Universität Leipzig

Institut für Informatik

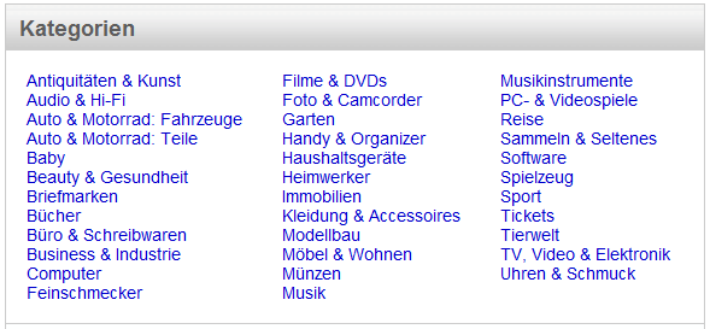
Kategorisierung von Produktangeboten in großen Katalogen

Hanna Köpcke
WDI-Lab

28.10.2010

Produktkatalog

- Systematisch geordnete Sammlung von Informationen zu Produkten oder Dienstleistungen
- Klassifikation anhand eines gemeinsames Merkmales, z.B. Hersteller, Nutzung, Preis, Farbe, Feature
- Beispiele:



Kategorisierung

- Einordnung von Produktangeboten in eine (oder mehrere) Kategorien des Katalogs
- Schnelleres Auffinden der gesuchten Information
- Herausforderungen:
  - Große Anzahl an Kategorien und Angeboten
  - Große Unterschiede zwischen Katalogen
    - Struktur, Bezeichnungen
  - Falsche Zuordnungen
  - Mehrere mögliche Kategorisierungen

(Semi-)automatische Kategorisierung

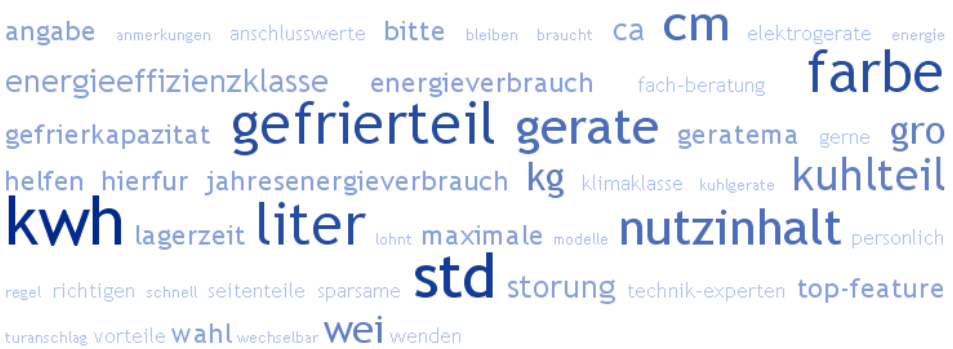
- 2-stufiger Prozess:

1. Training
   - Voraussetzung: Angebote für die die Kategoriezuordnung bekannt ist

2. Kategorisierung
   - Anwendung auf:
     - Kompletten Bestand eines Shops
     - Neue Angebote
     - Aktualisierte Angebote

Trainingsphase

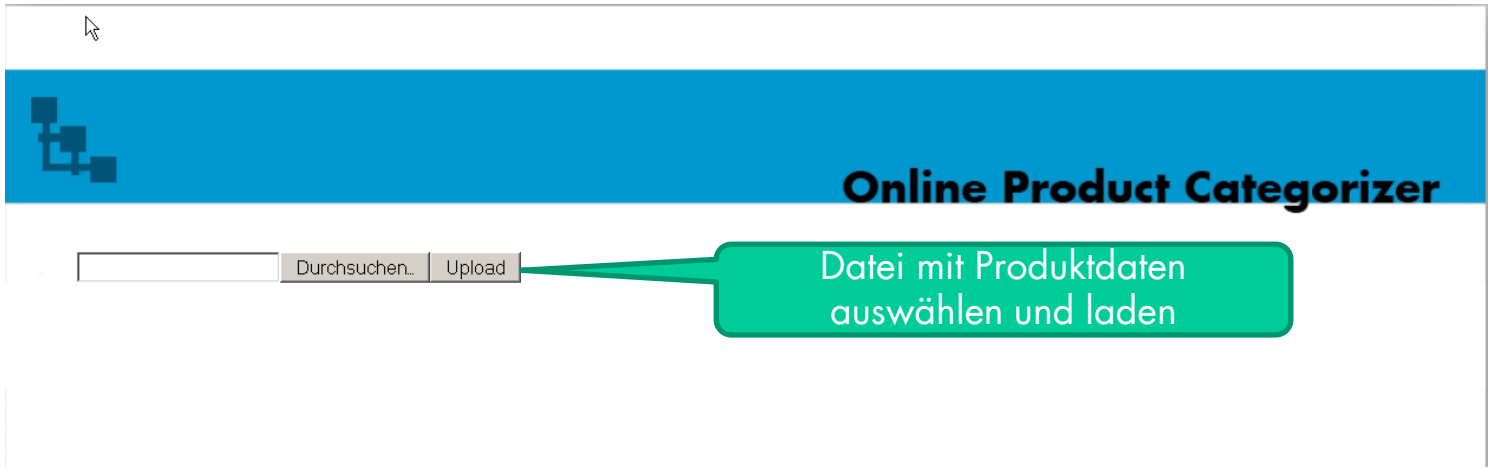
- Tokenisierung
- Entfernung von Stoppwörtern
- Modifizierter Naive Bayes Ansatz
- Verbesserung der Trefferquote durch inkrementelles Lernen



Vorteile

- Reduzierung des manuellen Aufwandes
- Überprüfung bestehender Kategorisierungen
- Unabhängig von Änderungen der Quellkataloge
- Anpassbar an eigenen Katalog
- Skalierbar für große Anzahl von Kategorien und Angeboten

Online Product Categorizer



Starten der Kategorisierung

E:\workspace_hk\Product | Durchsuchen... | Upload | Categorise

Anzeige der Produkte

Anzeige des Katalogs

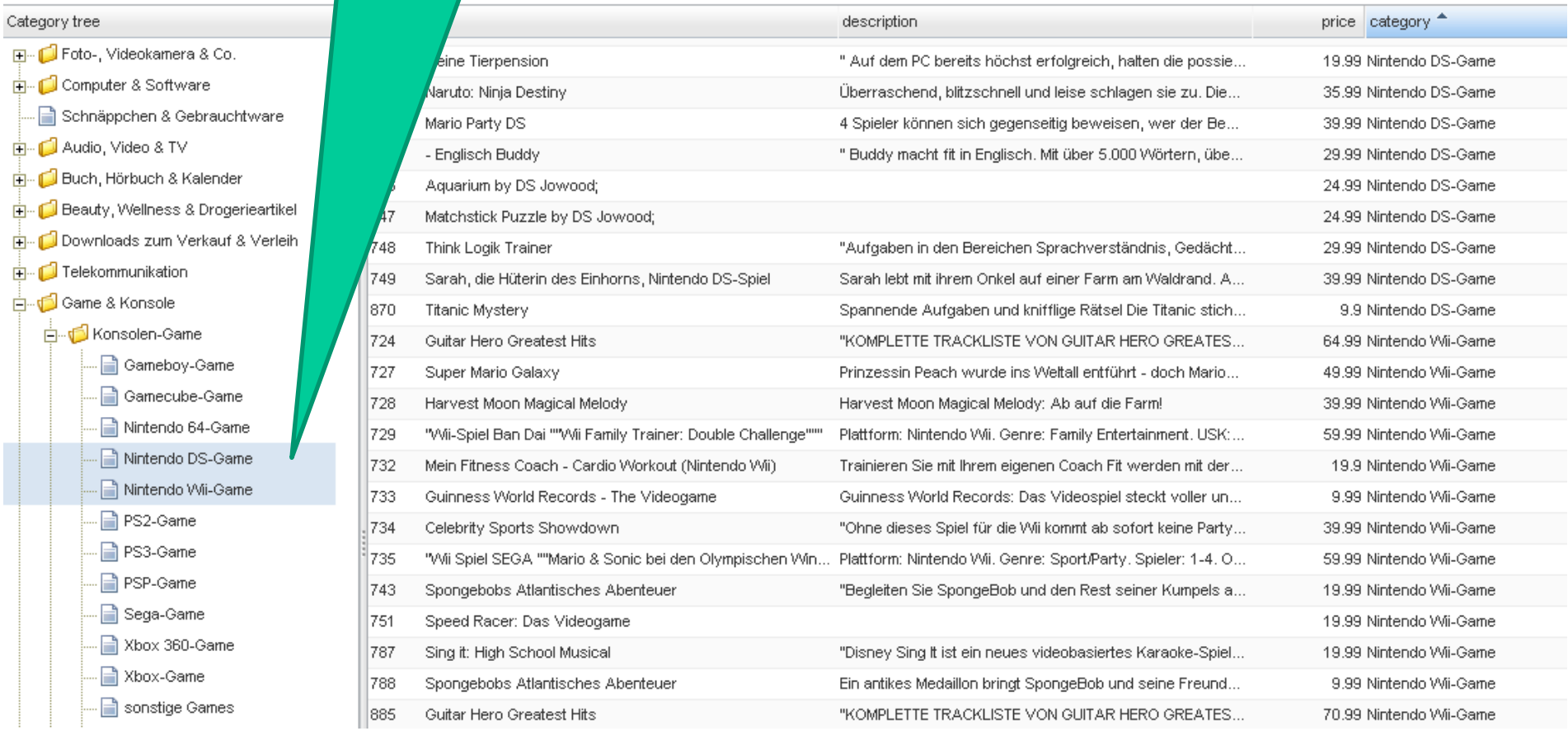
- Category tree
  - Foto-, Videokamera & Co.
  - Computer & Software
  - Schnäppchen & Gebrauchtware
  - Audio, Video & TV
  - Buch, Hörbuch & Kalender
  - Beauty, Wellness & Drogerieartikel
  - Downloads zum Verkauf & Verleih
  - Telekommunikation
  - Game & Konsole
  - Hobby & Spiel
  - Haushalt & Wohnen
  - Sport & Freizeit
  - Mode & Accessoires
  - weitere Produkte

		description	price
		Kino und Hermes reisen durch ...iedene Länder, in denen sie ihnen...	7.5
		...undkurs Marmorieren Anne Pieper;	Verständlich erklärt!Mit Mat...alkundeFotogrundkursenTipps und Tricks...	8.95
2	"Schmidt - Kniffel ""Kniffel-Block"", 3er-Set, für 1440 Spiele "	Mit drei Zusatz-Blocks für 1.440 Spiele.	6.49
3	Hört mal, was da klingt!, m. Audio-CD Conny Frühauf;Christine Werner;	"Hören und Zuhören sind für Kinder Grundvoraussetzungen, um erfolgr...	28.9
4	Aktmalerei, m. DVD Martin Thomas;	In diesem Aufbaukurs zeigt Martin Thomas Schritt für Schritt, wie man a...	19.9
5	Das große Buch der Gouachemalerei Ute Schmidt;	Kreative Entdeckungsreisen Das große Buch der Gouachemalerei lädt...	16.8
6	Mein erstes Bildlexikon Haustiere	"Ein Haustier zu haben ist der größte Wunsch vieler Kinder. Sie finden...	9.95
7	Fenstersticker aus Windowcolor Stefanie Thomas;Heike Roland;	Vom Bauernhof zur Baustelle, aus der Prinzessinen-Welt in die Südsee...	7.5
8	Kinos Reise Keiichi Sigsawa;		7.5
9	Chibi-Manga zeichnen und malen Christopher Hart;	Chibi sind klein, süß und meistens gut gelaunt. Mit Kulleraugen, starken...	14.95
10	Mein erstes Bildlexikon, Tiere im Wald Antje Kleinelümern-Depping;	Die Tierwelt des Waldes ist für Kinder spannend und geheimnisvoll und...	9.95
11	Die Klarinette, m. Audio-CD Rudolf Mauz;	"""Musikinstrumente entdecken - Die Klarinette"" vermittelt anschaulich u...	19.95
12	Das Grüffelo-Puzzle-Buch Julia Donaldson;Axel Scheffler;	Den beliebten und erfolgreichen Bilderbuchklassiker von Axel Scheffler...	15.9
13	Eine kulinarische Entdeckungsreise durch den Chiemgau Petra Wagner;	Der Chiemsee verzaubert von jeher Dichter und Künstler, Einheimische...	29.9
14	Eine kulinarische Entdeckungsreise durch Münsterland und Osnabrück...	In der 1200 Jahre alten Friedensstadt Münster, die neben ihrer geschic...	29.9
15	Die Einkreisung Caleb Carr;	Das Psychogramm eines MördersNew York 1896: Unter Polizeichef Th...	9.95
16	Volevo i pantaloni Lara Cardella;		8.95
17	Die Unruhezone Jonathan Franzen;	Eine Geschichte von mir. Jonathan Franzen ist ein gefeierter Autor für...	19.9
18	Asterix, English editionPt.31 Asterix and the Actress; Asterix und Latra...	Während Asterix und Obelix mit ihren Freunden eine große Geburtstag...	11.2
19	Arcodamore Andrea De Carlo;	"""Nie wieder"", denkt Leo Cernitori nach seiner gescheiterten Ehe, bis...	9.9
20	Xenophobe's Guide to the Americans Stephanie Faul;	"""An enlightened new series, good natured, witty and useful. The Xen...	6.5
21	Winnetou, Engl. ed. Karl May;		34.9
22	The Study of Language George Yule;	Unrivalled in its popularity, The Study of Language is quite simply the be...	21.3
23	Der Junge, der Anne Frank liebte Ellen Feldman;	"Was wäre, wenn ... Europa 1945. Der junge Peter van Pels hat als Ein...	9.95
24	Die Wirklichkeit der tropischen Mythen Fritz J. Raddatz;		12.5

Kategorisierung

id	title	description	price	category
1451	Tiger & Dragon Wang Du Lu;	400 Jahre alt ist das wertvolle Schwert, das den beiden...	5	Action-DVD
1455	Godzilla Terry Rossio;Dean Devlin;Roland Emmerich;Ted E...	In Manhattan herrscht Krieg. Gegen den furchterregendst...	99	Action-DVD
1642	Telekommunikation - Schnurloses Telefon »Gigaset A580...	Schnurloses Telefon »Gigaset A580«, Telefonbuch für bi...		Analog-Telefon
1643	Telekommunikation - Schnurgebundenes Telefon »Eurose...	Schnurgebundenes Telefon »Euroset 5035«, mit Anrufbe...	49.9	alog-Telefon
1644	Telekommunikation - Schnurloses Telefon »Gigaset C385...	Schnurloses Telefon »Gigaset C385 Duo«, mit digitalem A...	98.76	g-Telefon
1645	Telekommunikation - Schnurloses Telefon »Gigaset C380...	Schnurloses Telefon »Gigaset C380 Duo«, mit zusätzlich...	83.29	Analog-Telefon
1646	Telekommunikation - Schnurloses Telefon »Gigaset A585...	Schnurloses Telefon »Gigaset A585«, mit digitalem Anruf...	49.97	Analog-Telefon
1647	Telekommunikation - Zusatz-Mobilteil von Philips,	Zusatz-Mobilteil, für Magic5-Eco-Voice, mit hintergrundbel...	47.59	Analog-Telefon
1648	Telekommunikation - Mobilteil »Gigaset C47H« von Siemens,	Mobilteil »Gigaset C47H« (Mobilteil »Gigaset C47H«)	67.82	Analog-Telefon
1649	Telekommunikation - Schnurloses Telefon »Gigaset E360«...	Schnurloses Telefon »Gigaset E360«, mit großen Tasten,...	77.34	Analog-Telefon
1650	Telekommunikation - Mobilteil »Gigaset S68H« von Siemens,	Mobilteil »Gigaset S68H«, mit Anrufanzeige mit Bild, Farbdi...	71.39	Analog-Telefon
1651	Telekommunikation - Telefon »Euroset 5020« von Siemens,	Schnurgebundenes Telefon »Euroset 5020«, Display: 1-z...	40.45	Analog-Telefon
1652	Easy CA 22 T Home		43.99	Analog-Telefon
1890	DVB-T Antenne, aktiv	DVB-T Antenne, aktiv- aktive Stabantenne für digitalen T...	14.79	Antenne
900	ENDPOINT PROTECTION 11.0 GR	5USER BNDL BUSINESS PACK GR	250	Antiviren-
901	Norton Smartphone Security	5.0/ deutsch/ Vollversion/ DVD Case/ Datenträger: CD	22.4	Antiviren-
902	ANTIVIR PERSONAL EDITION D	PREMIUM 1 YEAR	21.8	Antiviren-
903	Total Protection for SMB Int	25 USER inkl. 1yr Gold Support	647.87	Antiviren-
904	Total Protection for SMB Int	5 USER inkl. 1yr Gold Support	158.06	Antiviren-
905	Total Protection for SMB Int	10 USER inkl. 1yr Gold Support	285.51	Antiviren-
906	Update CA Threat Manager r8.1	25 User - Upgrade from eTrust Antivirus, eTrust PestPatr...	677.92	Antiviren-
907	Total Protection SMB Adv 10U	BOX - 10 USER inkl. 1Jahr Gold Support/ Schutz für: Desk...	534.16	Antiviren-
908	Update CA Threat Manager r8.1	5 User - Upgrade from eTrust Antivirus, eTrust PestPatrol...	135.6	Antiviren-
909	Total Protection SMB Adv 25U	BOX - 25 USER inkl. 1Jahr Gold Support/ Schutz für: Desk...	1218.89	Antiviren-
1990	RAM DDR2-667 2GB/CL5/KIT-2x1024MB	DDR2 2GB / non-ECC / CL5 / 667MHz / ValueRAM / KIT-2x...	60.98	Arbeitsspeicher

Einschränken der Anzeige



		description	price	category
	eine Tierpension	" Auf dem PC bereits höchst erfolgreich, halten die possie...	19.99	Nintendo DS-Game
	Naruto: Ninja Destiny	Überraschend, blitzschnell und leise schlagen sie zu. Die...	35.99	Nintendo DS-Game
	Mario Party DS	4 Spieler können sich gegenseitig beweisen, wer der Be...	39.99	Nintendo DS-Game
	- Englisch Buddy	" Buddy macht fit in Englisch. Mit über 5.000 Wörtern, übe...	29.99	Nintendo DS-Game
	Aquarium by DS Jowood;		24.99	Nintendo DS-Game
47	Matchstick Puzzle by DS Jowood;		24.99	Nintendo DS-Game
748	Think Logik Trainer	"Aufgaben in den Bereichen Sprachverständnis, Gedächt...	29.99	Nintendo DS-Game
749	Sarah, die Hüterin des Einhorns, Nintendo DS-Spiel	Sarah lebt mit ihrem Onkel auf einer Farm am Waldrand. A...	39.99	Nintendo DS-Game
870	Titanic Mystery	Spannende Aufgaben und knifflige Rätsel Die Titanic stich...	9.9	Nintendo DS-Game
724	Guitar Hero Greatest Hits	"KOMPLETTE TRACKLISTE VON GUITAR HERO GREATES...	64.99	Nintendo Wii-Game
727	Super Mario Galaxy	Prinzessin Peach wurde ins Weltall entführt - doch Mario...	49.99	Nintendo Wii-Game
728	Harvest Moon Magical Melody	Harvest Moon Magical Melody: Ab auf die Farm!	39.99	Nintendo Wii-Game
729	"Wii-Spiel Ban Dai ""Wii Family Trainer: Double Challenge"""	Plattform: Nintendo Wii. Genre: Family Entertainment. USK:...	59.99	Nintendo Wii-Game
732	Mein Fitness Coach - Cardio Workout (Nintendo Wii)	Trainieren Sie mit Ihrem eigenen Coach Fit werden mit der...	19.9	Nintendo Wii-Game
733	Guinness World Records - The Videogame	Guinness World Records: Das Videospiel steckt voller un...	9.99	Nintendo Wii-Game
734	Celebrity Sports Showdown	"Ohne dieses Spiel für die Wii kommt ab sofort keine Party...	39.99	Nintendo Wii-Game
735	"Wii Spiel SEGA ""Mario & Sonic bei den Olympischen Win...	Plattform: Nintendo Wii. Genre: Sport/Party. Spieler: 1-4. O...	59.99	Nintendo Wii-Game
743	Spongebobs Atlantisches Abenteuer	"Begleiten Sie SpongeBob und den Rest seiner Kumpels a...	19.99	Nintendo Wii-Game
751	Speed Racer: Das Videogame		19.99	Nintendo Wii-Game
787	Sing it: High School Musical	"Disney Sing It ist ein neues videobasiertes Karaoke-Spiel...	19.99	Nintendo Wii-Game
788	Spongebobs Atlantisches Abenteuer	Ein antikes Medaillon bringt SpongeBob und seine Freund...	9.99	Nintendo Wii-Game
885	Guitar Hero Greatest Hits	"KOMPLETTE TRACKLISTE VON GUITAR HERO GREATES...	70.99	Nintendo Wii-Game