INSTITUT FÜR DEUTSCHE SPRACHE
OPAL
Online publizierte Arbeiten zur Linguistik

ISSN 1860-9422
Sonderheft
1/2008

Wolfgang Bock

Technische Aspekte des OWID-Portals

aus: Klosa (Hg.): Lexikografische Portale im Internet.

(= OPAL-Sonderheft 1/2008), S. 37-44.

OPAL – Online publizierte Arbeiten zur Linguistik
Herausgegeben vom Institut für Deutsche Sprache

Institut für Deutsche Sprache
Postfach 10 16 21
68016 Mannheim
opal@ids-mannheim.de

Technische Redaktion: Norbert Volz

Wolfgang Bock

Technische Aspekte des OWID-Portals

Abstract

Der Beitrag behandelt die technischen Gesichtspunkte unter denen die Neustrukturierung und -gestaltung des OWID-Portals erfolgte. Ausgehend von den definierten Anforderungen werden die grundlegenden Strukturen des Portals dargelegt. Die verwendeten technischen Standards (Oracle-Datenbank, XML/XSL, HTML/CSS) werden beschrieben und ihr Zusammenspiel erläutert. Exemplarisch wird der Weg eines Artikels von der Entstehung über das Einchecken ins System bis hin zur Aushabe als HTML-Seite dargestellt.

This contribution covers the technical aspects of the restructuring and the new design of the OWID-portal. Based on the requirements the basic structures of the portal are explained. The technical standards used (Oracle database, XML/XSL, HTML/CSS) are described and their interaction is illustrated. The way of an article starting from the emergence, via the checking into the system to the output as an HTML page is treated.

Inhalt:

Die Neustrukturierung und Neugestaltung des OWID-Portals war und ist immer noch eine herausfordernde Aufgabe. Sie bestand nicht nur in der Entwicklung und Umsetzung eines neuen Designs für einen zeitgemäßen Auftritt, sondern auch in der Optimierung des technischen Aufbaus. Der zentrale Punkt war jedoch die Einbindung neuer Module in die bestehende Infrastruktur des *elexiko*-Wörterbuchs. Durch die Erweiterung der Präsentation eines einzelnen Projektes zur Darstellung verschiedener Projekte in einem gemeinsamen Kontext wurde das lexikografische Portal OWID geschaffen.

Die Herausforderung, ein Portal wie OWID aus einer bestehenden Infrastruktur heraus zu erschaffen, spielt sich auf drei unterschiedlichen technischen Ebenen ab: Die Datenbankebene beinhaltet alle Daten, die in der Darstellungsebene angezeigt werden. Die Verbindung zwischen diesen beiden Ebenen übernimmt die Kommunikationsebene. Sie regelt, was wie und wann angezeigt wird. Jede dieser drei Ebenen musste erweitert und gegebenenfalls umstrukturiert werden und für jede gab es spezifische Anforderungen.

1. Anforderungen

1.1 Umbau zum Portal

Die wichtigste Aufgabe bestand in der Erweiterung der Datenbank von der Verwaltung eines einzelnen Wörterbuchs zu derjenigen mehrerer Wörterbücher, von denen sich manche in der Zielsetzung und damit auch im internen Aufbau völlig unterscheiden. Auf diese Struktur und ihre Umsetzung in der Datenbank gehe ich im Verlauf des Artikels noch ausführlich ein.

1.2 Ansprechendes, modernes und flexibles Design der Oberfläche

In den Anfängen von *elexiko* lag der Fokus bei der Gestaltung der Webseiten auf der klaren, einfachen und übersichtlichen Darstellung der Inhalte. Diese Anforderungen galten selbstverständlich auch für das neue Portal, denn der regelmäßige Besucher der Seiten musste sich auch im neuen Design sofort zurechtfinden. Darüber hinaus sollte die Site mit modernen Mitteln der HTML-Codierung aufgebaut werden, um so die Flexibilität zu erhöhen und eine möglichst weitgehende Barrierefreiheit zu erhalten.

1.3 Einfache Pflege

Ein Portal wie OWID ist kein starres Gebilde, das, einmal erstellt, für immer unverändert im Netz steht. Es soll sich verändern und entwickeln. Daher ist es von zentraler Bedeutung, dass Aufbau und Pflege des Portals so einfach wie möglich gestaltet werden. Um dieses Ziel zu erreichen, sollten, wo immer möglich, die einzelnen Bereiche so standardisiert aufgebaut werden, dass sie an einer zentralen Stelle des Systems erzeugt, gepflegt und aktualisiert werden können. Denn nur so können Änderungen zuverlässig und schnell portalweit umgesetzt werden. Zu diesem Zweck ist es wichtig, sich zuerst einmal die Voraussetzungen anzusehen, unter denen der Umbau in Angriff genommen wurde.

2. Ausgangssituation

2.1 Die Oberfläche

Die bereits bestehende Oberfläche des *elexiko*-Wörterbuchs sollte als Ausgangspunkt für die neue Gestaltung des OWID-Portals dienen. Sie bestand im Wesentlichen aus drei Bereichen (vgl. Abbildung 1): der Navigationsleiste im Kopfbereich (1), einer Spalte zur Darstellung von Wortlisten (2) und dem eigentlichen Inhaltsbereich (3), in dem die Wortartikel und weitere Informationen dargestellt wurden. Dieser Aufbau hatte sich bewährt, und die Nutzer des OWID-Wörterbuchs hatten sich daran gewöhnt. So bestand weder die Notwendigkeit noch der Wunsch, diesen Aufbau grundlegend zu ändern. Er sollte erhalten, modernisiert und erweitert werden (vgl. Abbildung 2).

Die drei beschriebenen Bereiche der Seiten wurden in separaten HTML-Dateien erzeugt und mittels eines Framesets (ein HTML-Konstrukt, das es ermöglicht, mehrere HTML-Dateien in einem Browserfenster darzustellen) angezeigt. Der Vorteil und ursprüngliche Grund für diese Technik liegt darin, dass einzelne Bereiche einer Seite separat geladen werden können, was in Zeiten geringer Datenübertragungsraten einen großen Vorteil darstellte. Die heute übliche Bandbreite der Datennetze rechtfertigt einen Einsatz von Frames nicht mehr.

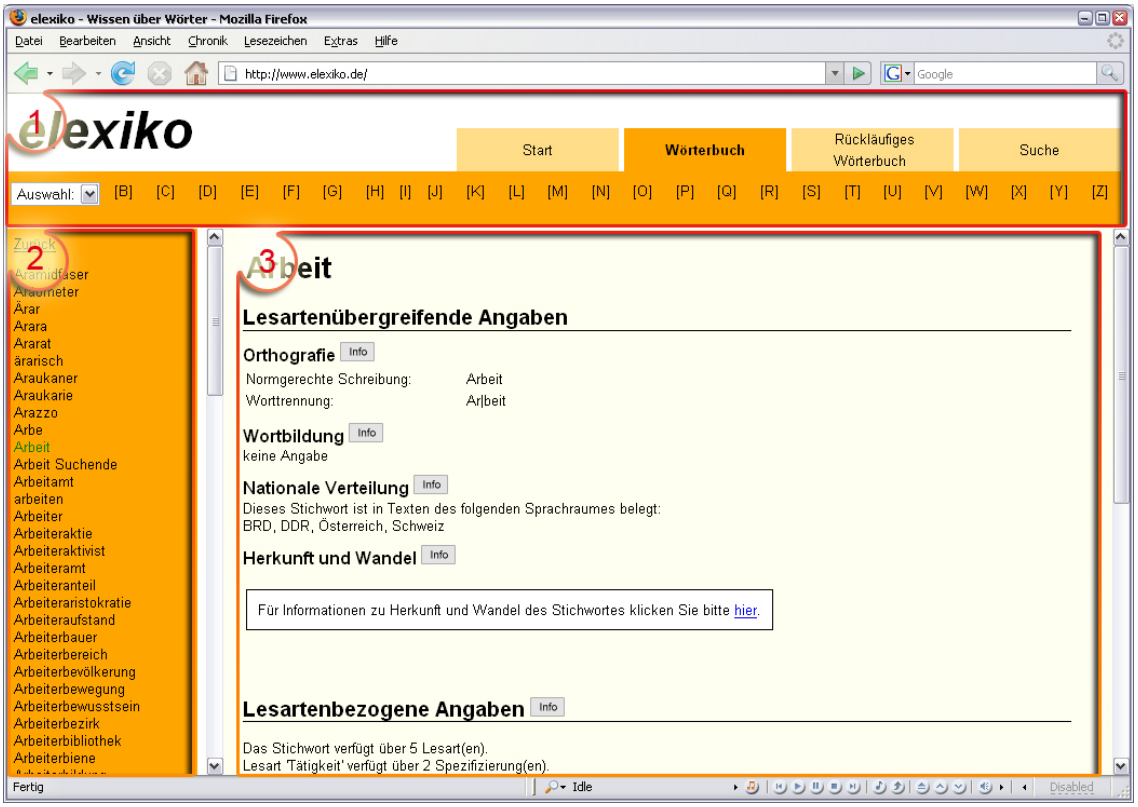

Abbildung 1: Aufbau der alten *elexiko*-Seiten

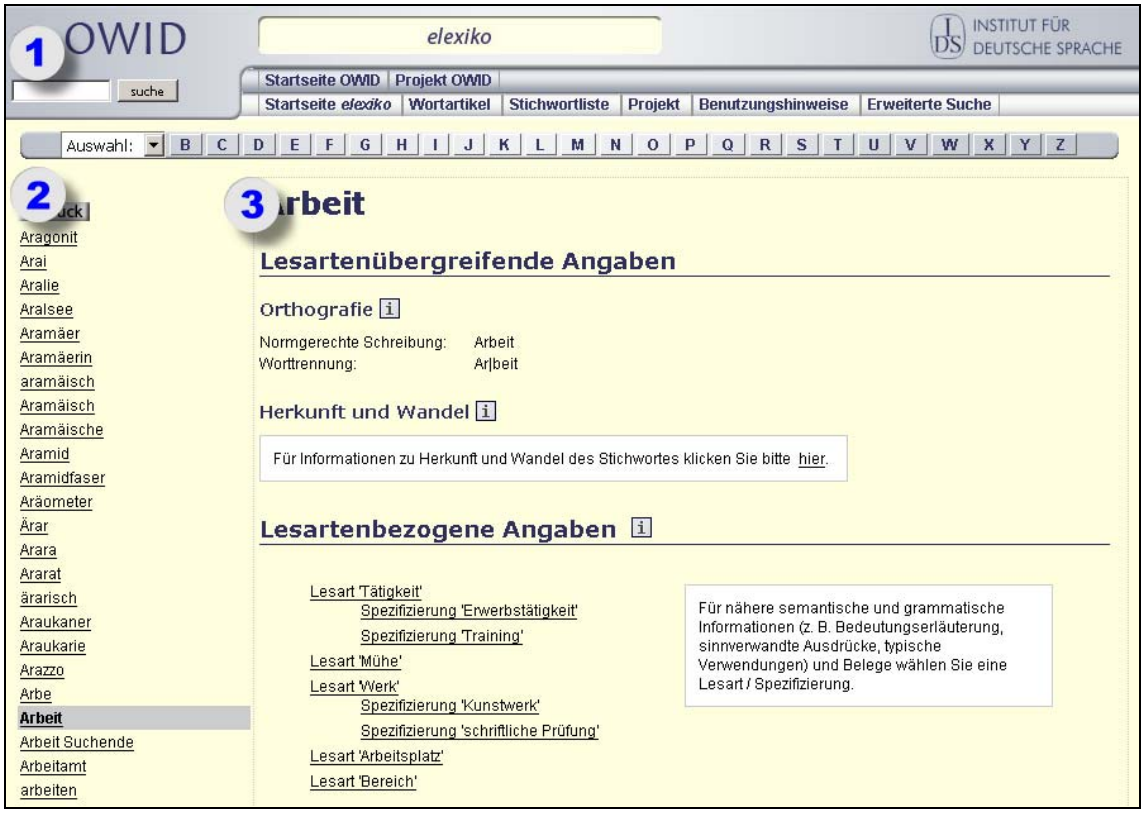

Abbildung 2: Der neue Aufbau der *elexiko*-Seiten

Die Nachteile von Frames liegen in der schlechteren Zugänglichkeit für Suchmaschinen und in der mangelnden Barrierefreiheit. So genannte Screenreader, die sehbehinderten Menschen Webinhalte vorlesen, können den Inhalt von Frames nicht erreichen. Daher sollte bei der technischen Neukonzeption des OWID-Portals auf Frames verzichtet werden.

Ebenso wie der strukturelle Aufbau der Site sollte auch der Aufbau der einzelnen HTML-Dokumente auf einer modernen, flexiblen Architektur basieren. Denn in der jüngeren, der ohnehin kurzen Geschichte des WWW setzt sich die Trennung von Inhalt und Formatierung immer mehr durch, und das aus gutem Grund.

Ursprünglich war das Internet zum schnellen Austausch strukturierter, meist wissenschaftlicher Daten gedacht. Diesen Zweck erfüllte es von Anfang an hervorragend. Nicht die Präsentation, sondern der Inhalt stand im Vordergrund. Mit der zunehmenden kommerziellen Nutzung des Webs änderten sich die Anforderungen, die an HTML gestellt wurden. Das Layout der Seiten, eine Funktion, die in HTML nur rudimentär integriert war, wurde immer wichtiger. Der HTML-Code einzelner Seiten wurde durch Formatierungsanweisungen extrem aufgebläht und gleichzeitig unübersichtlich, da Formatierungen für jedes Element auf jeder Seite einzeln eingesetzt werden mussten. Abhilfe schaffte die Einführung der *CascadingStyleSheets* (CSS), d.h. von Stilvorlagen, die flexibel auf bestimmte Bereiche angewendet werden können.

Um z.B. den Brottext einer Seite, also den Text, der die eigentlichen Informationen enthält, zu formatieren, reicht es aus, die Formateigenschaften in einem Stylesheet zu definieren. Änderungen in diesem Stylesheet werden sofort auf alle Textpassagen angewendet, die als Brottext ausgezeichnet sind. Dieses Konzept der Trennung von Formatierung und Inhalt ist nicht nur innerhalb einzelner Seiten, sondern auch über eine ganze Website wie z.B. das OWID-Portal anwendbar. In den HTML-Dateien findet sich also nur noch die reine Information in strukturierter Form. Die Formatierungsanweisungen finden sich in den CSS-Dateien, die auf die Strukturmerkmale der HTML-Dateien zugreifen.

2.2 Der Aufbau

Neben der Struktur der einzelnen Webseiten sollte auch der Aufbau der Funktionen innerhalb des Portals optimiert werden. Die einzelnen Seiten bestehen aus immer wiederkehrenden Elementen, die aber in verschiedenen Zusammenhängen unterschiedliche Inhalte haben können. Am einfachsten und besten wird das am Beispiel des Menüs deutlich.

Die Navigationsleiste, die ausnahmslos auf jeder Seite des Portals sichtbar ist, besteht aus zwei Zeilen, dem Hauptmenü mit den immer gleichbleibenden Punkten „Portal-Startseite“ und „Portal-Informationen“ und dem darunterliegenden Untermenü, das von Wörterbuch zu Wörterbuch unterschiedlich ist. Natürlich ist es möglich, das Menü für jede denkbare Situation vollständig und individuell in HTML zu erzeugen. Die Anzahl der verschiedenen Szenarien ist begrenzt. Sie liegt bei ca. drei verschiedenen Darstellungsformen für jedes Wörterbuch. Rechnet man noch drei Situationen für das Portal hinzu, kommt man bei vier Wörterbüchern auf 15 verschiedene Kombinationen. Da sich die einzelnen Szenarien nur minimal unterscheiden, wäre diese Arbeit per „Kopieren und Einfügen“ schnell erledigt. Dass diese Methode trotzdem nicht effizient ist, zeigt

sich spätestens bei nachträglichen Änderungen am Seitenaufbau, bei denen in diesem Fall 15 verschiedene Codeschnipsel berücksichtigt und gegebenenfalls geändert werden müssten. Diese Vorgehensweise ist nicht nur äußerst mühsam, sie ist darüber hinaus auch extrem fehleranfällig.

In der Erstprogrammierung zwar aufwendiger, aber in der Aktualisierung und Pflege einfacher ist die Erstellung eines eigenständigen Menümoduls, das von jeder Stelle des Portals aufgerufen wird und situationsabhängig auf die Anforderungen reagiert. Beim Aufruf des Menümoduls wird dem Programm ein Parameter übergeben, der beschreibt, für welches Wörterbuch das Menü erzeugt werden soll. Mithilfe dieses Parameters kann das Programm auf die gestellten Anforderungen reagieren. Durch die Übergabe weiterer Parameter können die möglichen Situationen weiter differenziert werden.

Zwar ist die Programmierung eines solchen Programmmoduls deutlich komplexer als die einfache Erstellung einzelner HTML-Szenarien, spätere Änderungen können dagegen viel einfacher und vor allem sicherer durchgeführt werden, da nur ein zentrales Modul für die Darstellung aller Menüs benutzt wird. Änderungen in diesem Modul werden sofort in allen Bereichen des Portals wirksam.

3. Vom Online-Wörterbuch zum Portal

Die dritte und größte Herausforderung lag darin, aus einem homogenen Wörterbuch ein Portal zu schaffen, dessen Inhalte aus ganz unterschiedlichen Quellen mit ebenso unterschiedlichen Zwecken geschaffen werden. Um diese teils gegensätzlichen Aufgaben unter einem Dach zu vereinen, bedurfte es einiger grundlegender Regeln.

3.1 XML als Standardformat

Die Auszeichnungssprache XML (*eXtensible Markup Language*) dient der „Darstellung hierarchisch strukturierter Daten in Form von Textdateien“ (Wikipedia, http://de.wikipedia.org/wiki/XML, Stand: 02.12.2007, 16:40 Uhr). Innerhalb der Datei werden also die Daten nicht formatiert, sondern lediglich strukturiert. Dabei unterliegt nur die Vorgehensweise der Strukturierung, die Struktur selbst unterliegt keinen strengen Regeln. Solange die Struktur formal richtig definiert ist, kann das XML-Dokument beliebige Strukturen enthalten. Der Vorteil von XML liegt in der verhältnismäßig einfachen maschinellen Verarbeitung.

Um jedoch eine große Gruppe gleichartiger Dokumente abarbeiten zu können, müssen diese auch zuverlässig auf derselben Struktur basieren. Minimale Abweichungen führen zwangsläufig zum Scheitern des gesamten Systems. Zur Definition und Prüfung von XML-Strukturen stehen *Document Type Definitions* (DTDs) zur Verfügung. Bei der Erstellung und Bearbeitung von XML-Dateien in dafür vorgesehenen Editoren wird die Struktur des Dokuments nach den in der DTD festgelegten Strukturen überprüft und bei Fehlern wird gewarnt. Selbst bei sehr komplexen XML-Konstrukten, wie z.B. den *elexiko*-Wortartikeln, sind dadurch einheitliche und fehlerfreie Dokumente garantiert, vorausgesetzt, die Wortartikel werden ausschließlich mit Editoren bearbeitet, die mit DTDs umgehen können.

Eine direkte Darstellung von XML-Dokumenten in Webbrowsern ist zwar möglich, in der Regel jedoch nicht ratsam, da Browser nicht nur die relevanten Informationen an-

zeigen, sondern auch die umgebenden Strukturelemente. Bei komplexeren Dateien geht die Übersichtlichkeit sehr schnell verloren. Abgesehen davon enthält XML per Definition keine Formatierungsanweisungen. Die Inhalte werden also listenartig und in immer gleicher Weise dargestellt, denn ein Browser kann nicht zwischen verschiedenen Gewichtungen der Elemente unterscheiden.

Um die Artikel korrekt in einem Browser darstellen zu können, müssen sie mit der *eXtensible Stylesheet Language* (XSL) in HTML transformiert werden. Mittels *XSL Transformation* (XSLT) lassen sich XML-Dateien in nahezu jedes andere Dateiformat überführen. Vor der Ausgabe auf den Bildschirm durchläuft das XML-Dokument die Anweisungen einer XSL-Datei und wird entsprechend den enthaltenen Anweisungen als HTML ausgegeben. Auf diese Weise können aus den streng strukturierten XML-Dateien beliebige HTML-Dateien generiert werden. Die Umwandlung von XML in HTML ist wahrscheinlich die häufigste Verwendung von XSLT. Andere Transformationen, wie diejenige ins PDF-Format oder in eine andere XML-Struktur, sind aber ebenfalls üblich.

Durch das Zusammenspiel von XML, DTD und XSLT steht eine hochflexible und gut verarbeitbare Dokumentenstruktur zur Verfügung, die universell in verschiedenste Umgebungen eingebunden werden kann.

3.2 Einbindung in die Datenbank

Zur Strukturierung beliebiger Datenarten ist XML hervorragend geeignet. Gleichzeitig ist es auf diese Aufgabe beschränkt. Es beinhaltet keinerlei Funktionen zur Verwaltung von Daten. Diese Aufgabe übernimmt für das Projekt OWID eine Datenbank aus dem Hause Oracle. In der Datenbank werden neben den verschiedenen Suchfunktionen innerhalb des Portals auch die gesamte Darstellung des Portals im Web und die Verwaltung der Daten auf dem Server definiert. Erst durch die logische Verbindung innerhalb der Wörterbücher und zusätzlich auch über die Wörterbuchgrenzen hinweg wird aus einer Ansammlung von Wörterbüchern ein zusammenhängendes Portal.

4. Von XML zur Webseite

Die Wortartikel werden mit einem Texteditor erstellt, der auf dem Rechner des jeweiligen Lexikografen installiert ist. Um einen Artikel dem Portal zur Verfügung zu stellen, muss er in das System eingecheckt werden.

4.1 Einchecken eines Wortartikels

Beim Einchecken durchläuft der Wortartikel mehrere Prozeduren. Nachdem der Autor die entsprechenden Daten auf seinem lokalen Rechner ausgewählt und hochgeladen hat, wird mittels der im System hinterlegten DTDs die Struktur des Artikels geprüft. Entspricht der Artikel den Vorgaben nicht, wird eine entsprechende Meldung abgegeben und eine Speicherung findet nicht statt.

Ist der Artikel fehlerfrei, wird er mittels PL/SQL, einer Oracle-eigenen Programmiersprache, und XPath, einer standardisierten XML-Abfragesprache analysiert. Bestimmte Elemente, wie z.B. die Lemmazeichengestaltangabe oder bestimmte Formangaben, werden aus dem Artikel extrahiert und in einzelne Tabellen in der Datenbank geschrie-

ben (vgl. Abbildung 3). Diese extrahierten Daten werden von der Datenbank indiziert und können so um ein Vielfaches schneller durchsucht werden, als direkte Abfragen des Portals auf den XML-Strukturen es ermöglichen würden. Darüber hinaus wird jeder Artikel mit einer eindeutigen ID versehen, anhand deren die in verschiedenen Tabellen abgelegten Informationen den jeweiligen Wortartikeln zugeordnet werden können.

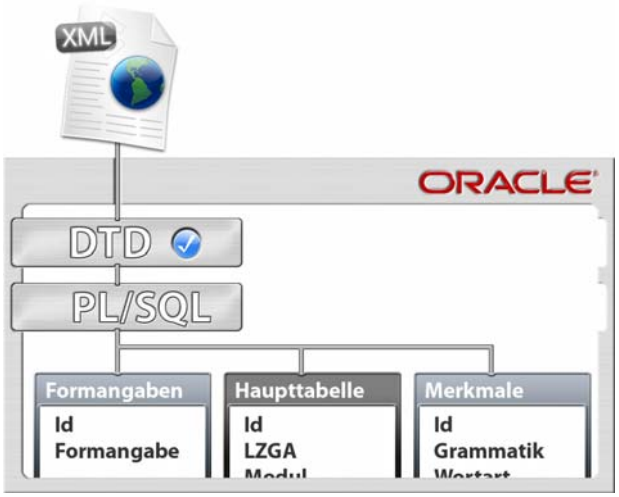

Abbildung 3: Speicherung der XML-Instanzen in die Datenbank

4.2 Ausgabe eines Wortartikels

Sobald der Artikel eingecheckt ist, steht er dem gesamten System zur Verfügung. Die Abfragemöglichkeiten im Portal sind so gestaltet, dass die Suchkriterien ausschließlich auf extrahierte Daten zugreifen. So ist trotz großer Datenmengen eine akzeptable Zugriffszeit gewährleistet. Wird anhand der gestellten Abfrage ein eindeutiger Treffer innerhalb aller relevanten Tabellen gefunden, identifiziert das System den Artikel anhand der ID und gibt ihn aus. Während der Ausgabe wandelt das im System abgelegte XSLT die XML-Struktur des Artikels wie oben beschrieben in HTML um. Das generierte HTML enthält jedoch keine Formatierungsanweisungen. Diese werden durch ein externes Stylesheet gewährleistet (vgl. Abbildung 4).

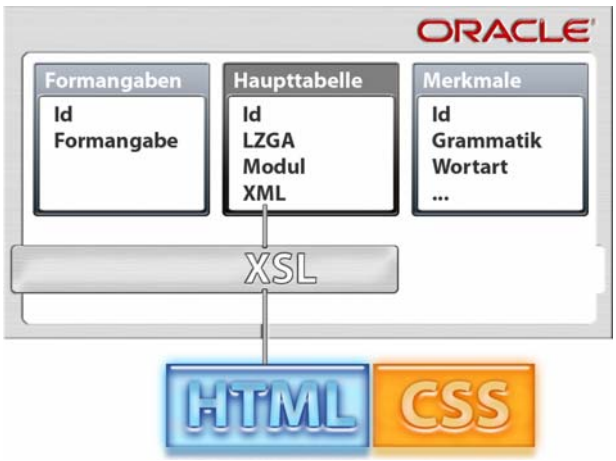

Abbildung 4: Ausgabe der Wortartikel aus der Datenbank

Durch die konsequente Verwendung offener Standards in Verbindung mit einem hochleistungsfähigen Datenbanksystem stellt das OWID-Portal ein flexibles und zukunftsfähiges System dar. Es kann weiter wachsen und bleibt für Veränderungen, Erweiterungen und Verbesserungen offen.