichnungen für Produkteigenschaften wie Produktname, Produktnummer, international eindeutige Produktkennzeichnungen wie EAN (*International Article Number*), UPC (*Universal Product Code*) usw. definiert. Im <Mappings>-Element (siehe Listing 4) werden die Spalten der individuellen CSV-Dateien den einheitlichen Bezeichnungen der Produkteigenschaften zugeordnet. Über diese Zuordnung lassen sich die Produktdaten und ihre jeweiligen Eigenschaften automatisiert in das System der Shopping-Portale einlesen.

**Integration der Produktdaten über eine Echtzeitschnittstelle.** Die Integration der Daten über eine Echtzeitschnittstelle erfolgt über den Aufruf des in <OnlineRequest> definierten Elements <Processor> mit den angegebenen Parametern (siehe Listing 1). So nennt das Element <ParamProduct> den Namen des Parameters (hier: product), der für eine Suche nach dem Produktnamen eingesetzt wird. Dieser Parameter wird zusammen mit den URL-kodierten Suchbegriffen an die im <Processor>-Element angegebene Schnittstelle geschickt, z. B.:

```
http://testshop.de/servlets/search?product=canon+ixus+65
```

Als Ergebnis wird ein XML-Dokument geschickt, welches gemäß dem XML-Schema <http://elektronischer-markt.de/schema/products.xsd> aufbereitet ist. Die eingesetzte Technik (PHP, ASP, JSP, ...) aufseiten des Webserverns spielt keine Rolle. Nachfolgend ein Beispiel für das zurück gelieferte Suchergebnis:

```
<?xml version = "1.0" encoding="ISO-8859-1"?>
<osp:ProductList xmlns:osp="http://elektronischer-markt.de/schema" ... >
  <Common> ... </Common>
  <Item>
    <Name>Canon Ixus 65</Name>
    <PriceDetails type="withTax">
      <Price><Amount>699</Amount></Price>
    </PriceDetails>
  </Item>
</osp:ProductList>
```

Listing 7: Beispiel für ein Ergebnis einer Echtzeitsuchanfrage

Ein solches Ergebnis lässt sich, wie bereits bei der Integration der Shopdaten verdeutlicht, leicht in die eigenen Systeme integrieren.

**Datenaufbereitung auf Seiten der Online-Shops.** Zur Förderung der Verbreitung des Standards wurden für zahlreiche Shopsysteme Module entwickelt, die zum einen den Shopbetreibern bei der Erstellung der shopinfo.xml helfen und zum anderen die Produktdaten in Form einer CSV-Datei sowie per Echtzeitschnittstelle bereitstellen (siehe <http://elektronischer-markt.de/nav?dest=impl.shopsystems.index>).

### **3 Zusammenfassung und Ausblick**

Der shopinfo.xml-Standard erleichtert den Datenaustausch zwischen Online-Shops und Shoppingportalen. Aktiv nutzen bereits über 2000 Shops und mehr als ein Dutzend kommerzieller Betreiber bzw. Vermarkter von Shopping-Portalen den Standard. Mithilfe der deutschen Ableger international operierender Shopping-Portale wird sich der shopinfo.xml-Standard weltweit verbreiten. Die derzeit noch in der Entwicklung befindliche Version 2.0 des shopinfo.xml-Standards wird gegenüber der Vorgängerversion daher insbesondere bzgl. international notwendiger Daten erweitert. So können bspw. sämtliche Texte in mehreren Sprachen abgefasst werden.

Die zunehmende Verbreitung des shopinfo.xml-Standards und die damit einhergehende Vereinfachung des Datenaustausches erhöhen den Wettbewerbsdruck unter den Shopping-Portalen. Eine einfache Volltextsuche in den Produktdaten genügt nicht mehr, um für Kunden attraktiv zu sein. Zusatzangebote wie detaillierte Produktkategorisierungen, Produktinformationen usw. spielen daher eine entscheidende Rolle bei der Kundenbindung. Nicht zuletzt ergeben sich auch weitere Einsatzmöglichkeiten für die im Standard angelegte Echtzeitsuche in Verbindung mit AJAX [Ge05].

Die im shopinfo.xml-Standard angelegten Elemente zur ausführlichen Beschreibung der Online-Shops hinsichtlich Liefer-, Zahlungs-, Abholmöglichkeiten usw. bieten Shops genügend Spielraum, sich von der Konkurrenz nicht nur über den Preis zu differenzieren. In letzter Konsequenz können alle Shop- und Produktdaten mithilfe von Intermediären wie Shopping-Portalen gezielt recherchiert werden. Nachfrager finden, was sie suchen. Unter Berücksichtigung der Servicedaten werden die günstigsten (und nicht nur die billigsten) Angebote gefunden.

### **Danksagung**

Die Autoren bedanken sich bei der Medien- und Filmgesellschaft Baden-Württemberg (MFG) für die Förderung durch ein Karl-Steinbuch-Stipendium. Des Weiteren danken wir den anonymen Gutachtern für hilfreiche Anregungen.

### **Literaturverzeichnis**

- [Ge05] Garrett, J.J.: Ajax – A New Approach to Web Applications, Adaptive Path LLC, Februar 2005, <http://www.adaptivepath.com/publications/essays/archives/000385.php>
- [HKS02] Hansch, M.; Kuhlins, S.; Schader, M. (2002): XML-Schema. In: Informatik Spektrum, Band 25, Heft 5, S. 363–366.
- [HM04] Harold, E.R.; Means, W.S.: XML in a Nutshell, 3. Aufl., O'Reilly, 2004
- [Ph06] PHP Group: PHP Handbuch, 2006, <http://www.php.net/manual/de/ref.simplexml.php>