

Entwicklung einer Suchmaschine für Verbraucher im Bundesamt für Verbraucherschutz und Lebensmittelsicherheit

Dr. Thomas Firley

Dr. Soumaya Lhafi

Bundesamt für Verbraucherschutz und Lebensmittelsicherheit

www.bvl.bund.de

Entwicklung einer im Internet jedem Verbraucher zugänglichen Suchmaschine für die Bereiche Verbraucherschutz und Lebensmittelsicherheit mit folgenden Zielen

- Verbesserung der Verbraucherinformation
- Besseres Verständnis der Verbraucherwahrnehmung
- Informationsgewährung, §5 VIG
- Gezielte und regelmäßige Auswertung des Informationsverhaltens der Verbraucher zur
- Risikofrüherkennung zur
- Verbesserung des Verbraucherschutzes und der Lebensmittelsicherheit
- Unabhängigkeit der Suchmaschine

Grundidee zum Aufbau der Suchmaschine

- **Moderierter Katalog der zu durchsuchenden Internetseiten**
- **Themen:**
 - Lebensmittel, Kosmetik, Textilien, Tabak, Bedarfsgegenstände, Futtermittel, Pflanzenschutz, Tiergesundheit, Gentechnik, Verbraucherschutz
- **Auswahlmöglichkeit nach verschiedenen Kategorien**
 - Behörden, wissenschaftliche Einrichtungen, Wirtschaft, Umwelt- und Verbraucherverbände, Medien
- **Suchergebnisse möglichst aktuell**
- **Gezielte Platzierung auf Verbraucherseiten**

Projektpartner

- BVL und
- RRZN (Regionales Rechenzentrum für Niedersachsen der Universität Hannover)

Projektfinanzierung

- Finanziert durch das BVL (Bereitstellung von 78.000 € im ersten Jahr und von 20.000 € in den Folgejahren)

Zeitplan

- Beginn: Dezember 2007
- Beginn Pilotbetrieb: Januar 2008
- Offizielle Vorstellung: 14. März 2008

Entwicklung und Betrieb der Suchmaschine „Clewwa“

- **BVL**
 - Projektleitung und –beratung
 - Erstellung und Bereitstellung der Web-Site-Listen
 - Erstellung und Bereitstellung von Synonymlisten (geplant)
 - Entwicklung der Auswertungsroutinen der Logfiles
 - Auswertung der Logfiles
 - Kontakt für Verbraucher und Site-Betreiber
- **RRZN**
 - Technische Umsetzung (inkl. Programmierung und Betrieb)
 - Erfassung der Logfiles
 - Routineauswertungen
 - Tägliche Übermittlung an das BVL

Hardware

- **2 Dual Core Server
(Crawler, Beantwortung von Suchanfragen)**

Software

- **Java**
- **Nutch / Lucene**

Zeitplan

- **Crawl-Interwall ca. eine Woche**
- **Bei Medien ca. einen Tag**

Allen Interessierten über das Internet frei zugänglich

Aufrufbar über die Internetseite <http://www.clewwa.de>

Beschränkung der Suche auf Internetseiten mit verbraucherrelevanten Inhalten und konsistenten Informationen

Quellen aus fünf ausgewählten Bereichen

Flexible Vorauswahl von Quellen

Suche in Volltexten und Dokumententiteln

Kombinationen von Suchbegriffen über Boolesche-Verknüpfungen

Clewwa, die schlaue Suchmaschine für Verbraucher - Microsoft Internet Explorer bereitgestellt von IT-Service BWL

Datei Bearbeiten Ansicht Favoriten Extras ?

Zurück Zurück Suchen Favoriten

Adresse <http://www.clewwa.de/>

 Bundesamt für
Verbraucherschutz und
Lebensmittelsicherheit

[Wie sucht Clewwa?](#) [Erweiterte Suche](#) [Glossar](#) [Kontakt](#) [Impressum](#)

 **Clewwa**
Die schlaue Suchmaschine für Verbraucher

Suchen auf den Internetseiten von:

- Behörden
- wissenschaftlichen Einrichtungen
- Wirtschaft
- Umwelt- und Verbraucherverbände
- Medien

Zeitstempel (Datum und Uhrzeit der Eingabe)

Anfragende IP-Adresse (anonymisiert, Hash-Wert)

Suchbegriffe und -kombinationen

Ausgewählte Suchbereiche

Art der Suche (einfache / erweiterte Suche)

Ergebnisse der Suche (Fundstellen)

Vom Nutzer aufgerufene Fundstellen in Verbindung mit Suchbegriffen

Reihenfolge der Fundstellenaufrufe

Ergebnisse der Parallelsuche im Bereich „Medien“

Routineauswertungen:

Sortieren der Suchbegriffe nach Häufigkeit der Eingabe

Kombination von Suchbegriff, Zeitstempel, IP-Hash und Reihenfolge der Fundstellenaufrufe

Eingruppierung der Suchbegriffe anhand der BVL-Synonymlisten

Suchbegriffe mit Ergebnissen aus dem Bereich „Medien“

Suchbegriffe ohne Ergebnisse aus dem Bereich „Medien“

Suchbegriffe ohne Ergebnisse

Erkennen der Suchmotivation der Nutzer unter Berücksichtigung der Schwarmtheorie

- **Auswertung der Begriffsinhalte und deren Bezug**
- **Vergleich der Suchbegriffe mit der Reihenfolge der Fundstellenaufrufe**
- **Abgleich der Anfragen / Ergebnissen mit in den Medien aufgetretenen Begriffen**
- **Berücksichtigung der Aktualität der Fundstellen**
- **Hinweise durch häufige Suchanfragen oder Anfragen ohne Suchergebnis**
- **Erkennung von Trends im Suchverhalten der Nutzer durch tägliche Erfassung und Auswertung**

→ **Risikofrüherkennung**

Katalogumfang

- **Ca. 2,3 Millionen indizierte Seiten**

Zugriffszahlen

- **Gesamt 166 576 Zugriffe (14. März bis 13. Mai)**
- **Zwischen 10 000 und 1000 Zugriffe pro Tag**
- **Durchschnitt 2730 pro Tag**

Besonderheiten

- **Dynamische URLs**
- **Robots.txt mit Crawl-Delay**
- **Unerwünschte Seiten**
- **Verärgerte Nutzer**
- **Falsche Erwartungen**
- **Presse**



**Danke für Ihre
Aufmerksamkeit!**

