

## Lebensmittelkontrolle 2.0

Alexandra Krewinkel<sup>1</sup> · Sebastian Sünkler<sup>2</sup> · Dirk Lewandowski<sup>2</sup> ·  
Niklas Finck<sup>1</sup> · Boris Tolg<sup>1</sup> · Lothar W. Kroh<sup>3</sup> · Georg A. Schreiber<sup>4</sup> ·  
Jan Fritsche<sup>1</sup>

Published online: 11 November 2015

© Bundesamt für Verbraucherschutz und Lebensmittelsicherheit (BVL) 2015

**Zusammenfassung** Der Lebensmittelhandel im Internet erhält eine stetig zunehmende Relevanz im E-Commerce und stellt Überwachungsbehörden vor neue Herausforderungen. Behörden sind dazu verpflichtet, deutsche Lebensmittelhändler nicht nur im stationären Handel, sondern auch im Online-Handel zu kontrollieren. Dabei entstehen für die Kontrollbehörden Schwierigkeiten im Umgang mit der enormen Menge an Daten und der Schnelllebigkeit der Angebote im World Wide Web. Für eine effiziente Kontrolle wird daher eine automatisierte IT-basierte Lösung benötigt, die auf einem detaillierten Konzept zur Online-Überwachung des virtuellen Lebensmittelmarktes basiert.

Das AAPVL-Forschungsprojekt beschäftigt sich mit den inhaltlichen und technischen Anforderungen für die Realisierung einer solchen Lösung. Im Rahmen des Projekts wurde ein neuartiges theoretisches Konzept zur Lebensmittelüberwachung im Internet entwickelt und ein Software-Prototyp zur automatisierten Überwachung von efoods realisiert.

### 1 Hintergrund

Das Warenangebot im Internet ist nahezu unbegrenzt. Neben den Klassikern wie Kleidung und Bücher lassen sich seit einigen Jahren auch Lebensmittel bequem über das Internet direkt nach Hause bestellen. Hierbei ist die Vielfalt des Warenangebots ebenso breit gefächert wie die Größe. Es gibt Anbieter mit großen Internetpräsenzen von mehreren tausend Seiten und Produkten, aber auch Händler mit spezialisierten Angeboten mit einem einzigen Produkt. In den letzten Jahren zeigte sich, dass der Lebensmittelhandel im Internet eine stetig wachsende Branche ist. Aktuelle Prognosen zufolge wird der Marktanteil bis zum Jahr 2020 von 0,3 auf 10 % ansteigen und somit 2020 einen Umsatz von 20 Mio. € im E-Food-Bereich erzielt werden (Wagner und Wiehenbrauk 2014). Eine solche Entwicklung bleibt auch von der deutschen Lebensmittelüberwachung nicht unbemerkt. Die Überwachungsbehörden sind gemäß EU-Verordnung 882/2004 verpflichtet, die Kontrollen von Lebensmittelunternehmern auf Basis einer gesundheits- und täuschungsbezogenen Risikobewertung durchzuführen. Hierzu zählen, wenn auch nicht explizit erwähnt, Lebensmittelhändler, die ihre Ware über das Internet anbieten. Im

---

Der “Food Science Dialog” fand vom 15. bis 18. September 2015 in Hamburg statt und wurde von der Hochschule für Angewandte Wissenschaften Hamburg (HAW), der Behörde für Gesundheit und Verbraucherschutz (BGV) und der Behörde für Wissenschaft und Forschung (BWF) und in Kooperation mit der Universität Wageningen (NL) organisiert und durchgeführt.

---

✉ Jan Fritsche  
Jan.Fritsche@haw-hamburg.de

- <sup>1</sup> Fakultät Life Sciences, Hochschule für Angewandte Wissenschaften Hamburg, Ulmenliet 20, 21033 Hamburg, Germany
- <sup>2</sup> Fakultät Design, Medien, Information, Hochschule für Angewandte Wissenschaften Hamburg, Finkenau 35, 22081 Hamburg, Germany
- <sup>3</sup> Institut für Lebensmitteltechnologie und Lebensmittelchemie, TU Berlin, Gustav-Meyer-Allee 25, 13355 Berlin, Germany
- <sup>4</sup> Bundesamt für Verbraucherschutz und Lebensmittelsicherheit, Mauerstraße 39–42, 10117 Berlin, Germany

Lebensmittel- und Futtermittel-Gesetzbuch wird in Paragraph 3 Nr. 1 in Verbindung mit Art. 3 Nr. 8 der Lebensmittel-Basisverordnung (EG) Nr. 178/2002 definiert, dass das Inverkehrbringen unter anderem ein Anbieten zum Verkauf ist. Dies verpflichtet, laut Europäische Kommission im Leitfaden für die Durchführung einzelner Bestimmungen der Lebensmittelhygieneverordnung, auch Onlinehändler, die Hauptverantwortung für die Sicherheit des von ihnen angebotenen Lebensmittels zu tragen, und sich als Lebensmittelunternehmer bei den Lebensmittel- und Veterinärämtern zu registrieren. Dies stellt vor allem die Überwachungsbehörden aufgrund der neuen Abläufe des Internethandels, die nicht grundsätzlich mit denen des stationären Handels vergleichbar sind, vor neue Herausforderungen. Eine der größten Schwierigkeiten ergibt sich durch die im Internet nahezu unbegrenzte Datenmenge, die mit herkömmlichen Mitteln nur durch erheblichen Arbeits- und Zeitaufwand zu bewältigen ist. Für eine effiziente Kontrolle bedarf es einer IT-basierten automatisierten Lösung, um eine Kontrolle des stetig wachsenden und schnelllebigen Marktes zu realisieren, und nicht zuletzt den Verbraucher zu schützen.

## 2 Das Forschungsprojekt AAPVL

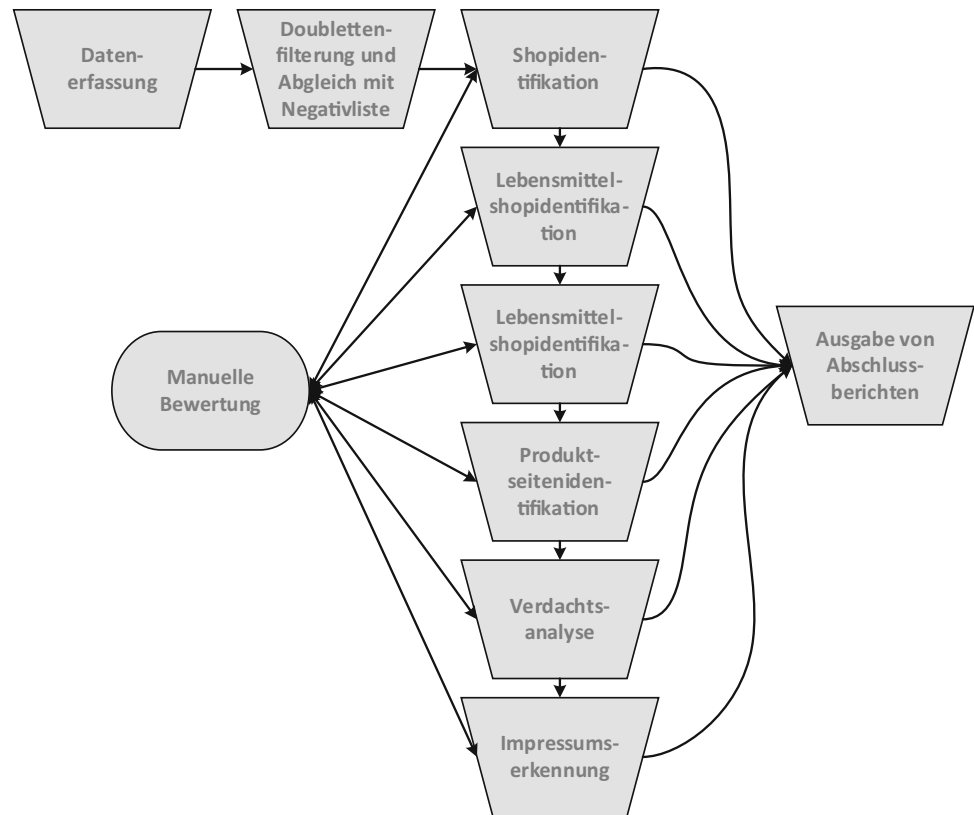
An der Hochschule für Angewandte Wissenschaften in Hamburg wurde in Kooperation mit dem Bundesamt für Verbraucherschutz und Lebensmittelsicherheit (BVL) das interdisziplinäre Forschungsprojekt „Entwicklung von automatisierten Analyseverfahren zur Identifizierung und Bewertung von nicht verkehrsfähigen Produkten des virtuellen Lebensmittelmarktes“ (AAPVL) ins Leben gerufen. Die Forschungsgruppe hat sich zum Ziel gesetzt, ein System zur automatisierten Identifizierung und Bewertung von potentiell nicht verkehrsfähigen Lebensmitteln, die über das Internet vertrieben werden, zu entwickeln. Die Förderung des Vorhabens erfolgt aus Mitteln des Bundesministeriums für Ernährung und Landwirtschaft (BMEL) aufgrund eines Beschlusses des deutschen Bundestages. Die Projektträgerschaft erfolgt über die Bundesanstalt für Landwirtschaft und Ernährung (BLE) im Rahmen des Programms zur Innovationsförderung. Am Ende des Innovationsprojekts sollen eine Strategie, sowie ein Softwareprototyp zur automatisierten Lebensmittelkontrolle im Internet stehen. Neben der automatisierten Recherche nach Anbietern potentiell nicht verkehrsfähiger Produkte sollen auch verbrauchertäuschende Werbeaussagen sowie Siegelmissbräuche aufgedeckt werden können.

### 2.1 Konzept zur automatisierten Kontrolle

Das Konzept zur Kontrolle der über das Internet vertriebenen Lebensmittel sieht einen mehrstufigen Prozess mit verschiedenen Filterstufen vor (Abb. 1) (Krewinkel et al. 2015). Jede Filterstufe besteht aus einem autonomen Modul und dient dazu, auf flexible Art automatisiert die Relevanz der Daten zu erhöhen. Schlussendlich sollen so den Mitarbeitern der Überwachungsbehörden potentielle Verstöße gegen das Lebensmittelrecht zur weiteren rechtlichen Prüfung und ggf. Durchführung von behördlichen Maßnahmen übergeben werden. Zunächst wird über variable Suchanfragen im Datenerfassungsmodul ein Datenpool erfasst. Die Suchanfragen werden automatisiert an verschiedene Suchmaschinen geschickt, Duplikate herausgefiltert und gegen eine Negativliste abgeglichen. Die Ergebnisseiten mit Hauptdomain und Impressumseite werden in der Datenbank gespeichert. Alle weiteren Analysen laufen nun auf den gespeicherten Seiten in der Datenbank ab. So wird zum einen ein Mindestmaß an rechtlicher Dokumentation gewährleistet, und zum anderen gibt es dadurch auch keine Ergebnisverzerrungen durch Änderungen oder Aktualisierungen auf den Webseiten während der Analyse. Im Weiteren sollen nun alle Seiten herausgefiltert werden, bei denen es sich um einen Shop und in der nächsten Stufe um einen Lebensmittelshop handelt. Dies dient vor allem der Filterung von Blogs, Foren, Informationsartikeln und Werbeseiten, die zwar den gesuchten Stoff oder das gesuchte Produkt behandeln, aber kein eigentliches Produktangebot darstellen.

Ein weiteres Modul, die Produktseitenidentifikation, soll dem Herausfiltern von Übersichtseiten in einem Shop dienen. So soll sichergestellt werden, dass eine spätere Analyse auch auf der tatsächlichen Angebotsseite mit allen relevanten Informationen über das Produkt stattfindet. Im Verdachtsanalysemodul werden geeignete lebensmittelrechtliche Kriterien hinterlegt, auf Basis derer der spätere Softwareprototyp Hinweise auf einen potentiellen Verstoß ausgeben wird. Über eine Impressumserkennung kann der Verstoß dem entsprechenden Händler zugewiesen werden. Dabei erfolgt auch eine Zuweisung des zuständigen Bundeslandes und Kreises. Eine Zusammenfassung aller ermittelten Daten findet über die Ausgabe eines variablen Schlussberichts statt. Der User wird zudem an jeder Stelle des automatisierten Prozesses die Möglichkeit haben, manuell einzugreifen und einzelne Daten nach zu bewerten, oder durch die Software getroffene Entscheidungen zu berichtigen. Hierzu wird ein Interface zur manuellen Bewertung entwickelt,

**Abb. 1** Konzept zur automatisierten Lebensmittelkontrolle im Internet (modifiziert nach Krewinkel et al. 2015)



durch das der User jederzeit die Kontrolle darüber hat, wie groß der Grad der Automatisierung im jeweiligen Anwendungsfall sein soll.

### 3 Technische Umsetzung des Konzepts

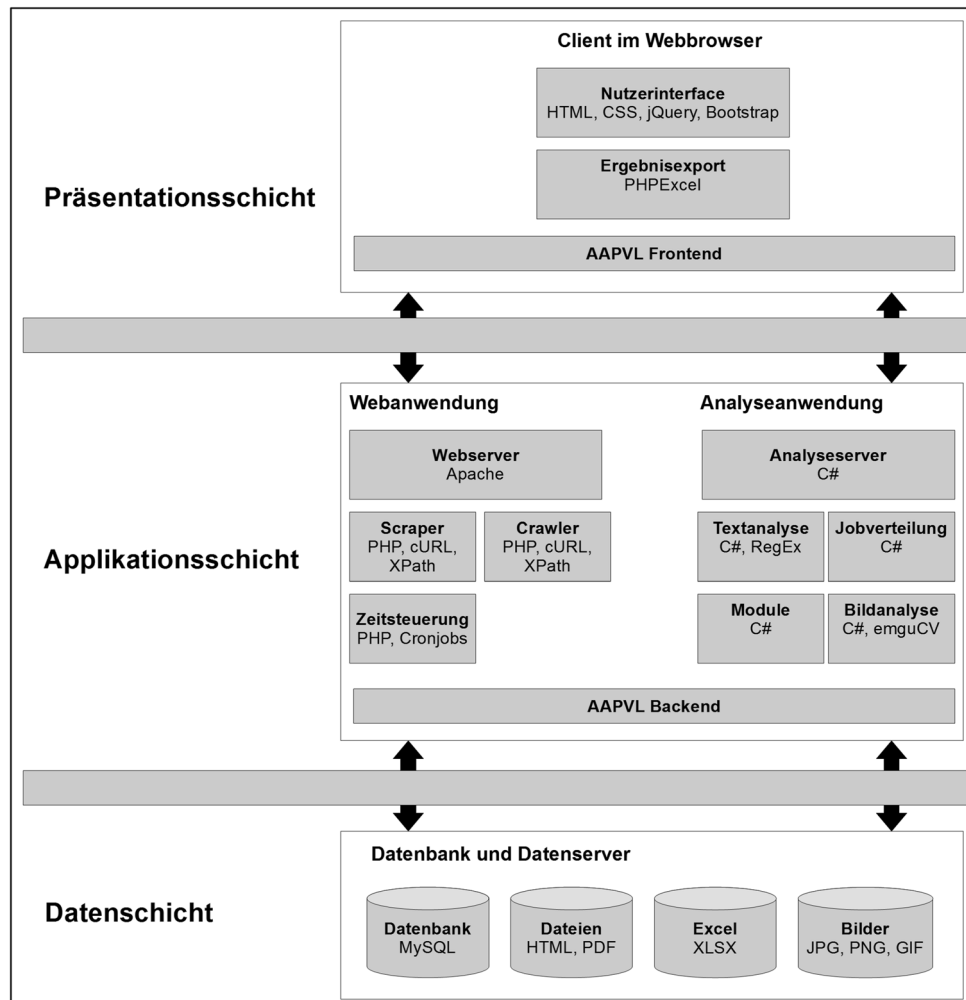
Die Umsetzung des Konzepts zur Lebensmittelkontrolle im Internet benötigt eine Software, die Komponenten und Prozesse anbietet, die die Unmengen an Daten im World Wide Web erfassen und analysieren können. Dafür werden sowohl etablierte Webtechnologien, als auch Analysewerkzeuge benötigt, die in einer flexiblen Software-Architektur zusammenspielen. Für die Sicherstellung der (Weiter-)Entwicklung und Wiederverwendbarkeit des Prototyps wurde daher ein so genanntes Drei-Schichten-Modell gewählt, das Bereiche einer Software in drei Ebenen unterteilt (Abb. 2). Dieses Modell ermöglicht einen flexiblen Austausch von Technologien, ohne die Gesamtfunktionalität zu gefährden (Eckerson 1995; Jing-Feng et al. 2006). So wäre z. B. der Austausch des Frontends auf der Präsentationsschicht durch eine andere grafische Schnittstelle ohne Folgen für die darunterliegenden Komponenten denkbar. Die

folgende Beschreibung der Architektur bezieht sich auf prototypische Realisierungen einzelner Komponenten als Grundlage für die komplette Umsetzung der Software. Bisherige umgesetzte Filterstufen sind die Impressumserkennung, die Shop- und die Lebensmittelshop-Identifikation.

#### 3.1 Datenschicht

Auf der Datenschicht befindet sich die Datenbank, die die zentrale Kommunikationsschnittstelle des Prototyps und den Speicher aller Analyseergebnisse darstellt. Ferner enthält die Datenbank auch diverse Konfigurationen, die für automatisierte Abfragen von Suchmaschinen geeignet sind, z.B. für die zeitgesteuerte Suche. Für die Analysen werden Quelltexte von Webseiten und Bilder benötigt, die ebenfalls dieser Schicht zugeordnet sind. Technologien auf dieser Schicht sind MySQL (Datenbankmanagementsystem), HTML (Beschreibungssprache für die Strukturierung von Webseiten), PDF, XLSX (Dateiformat von Excel-Tabellen) und diverse Bildformate. Excel-Tabellen sind in der Software das Ausgabeformat der Ergebnisberichte, die auf Basis von Datenbankabfragen zusammengestellt werden.

**Abb. 2** Software Architektur des AAPVL Prototyps (modifiziert nach Krewinkel et al. 2015)



### 3.2 Applikationsschicht

Die Applikationsschicht der Software besteht aus zwei Anwendungen, die gemeinsam das Backend der Software bilden. Im Backend laufen alle Prozesse zur Datenerfassung (Speichern von Suchmaschinenergebnissen) und Datenanalyse (Filterstufen) ab. Technisch betrachtet handelt es sich um Serveranwendungen, die über die MySQL-Datenbank kommunizieren und Daten in beide Richtungen austauschen. Während die Webanwendung bisher nur als eine Serveranwendung realisiert ist, können bei den Analyseservern Rechencluster gebildet werden, die zeit- und rechenintensiv Vorgänge durch Verteilung der Last auf mehrere Server beschleunigen. Auf der technischen Ebene der Webanwendung wird ein Apache-Server mit PHP- und MySQL-Unterstützung eingesetzt. Durch die Nutzung der cURL<sup>1</sup>-Programm-bibliothek ist es möglich, Anfragen an

Webseiten zu simulieren und zurückgegebene Inhalte (Quelltexte der Webseiten) zu speichern und auszuwerten. Diese Prozesse werden benötigt, um die Suchergebnisse von Suchmaschinen automatisiert zu erfassen. Dafür wird mit Hilfe von cURL eine Anfrage abgesendet und die Suchergebnisse jeder Suchergebnisseite werden von der ersten bis zur letzten Seite mit Hilfe von XPath<sup>2</sup> (einer Sprache zur Extraktion von Inhalten einer Webseite anhand der DOM-Struktur<sup>3</sup>) gespeichert. Über die grafische Benutzeroberfläche auf der Präsentationsschicht können diese Seiten an die Analyseserver übergeben und ausgewertet werden. Die Analyseserver nutzen Technologien wie reguläre Ausdrücke.<sup>4</sup> Reguläre Ausdrücke sind Zeichenketten, mit denen Muster und Mengen dieser Muster in Texten erkannt werden können. Konkret werden reguläre Ausdrücke im

<sup>1</sup> <http://curl.haxx.se/libcurl/php/>.

<sup>2</sup> <http://www.w3.org/TR/xpath20/>.

<sup>3</sup> <http://www.w3.org/DOM/>.

<sup>4</sup> <http://pubs.opengroup.org/onlinepubs/009695399/toc.htm>.

Prototyp zur Extraktion von Händlerkontaktdaten wie Straße, PLZ, Ort eingesetzt. Andere Analysevorgänge basieren auf Algorithmen aus dem Bereich des maschinellen Lernens, die technisch mit der Programmiersprache C# umgesetzt sind. Dabei werden so genannte überwachte Methoden verwendet (MacKay 2003). Der Software ist eine Menge an klassifizierten Webseiten bekannt (z.B. Shops, Lebensmittelshops, Informationsseiten), die als Wissensbasis bezeichnet wird. Anhand dieser Wissensbasis werden Begriffe im Kontext, wie sie auf den Webseiten vorkommen, extrahiert und als Indikatoren zur Klassifizierung unbekannter Webseiten bestimmt. Dabei richtet sich die Güte des Indikators nach Häufigkeiten und nach Kombinationen der Indikatoren untereinander, für die Wahrscheinlichkeitswerte bestimmt werden. In Bezug auf Online-Shops sind Indikatoren z. B. die Begriffe „Onlineshop“ „Warenkorb“ und „zzgl. MWST.“. Der Vorteil der genannten Methode ist ein selbst lernendes System, was mit jeder richtig klassifizierten Webseite weiterlernt und genauere Ergebnisse bei der Klassifizierung unbekannter Webseiten erzeugt.

### 3.3 Präsentationsschicht

Die Präsentationsschicht ist das, was der Nutzer der Software sieht und bildet damit das Frontend des Prototyps. Alle anderen Komponenten auf den vorgestellten Schichten laufen im Hintergrund der Anwendung. Technisch betrachtet ist die Präsentationsschicht ein Webinterface, das mit HTML, CSS (Beschreibungssprache zum Layouten von Elementen auf einer Webseite, z. B. Farben von Text) und jQuery<sup>5</sup>

(Programm-bibliothek basierend auf JavaScript für Echtzeit-Webapplikationen) entwickelt wird. Nutzer können sich über das Interface authentifizieren und Rechercheaufträge anlegen, Suchmaschinen automatisiert abfragen und Analyseprozesse starten. Sie können Ergebnisse nachbewerten und Ergebnisberichte anfordern.

## 4 Fazit

Das AAPVL Forschungsprojekt hat die technologischen Grundlagen für die Entwicklung eines automatisierten Analyseverfahrens für die Lebensmittelkontrolle im Internet geschaffen. Das APPVL Konzept liefert einen effektiven Beitrag, um auf effizienter Weise eine hohe Lebensmittelsicherheit im Onlinehandel zukünftig gewährleisten zu können.

## Literatur

- Eckerson W (1995) Three tier client/server architecture: achieving scalability, performance, and efficiency in client server applications. *Open Inform Syst* 10(1):46–50
- Jing-Feng L, Li Y, Chen P (2006) A unified architecture model of web applications. *J Shanghai Univ* 6(3):221–227
- Krewinkel A, Sünkler S, Lewandowski D, Finck N, Tolg B, Kroh LW, Schreiber GA, Fritsche J (2015) Concept for automated computer-aided identification and evaluation of potentially non-compliant food products traded via electronic commerce. *Food Control*. <http://dx.doi.org/10.1016/j.foodcont.2015.09.039>. Zugriff am 9. Oktober 2015
- MacKay D (2003) *Information theory, inference and learning algorithms*. Cambridge University Press, Cambridge
- Wagner W, Wiehenbrauk D (2014) *Study: cross channel revolution im Lebensmittelhandel*. Ernst & Young GmbH, Germany

<sup>5</sup> <https://jquery.com/>.