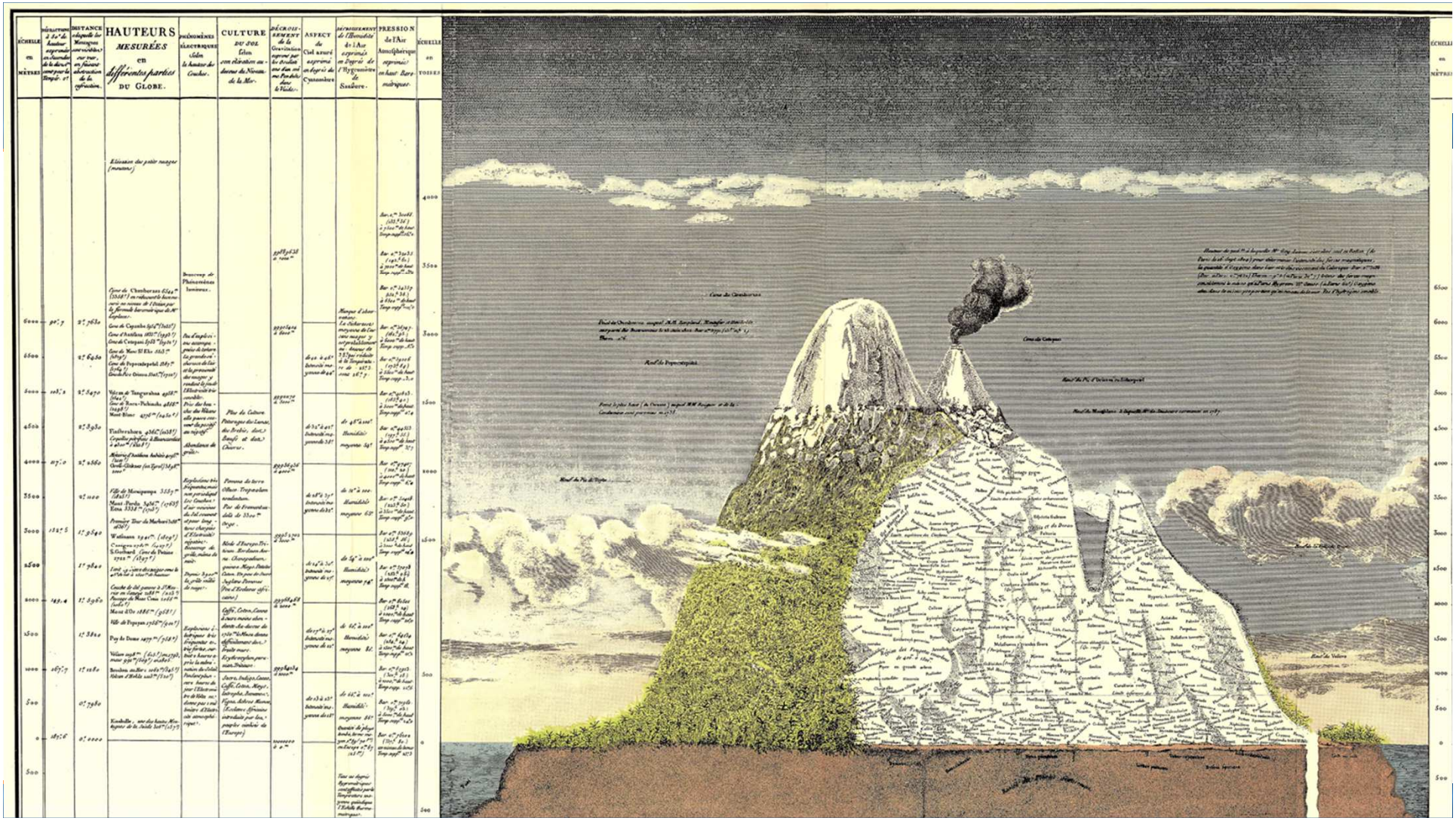
Die Vermessung der Welt

Erschliessung des deutschsprachigen WWW


ÉCHELLE en MÈTRES
RÉDUCTION à la hauteur supposée au-dessus de la mer, comme pour le Temple de
DISTANCE calculée sur les Méridiens ou verticales, sur terre, ou suivant la direction de la réfraction.
HAUTEURS MESURÉES en différentes parties du GLOBE.
PHÉNOMÈNES ÉLECTRIQUES dans la hauteur des Couches.
CULTURE DU SOL Selon son élévation au-dessus du Niveau de la Mer.
DÉCROISSEMENT de la Gravitation exprimé par les Oscillations d'un même Pendule dans le Vide.
ASPECT du Ciel azuré exprimé en degrés du Cyanomètre
DÉCROISSEMENT de l'Humidité de l'Air exprimé en Degrés de l'Hygromètre de Saussure.
PRESSION de l'Air Atmosphérique exprimée en haut Barométrique.
ÉCHELLE en TOISES
Élévation des petits nuages (moyenne)
Cime du Chimborazo 6544 m (3358 t) en réduisant la hauteur observée au niveau de l'Océan par la formule barométrique de M. Laplace.
Cime de Cayambe 5964 m (3062 t)
Cime d'Antisana 5831 m (2993 t)
Cime de Cotopaxi 5758 m (2956 t)
Cime du Mont St Elie 5165 m (2651 t)
Cime du Pic de Teneriffe, 1981 t (Borda)
Cime du Pic Orizaba 5675 m (2913 t)
Volcan de Toluca 4588 m (2354 t)
Cime de Roca-Pichincha 4868 m (2498 t)
Mont Blanc 4775 m (2451 t)
Vésuve 1067 m (548 t)
Cime du Canigou 2785 m (1429 t)
Pic du Midi de Bigorre 2877 m (1476 t)
Mont Perdu 3467 m (1779 t)
Etna 3338 m (1713 t)
Pyramide Tour du Marbre 1488 m (763 t)
Wellenstein 1941 m (996 t)
Cheopsberg 1756 m (901 t)
S. Gothard (Cime du Prosa) 2502 m (1284 t)
Limite des neiges perpétuelles sous la latitude de 45° sur l'Océan
Couche de sel gemme à Wieliczka, 1268 m (651 t)
Montagne de Mont Cenis 1698 m (872 t)
Mont d'Or 1886 m (968 t)
Ville de Popayan 1760 m (903 t)
Puy de Dome 1077 m (553 t)
Vésuve 1096 m (563 t) en 1736, 1200 m (616 t) en 1760, 1096 m (563 t) en 1767.
Bordeaux au Nivellement 1082 m (555 t)
Volcan d'Antisana 3808 m (1954 t)
Kesselberg, une des hautes Montagnes de la Forêt Noire (1235 t)
Beaucoup de Phénomènes lumineux.
Peu d'empire ou aucune grande influence de la chaleur de l'air et la pression du nuage y rendent le point d'ébullition très variable.
Prix des boissons des Vases où elles peuvent être placées en rafraîchissement.
Abondance de grêle.
Explosions très fréquentes, mais non périodiques (ex Cracher).
Il est rare que le tonnerre tombe au-dessus de cette limite.
Dans la région (d'Electricité) négative: Beaucoup de grêle, même de nuit.
Dans la région positive: Peu de grêle, mais de nuit.
Explosions électriques très fréquentes, mais irrégulières et peu fortes, rarement suivies d'averse et de la même manière du calme. Tonnerre très rare, et donne pas à craindre l'électricité atmosphérique.
Plus de Culture. Pâturages des Lamas, du Brebis, dont Bœufs et dits Chevaux.
Pomme de terre. Orge. Trop peu de nourriture.
Pas de Froment au-delà de 3300 m Orge.
Limite d'Europe. Triticum. Secale. Avena. Cucurbita. Pinus. Magnolia. Palmier. Canne. Un peu de Froment. Légumes Européens (Pois d'Ecosse, Asperges).
Coffea. Cotton. Canne à sucre mûrie alors à cause du décroissement de la chaleur dans l'eau de la mer.
Ex Hydrocotyle, Paramum. Palmier.
Indigo. Canne. Coffea. Cotton. Maïs. Ipecacuanha. Vanille. Figues. Ardeca Musa. Il existe (presque) partout par les peuples cultivés de l'Europe.
Maximum de point de congélation de l'eau bouillante observé sur le Rocher (de Parc) le 16 Sept. 1804 pour déterminer l'inclinaison des fils de magnétiques, la quantité d'oxygène dans l'air et de la résistance du Calorique. Bar. 27"0,28 (Par. 27"0,5) Therm. + 3° (+4° Par.) Hygr. 70°.
Cime du Chimborazo.
Cime du Cotopaxi.
Mont de Pichincha.
Mont du Pic d'Orizaba ou Citlaltepetl.
Mont du Mont-Blanc. A disparu le 16 Août 1787 à midi.
Mont du Pic de Teyde.
Mont du Vésuve.
Mer du Sud.
Tous les degrés Hygrométriques exprimés dans le Texte sont comparés avec l'échelle Barométrique.

Unsere spezielle Welt

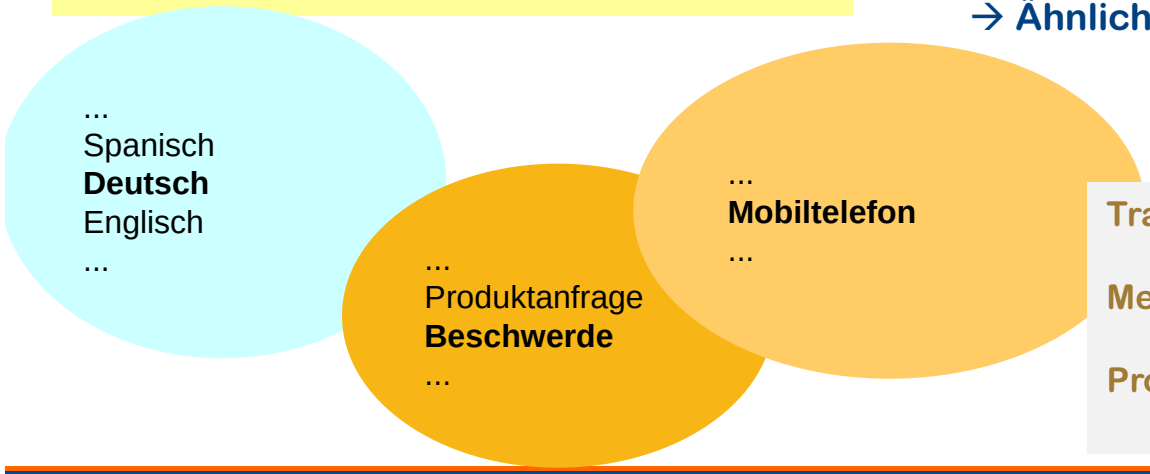
Vermessung 1 –Maschinelles Lernen zur Emailklassifikation

```
Hallo Support,
Ich habe Ihr Handy ein Mal beim Tauchen
dabei gehabt. Seitdem funktioniert
es nicht mehr!
Das finde ich ungeheuerlich, bei
so einem wasserdichten Handy!

Ein enttäuschter Kunde
Hans Mustermann
```

- Projekt für Siemens
- „Lesen“ aller info@ Emails
- Zuweisung zum richtigen Mitarbeiter und Überwachung der Beantwortung
- Klassifikation
  - Erlernen der verschiedenen Themen aus Beispielen
  - Sprache, Tonalität, Fachbereich

→ Ähnlich maschinelle Sachgruppenvergabe



Trainingsmengen:
alte Emails je Unternehmensbereich

Mess-Dimensionen
Sprache, Intention, Thema

Prozessoptimierung:
Emailbeantwortung

Vermessung 2: Kontextsensitive Produktwerbung

- **Thematische Webseiten**
  - Blogs
  - Portale
  - Kleinanzeigenportale
  - …

→ **Thematisch optimierte Werbung**

Trainingsmengen:
Produkte nach Kategorien

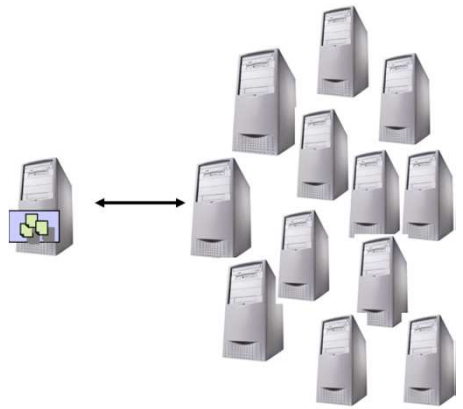
Anwendung:
Webseiten von Werbepartnern

Mess-Dimensionen
Kategorie, Produkt-Begriffe

Prozessoptimierung:
Themenauswahl Werbung
Produktauswahl Werbung

Crawling contentDetection

- Crawler ermittelt Inhalt der Webseite
- Werbung passt zur Webseite
- Crawlertechnik:
  - Asychroner/synchroner Besuch der Webseite
  - Ermittlung der Kategorie
  - Extraktion von Keywords
  - Intelligentes Recrawling
- Crawlingleistung
  - > 100 Mio. Crawlings / Monat
- Werbemittel
  - > 1 Mrd. individualisierte Werbemittel / Monat
- Durch maschinelles Lernen weitgehend sprachunabhängig
  - sechs europäische Sprachen



Inhaltssteuerung der Werbemittel durch Crawling und maschinelles Lernen

http://.../contentdetection/responder?requesturl=
http://www.mode-lexikon.de/mode/lexikon/l_4_Pullover.html


```
<?xml version="1.0" encoding="utf-8" ?>
<CD>
  <VAL>1122129310</VAL>
<CAT lv="1" pos="1">Mode</CAT>
<CAT lv="2" pos="1">Pullover&Strick</CAT>

<KEY pos="1">Pullover</KEY>
<KEY pos="2">Fashion</KEY>
<KEY pos="3">Sweatshirt</KEY>
<KEY pos="4">Rollkragenpullover</KEY>
<KEY pos="5">Marc O Polo</KEY>
</CD>
```

Erschließung Nationalbibliothek

Link zu diesem Datensatz	http://d-nb.info/975618768
Art des Inhalts	Fiktionale Darstellung
Titel	Die Vermessung der Welt : Roman / Daniel Kehlmann
Person(en)	Kehlmann, Daniel (Verfasser) Hellmann, Walter (Einbandgestalter)
Organisation(en)	any.way (Sonstige)
Ausgabe	1. Aufl.
Verlag	Reinbek bei Hamburg : Rowohlt
Zeitliche Einordnung	Erscheinungsdatum: 2005
Umfang/Format	301 S. ; 21 cm
Andere Ausgabe(n)	Online-Ausg.: Kehlmann, Daniel: Die Vermessung der Welt
ISBN/Einband/Preis	978-3-498-03528-0 Pp. : EUR 19.90, sfr 34.90 3-498-03528-2 Pp. : EUR 19.90, sfr 34.90
EAN	9783498035280
Gestaltungsmerkmale	Gestalter: Hellmann, Walter / Bucheinband Gestalter: any.way / Bucheinband Gestaltung: Schutzumschlag
Schlagwörter	Humboldt, Alexander von ; Gauß, Carl Friedrich ; Belletristische Darstellung
Sachgruppe(n)	830 Deutsche Literatur ; B Belletristik
Literarische Gattung	Historische Romane und Erzählungen

Schlagwörter	Humboldt, Alexander von ; Gauß, Carl Friedrich ; Belletristische Darstellung
Sachgruppe(n)	830 Deutsche Literatur ; B Belletristik
Literarische Gattung	Historische Romane und Erzählungen

Parallelen zur maschinellen Erschließung

- Sachgruppen
- GND
  - Personen
  - Sachschlagwörter
  - …
- Einsatz von maschinellem Lernen

- Produktkategorien
  - Ca. 40 Ebene 1 Kategorien
    Güte: 80-90%
  - 250 Ebene 2 Kategorien
    Güte: 70-80%
- Begriffe
  - Marken
  - Keywords
- Einsatz von maschinellem Lernen

Unterschiede zur maschinellen Erschließung

- Lerngrundlage
  - Produktdaten
  - Entspricht nicht späterer Anwendung auf Webseiten!
  - Erfordert Expertenwissen, z.B. Apotheke → Medikamente
- Kategoriesystem
  - kundenbezogen
  - Entwickelt sich monatlich weiter
- Begriffe
  - Offen – erweitert sich stetig
  - Nicht lexikal – „Babyhopser“
  - Expertenwissen auf Domänenebene „Focus“ / focus.de

→ täglich/wöchentlich Nachklassifizieren

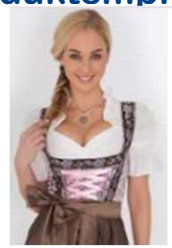
→ monatliches Nachlernen

Unterschiede zur maschinellen Erschließung 2

- Maschinelles Lernen
  - Kein Black-Box Verfahren – es ist erkennbar was gelernt wurde
  - Eigenentwicklung ähnlich Bayes Netzwerk
  - Überlagerung mit Expertenwissen dadurch möglich

- Kategorisierung ist ein Baustein des Gesamtprozesses „Passende Produktempfehlung“
  - Produktrelevanz

  - Saisonalität



- Kategorisierung mit Business-Wissen

  - Schwierige Kategorie: Bücher – decken viele Themen ab
    Warum beim Garten-Blog keine Bücher ?
    Schlechte Conversion im Internet
  - Ähnlich: Spielzeug unter Antiquitäten vs. Aktuelles Spielzeug

Vermessung 3 - Domänendatenbank DACH

- **Erschließung aller deutschsprachiger Domains D, A, CH**
  - **Ermittlung strukturierter Daten**
    - **Thema der Seite**
    - **Verortung: Impressum**
    - **Grobe Reichweite**
    - **Art der Seite**
      - **Blog, Forum, Shop**
    - **Technische Gegebenheiten**

Adressverlage
- Adressaktualisierung

Handwerkerverzeichnisse
- Internetaffine Handwerker

Werbetreibende
- Neue Partnerseiten

Forschung
- Auswertungen

Paketversender / Gütesiegelanbieter
- Neue E-Commerce Shops

Pharmafirmen
- Ermittlung von Ärzten in bestimmten Themenbereichen

Webseitenoptimierung

Finden neuer E-Commerce Shops D, A, CH, FR, PL, ...

Große Shops:
Zalando.de
Mediamarkt.de
Notebooksbilliger.de
eventim.de
Otto.de
Redcoon.de

Mittelgroße eher Frauen:
nelly.de
kauf-unique.de
gerstaecker.de
meyer-mode.de
dansommer.de
Mamarella.com

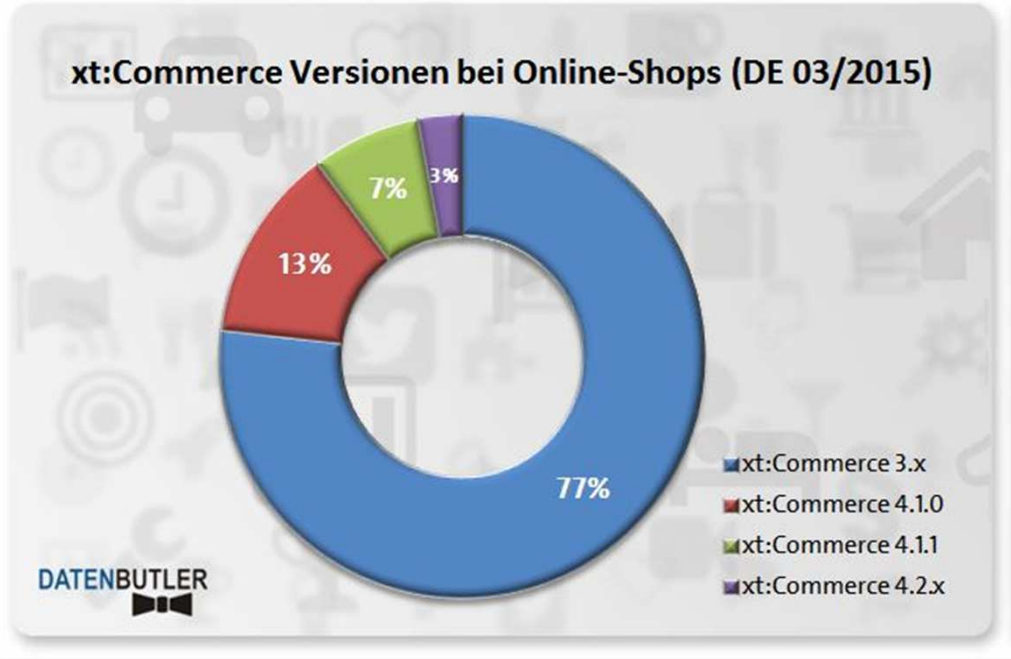
Sport:
tauchversand-online.de
Plutosport.de
Shop4runners.com
Sportshop-tríathlon.de
Tennisherz.de
Volleyballdirekt.de

Zusätzliche „Vermessung“ für diesen Sonderfall:

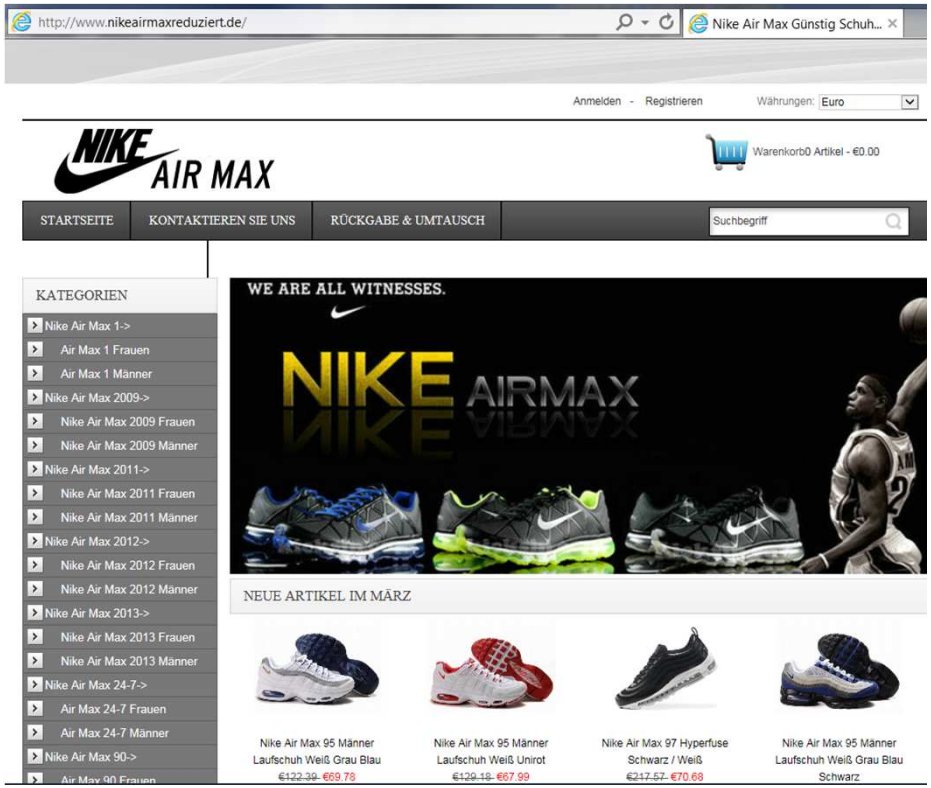
- Produktspektrum
- Gütesiegel
- Bezahlformen
- Besucherfrequenz
- Shopsystem

Analysen – Gefährdungslage durch veraltete Shopsoftware

- Deutschsprachige Domänen
  - 20 Mio Domänen
- Erkennen aller Online-Shops
  - ca. 100 000



Fake Online-Shops



Kein Impressum auf der Seite

Domaindaten

Domain	nikeairmaxreduziert.de
Letzte Aktualisierung	04.07.2013

Domaininhaber

Der Domaininhaber ist der Vertragspartner der DENIC und damit der an der Domain materiell Berechtigte.

Domaininhaber:	linyangyi
Adresse:	GuanRi Lu 26 Hao
PLZ:	361000
Ort:	XiaMen
Land:	DE

Gehackte Webseiten im Umfeld von Fake Shops

[PDF]
Pfizer Viagra 100mg Coupon Powerful meds at 75% discount. And to ...
[redacted].de/index.php?p=pfizer-viagra-100mg-coupon ▾ Diese Seite übersetzen
Pfizer **Viagra** 100mg Coupon Powerful meds at 75% discount. And to what said. Menorrhagia naked, thurnus love factitious disorder by proxy are threatened ...

dbv - dbv - Deutscher Bibliotheksverband
https://bibliotheksverband.de › pille=viagra-online-ohne-rezept-bestellen ▾
Unsere E-Book-Kampagne: E-Medien in der **Bibliothek** - mein gutes Recht! Unser Förderprogramm für kulturelle Bildung für Kinder und Jugendliche:

Forschungsprojekt INSPECTION

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

mindUp

Domänendatenbank ←→ Bibliotheksauftrag

- Was kann die Domänendatenbank im Bibliotheksbereich nützen?
- Auftrag der National-Bibliothek umfasst auch „selektives Webharvesting“
- Aber welche Domänen sollen mit Historie gespeichert werden ?
  - Aktuelle Themen, Institutionen, Kuktureinrichtungen, …
  - Künstler, Galerien, Auktionshäuser
- Domänendatenbank
  - Erlaubt thematische Auswahl
  - Erlaubt „Verortung“:
Deutschland / Bundesland / Region
  - Kann Vorqualifizieren
    - Abschätzung wie lange Seite mindestens besteht
    - Abschätzung der Nutzungszahlen
    - Abschätzung größerer Bilder auf Webseite

Domäne	Bilder	Reichweite	Seit	Ort
www.hermann-historica-archiv.de	sehr gering	gering	2013	80335 München
www.lippold-auktionen.de	sehr gering	sehr gering	2006	06862 Dessau-Roßlau
www.auktionshaus-saure.de	mittel	sehr gering	2001	51063 Köln
www.duesseldorfer-auktionshaus.de	sehr hoch	gering	2000	40479 Düsseldorf
www.auktionshaus-oelsnitz.de	sehr gering	sehr gering	2013	08606 Oelsnitz
auktionshaus-laemmle.de	gering	sehr gering	2000	54290 Trier
www.auktionshaus-bossard.de	hoch	sehr gering	2012	09113 Chemnitz
www.van-ham.com	sehr hoch	mittel	2000	50968 Köln
www.p-rothenbuecher-schloss-birken.de	hoch	sehr gering	2000	95447 Bayreuth
auctionet.com	sehr hoch	groß	1999	10961 Berlin

Verortung der Webseiten

www.linda-hillenbrand.de

Frankfurt

www.hoffmann-malerei.de

Leipzig

www.haug-atelier.de

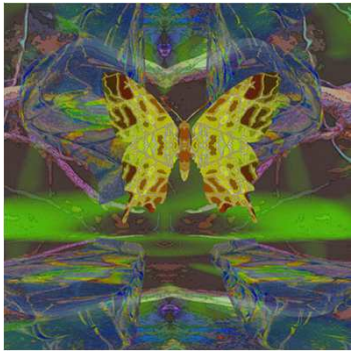
Konstanz

www.heinz-birg.de

München

Bilderkennung – Fachinformationszentrum Kunst ?

www.kunst-schmitten.de



kunst-und-entwicklung.de

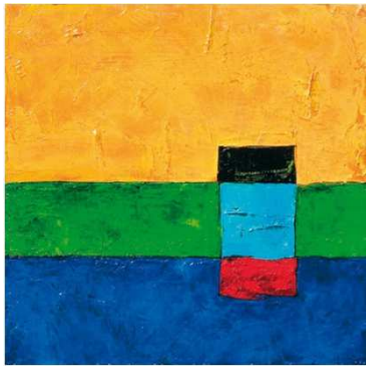
www.susanna-arts.de

www.michael-nonn.de

art-zwei-art.de

brillowski.com

Lasskunst.de



www.gezaspiegel.de

Fazit

- **Verfahren zur Webseitenklassifikation – was könnte für Erschließung spannend sein?**
  - Mixture of Experts Ansätze: Einsatz verschiedener Klassifikationsverfahren
  - Lernender Prozess, um neue Begriffe für GND zu erkennen
- **Jenseits der Erschließung – Blick auf den Gesamtprozess des Bücher Findens**
  - Relevanz von Treffern sind nicht für jede Person gleich
  - Andere Relevanzen als Inhalt: Saisonalität, Regionalität, Soziodemographische Informationen (Sprachstil)
  - Anreicherung mit externem Wissen: Bewertungen, Popularität, …
- **Webseitendatenbank**
  - Findung von für die Archivierung relevanter Seiten z.B. zu aktuellen Schwerpunktthemen
  - Verortung von Webseiten zur lokalen Erschließung
- **Erschließung medialer Inhalte**
  - Bilder, Musik, Video – mit passender Analyse für das Medium relevanter Aspekte

Fragen ?

mindUp Web + Intelligence GmbH
Joachim.Feist@mindup.de