Optimizing the Long Tail

neluand Fachtag 2017


Long Tail

NEW YORK TIMES BESTSELLER

CHRIS ANDERSON

WHY THE FUTURE OF BUSINESS IS SELLING LESS OF MORE

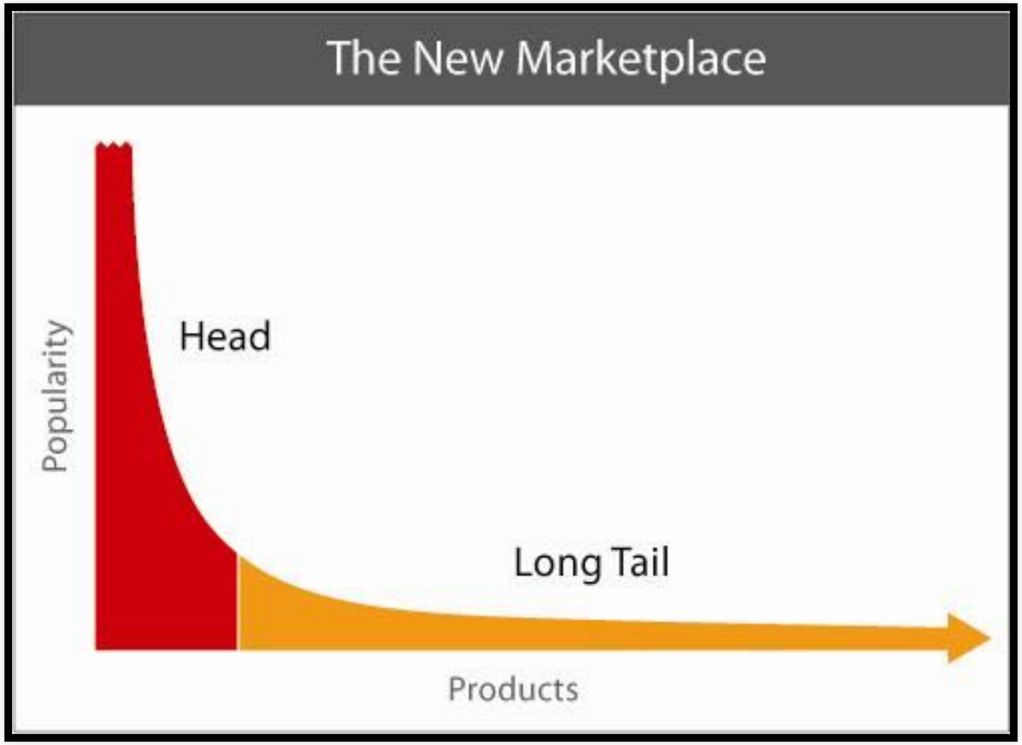
The Long Tail

LONGER

INCLUDES A NEW CHAPTER: THE LONG TAIL OF MARKETING

Long Tail Theory

- 20% der Produkte erzielen 80% der Verkäufe
- Ausstellungsraum kostet Geld
- Begrenzte Verkaufsfläche wird mit Topsellern bestückt
- Aber: Die restlichen 80% Produkte würden mehr Sales generieren, wenn man sie gezielt anbieten könnte


The New Marketplace
Head
Long Tail
Popularity
Products

Im Internet

- Ausstellungsraum kostet hier weniger Geld
- Die Reichweite an Kunden ist höher
- Produkte im Long Tail haben oft eine bessere Marge als Top-Seller
- Mehr Produkte können mehr Kunden günstig angeboten werden

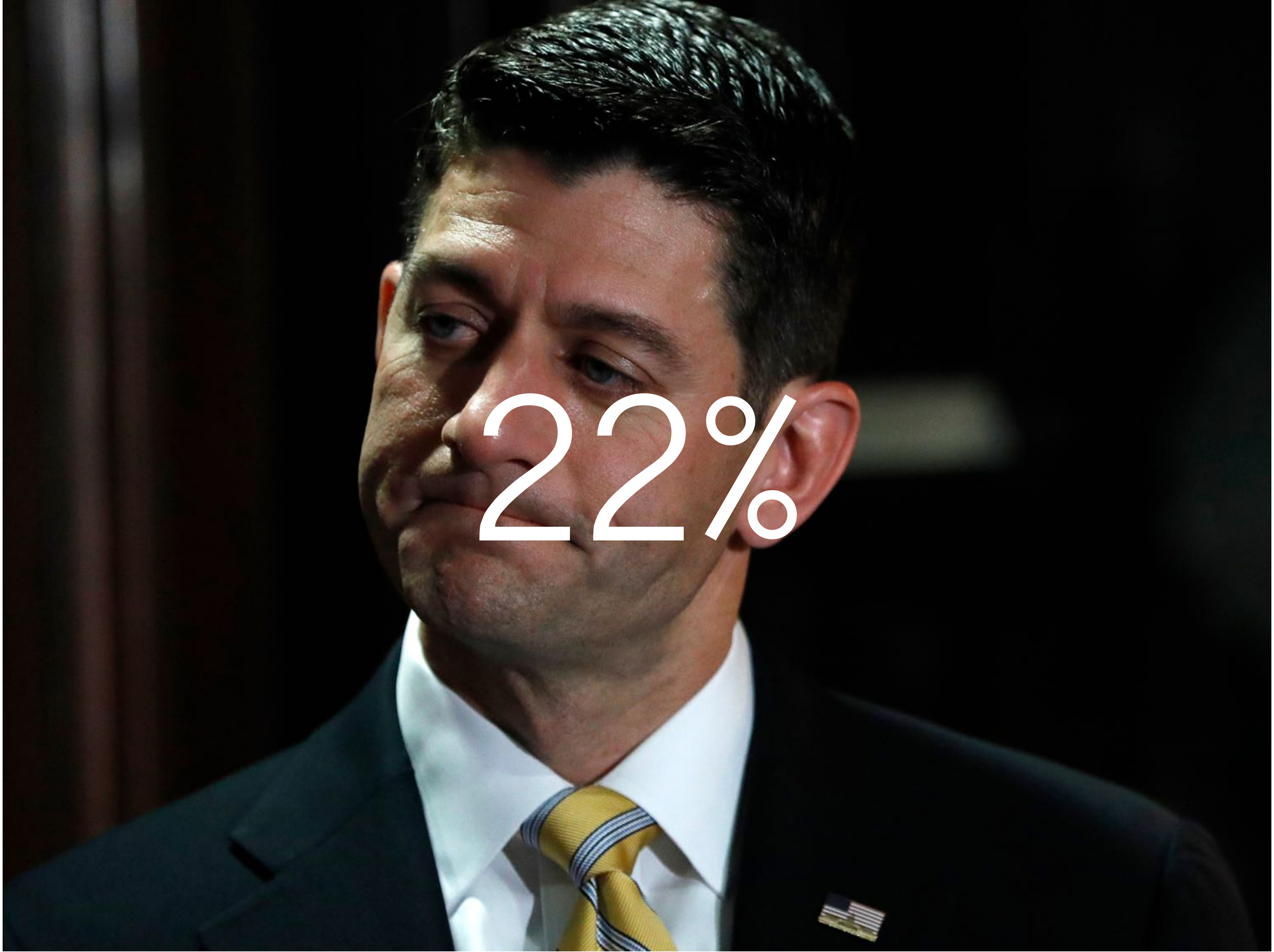

Personalisiertes
Einkaufserlebnis ?

The Consumer Conversation

The experience void between brands and their customers

in association with IBM ExperienceOne


81%
LOW MILES
LIKE NEW
SHARP
STOP
We FINANCE


22%

Kontaktpunkte

- Suchergebnisse
- Reco-Boxen
- Newsletter
- Retargeting

Aber auch:

- Kataloge
- Push-Emfehlungen
- Content-Empfehlungen
- ...

Empfehlungen

FACT-Finder®

Apache Solr

prudsys

rr richrelevance

FREDHOPPER

econda WEB SHOP CONTROLLING

"Big data will spell the death of customer segmentation and force the marketer to understand each customer as an individual within 18 months or risk being left in the dust."

IBM CEO Virginia Rometty

2013

78%

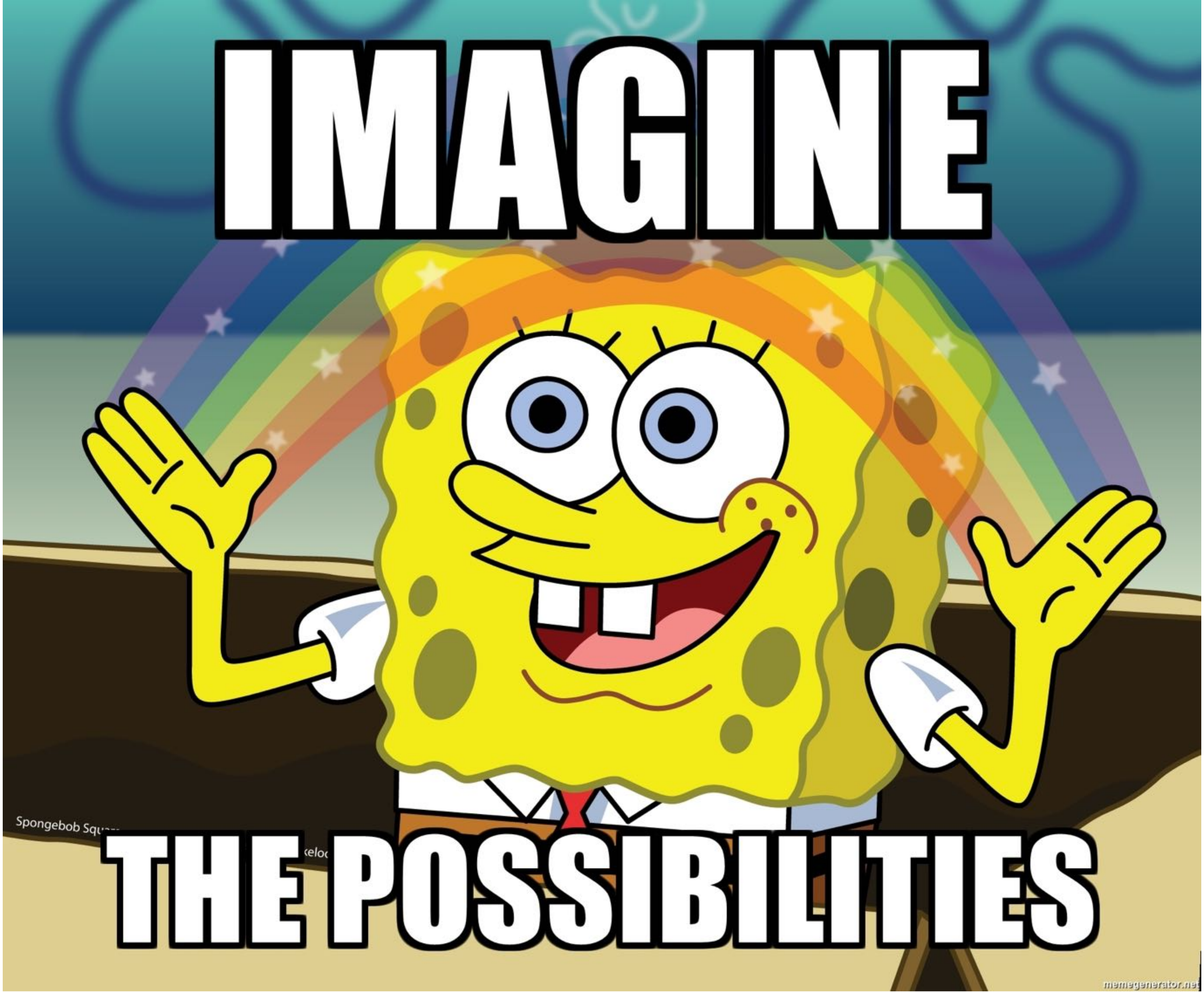
der Kunden fühlen sich nicht individuell verstanden

90 - 96%

verlassen den Shop, ohne zu kaufen

80 - 90%

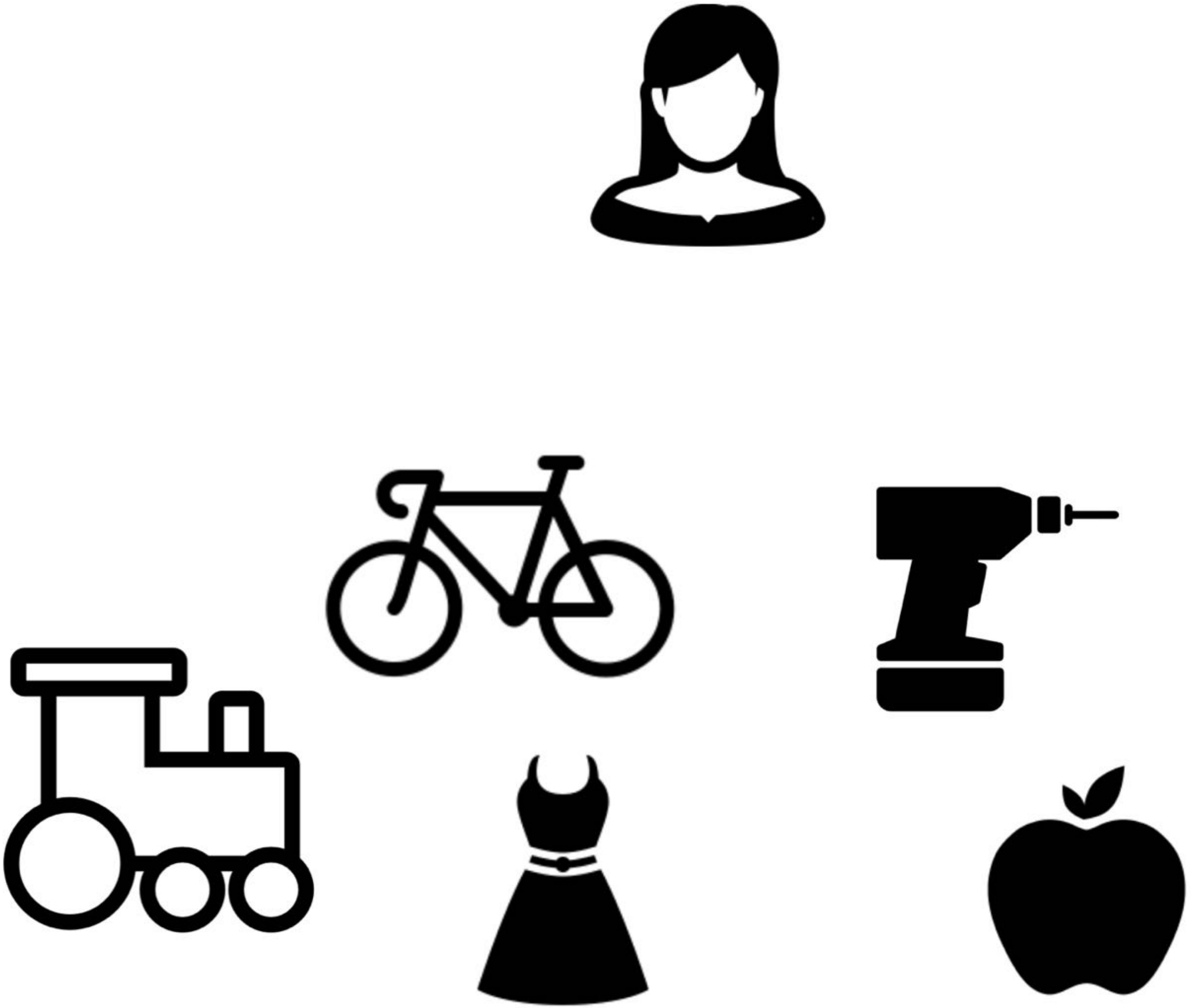
der Artikel verfügen über ungenutztes Verkaufspotenzial

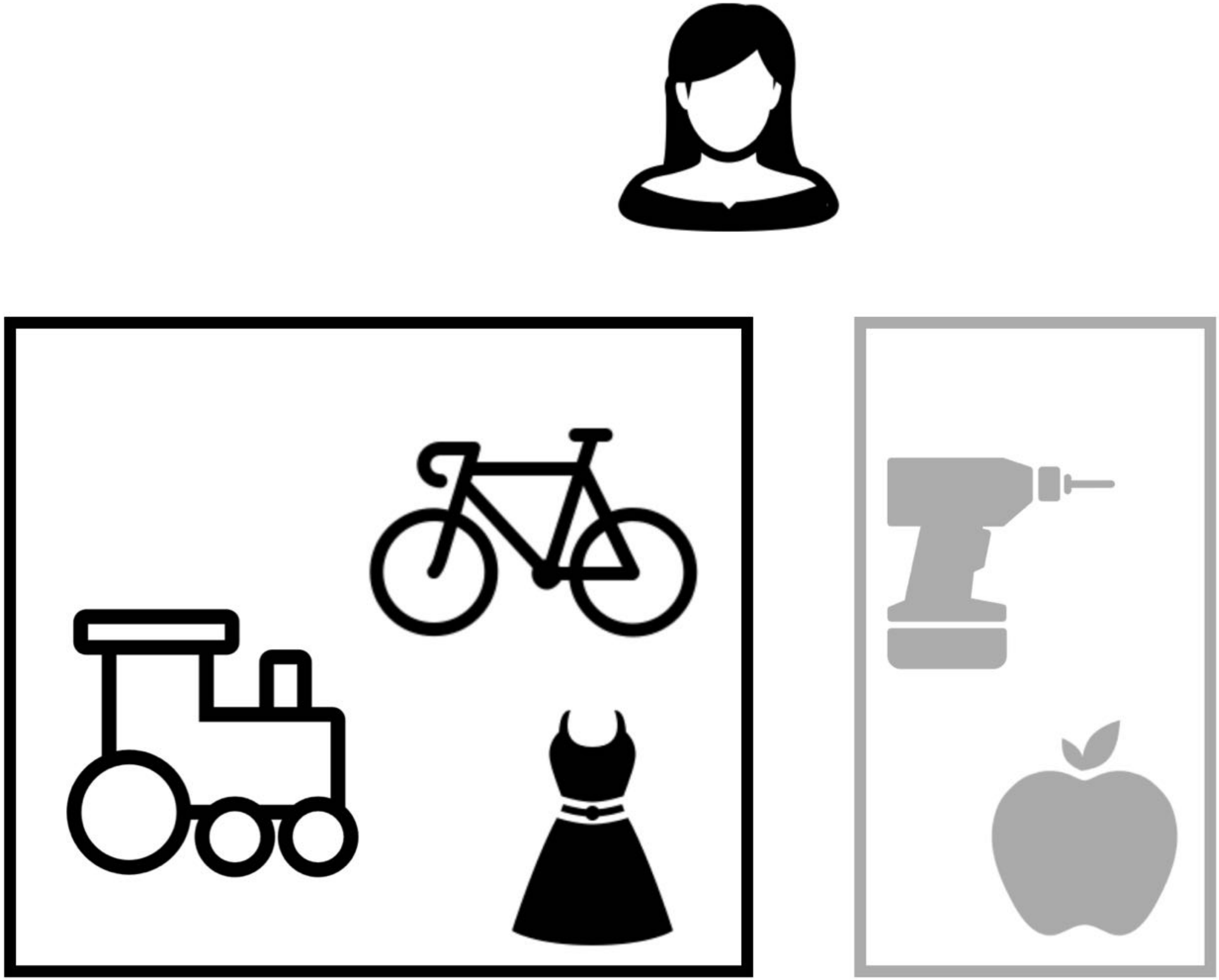

IMAGINE
THE POSSIBILITIES
memegenerator.net

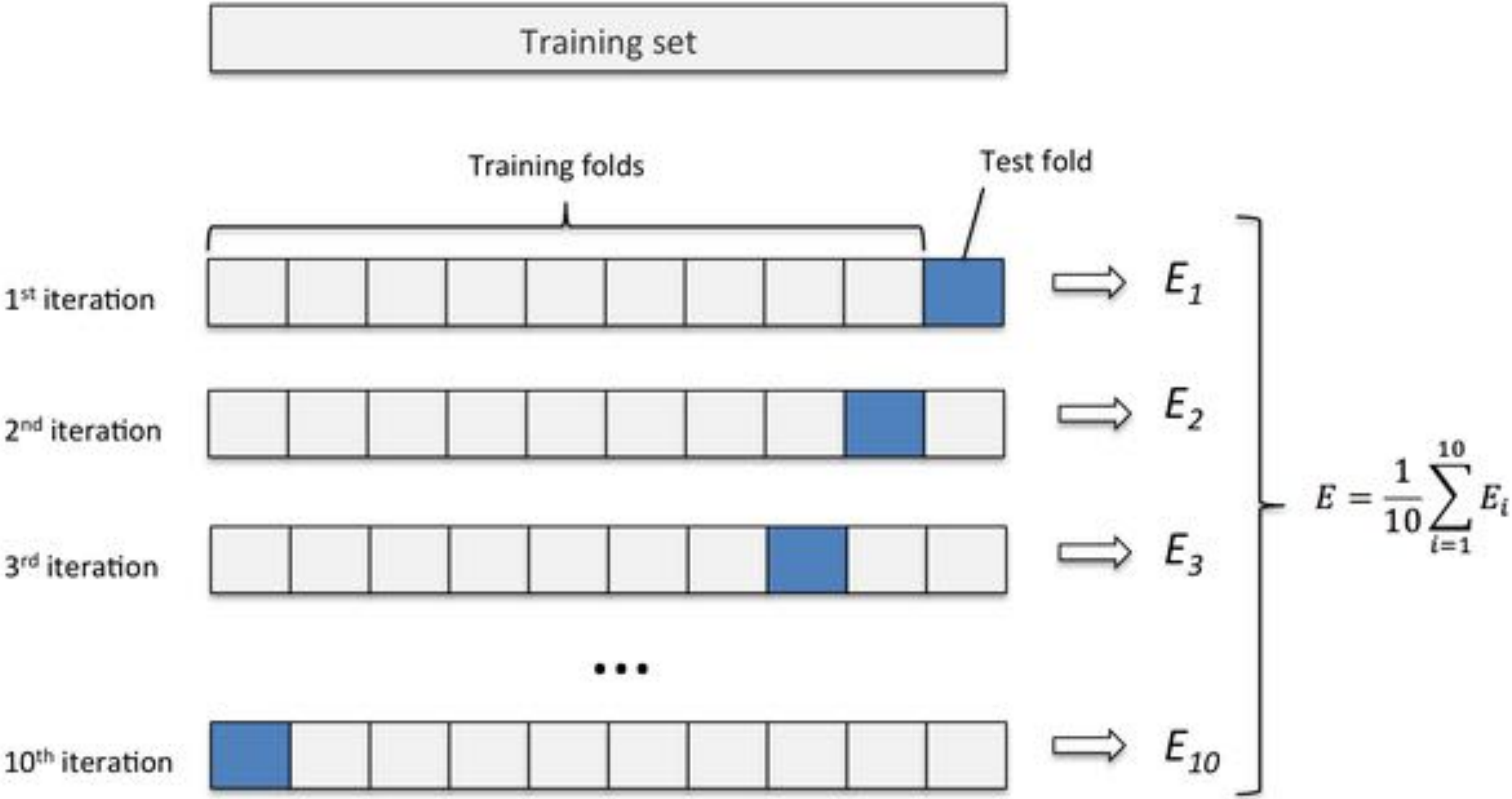

Delta TV

Accuracy

Welche Produkte sind für den User in der nahen Zukunft interessant?

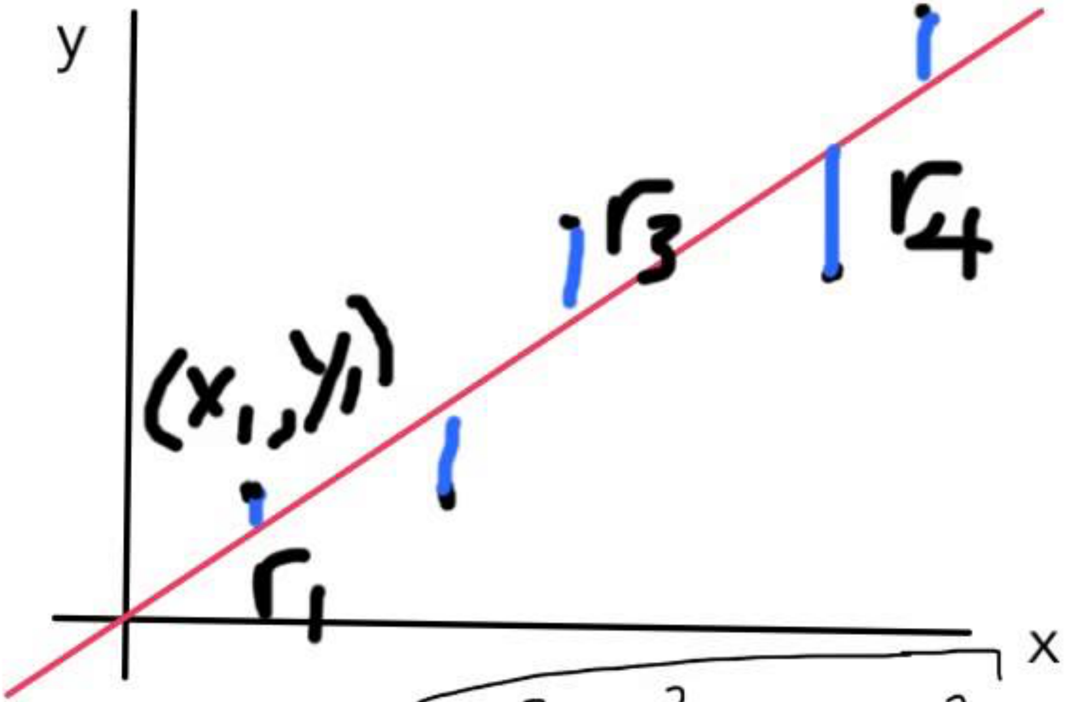






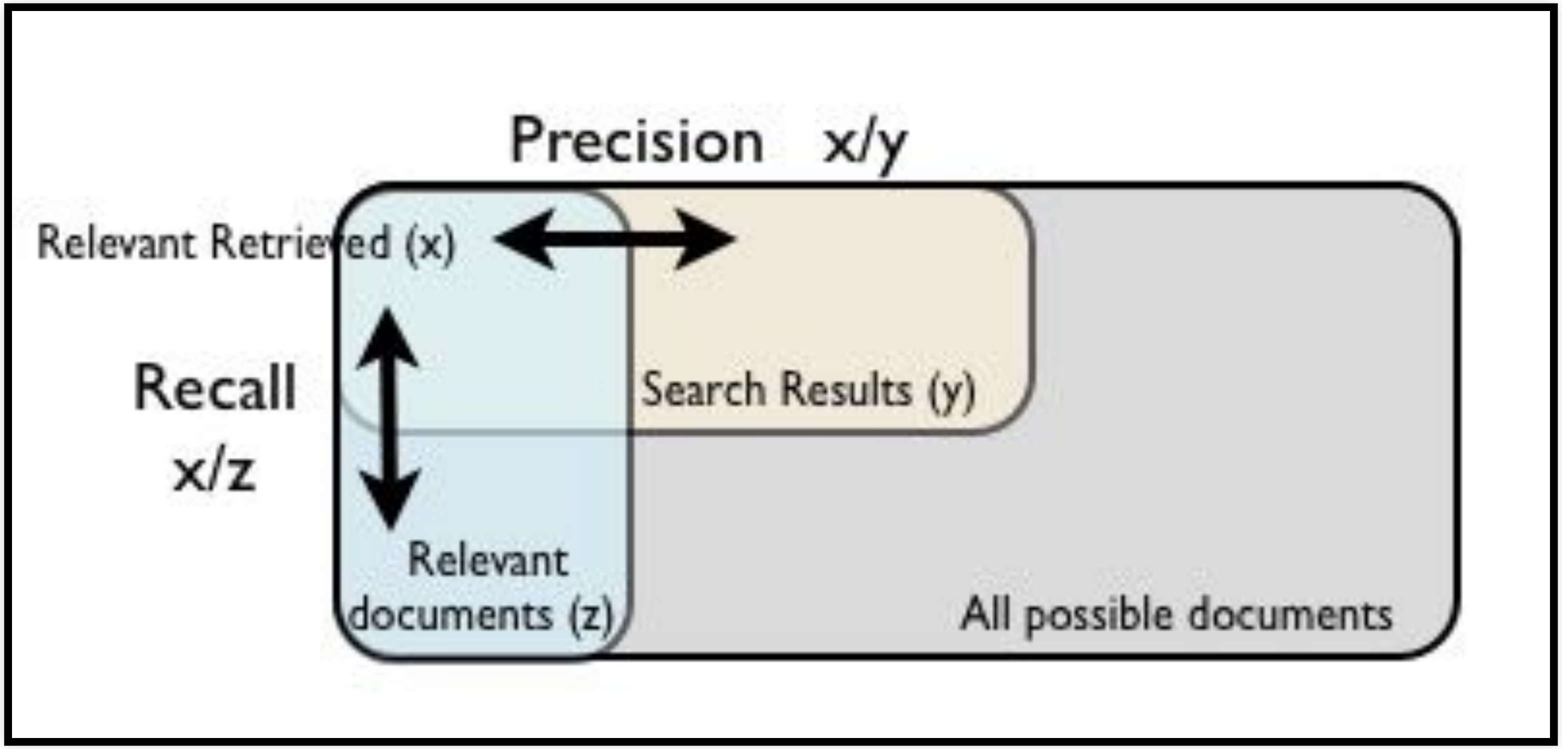
Training set
Training folds
Test fold
1st iteration
2nd iteration
3rd iteration
...
10th iteration
E_1
E_2
E_3
E_{10}
$E = \frac{1}{10}\sum_{i=1}^{10} E_i$

Root Mean Squared Error
(RMSE)



$$RMSE = \sqrt{\frac{r_1^2 + r_2^2 + \cdots + r_n^2}{n}}$$

Precision and Recall

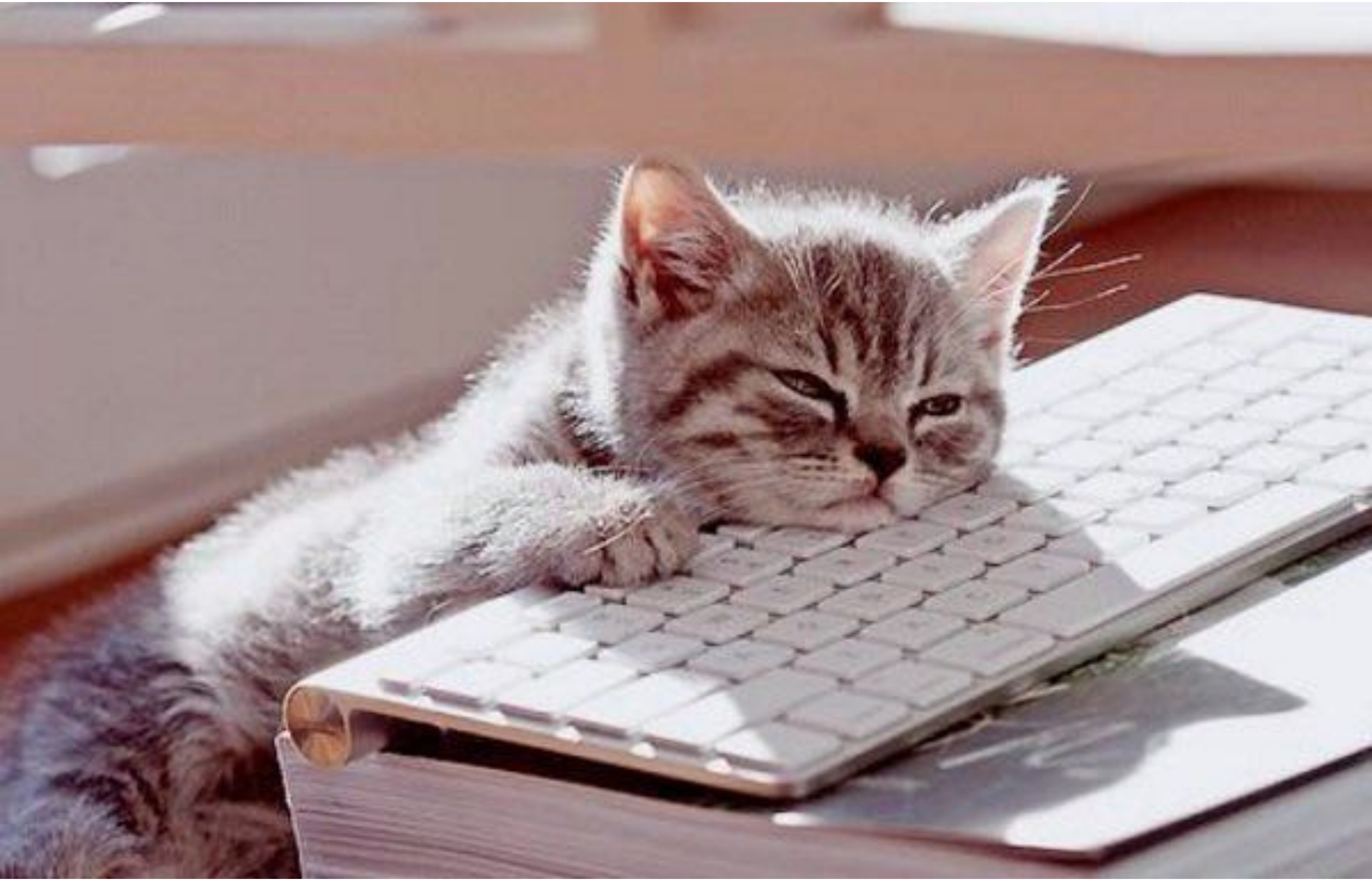



WELL ALRIGHT THEN
SOUNDS GOOD
memegenerator.net


There's more?!

Aber...

- Viele Produktempfehlungen stammen nicht aus einem Empfehlungssystem, sondern werden einfach, z.B. nach Umsatz, gerankt.
- Viele Recommender Algorithmen haben die Tendenz, vor allem populäre Produkte vorzuschlagen. (P. Cremonesi et al., 2010)



Accurate is not always good: How Accuracy Metrics have hurt Recommender Systems

Sean M. McNee
GroupLens Research
Department of Computer Science and Engineering
University of Minnesota
Minneapolis, MN 55455 USA
mcnee@cs.umn.edu

John Riedl
GroupLens Research
Department of Computer Science and Engineering
University of Minnesota
Minneapolis, MN 55455 USA
riedl@cs.umn.edu

Abstract
Recommender systems have shown great potential to help users find interesting and relevant items from within a large information space. Most research up to this point has focused on improving the accuracy of recommender systems. We believe that not only has this narrow focus been misguided, but has even been detrimental to the field. The recommendations that are most accurate according to the standard metrics are sometimes not the recommendations that are most useful to users. In this paper, we propose informal arguments that the recommender community should move beyond the conventional accuracy metrics and their associated experimental methodologies. We

"In essence, we reward a travel recommender for recommending places a user has already visited, instead of rewarding it for finding new places for the user to visit."

"As we have shown in previous work, user satisfaction does not always correlate with high recommender accuracy [McNee 2002, Ziegler 2005]. There are many other factors important to users that need to be considered."

Catalog Coverage

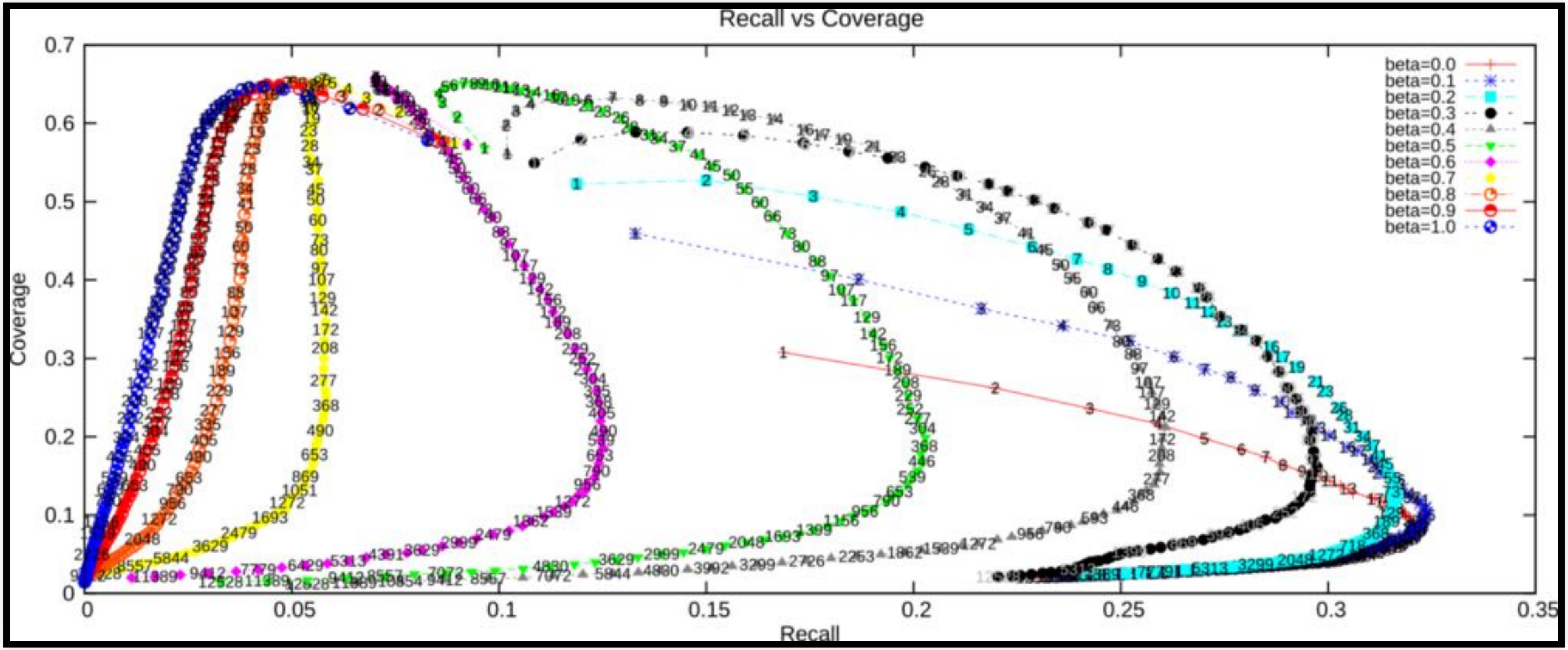
Wie viele Artikel aus unserem Sortiment spielen wir überhaupt als Recommendation aus?

Was tun?

- Nur die Top Seller anzeigen?
- Zufälliges Item anzeigen?

Messbarkeit

%-Coverage vs. Recall



Weighted Coverage

Wieviel % der potenziell nützlichen Produkte wurden als Empfehlung ausgespielt?

(Jannach et al., 2010)

Was heißt nützlich...?


ovelty

Novelty

Deutsch: Neuheit

Ein Produkt, dass der User bisher nicht gesehen hat oder von dessen Existenz er bisher nicht gewusst hat.

Aus Händlersicht folgene Vorteile:

- Produkte aus dem Long Tail platzieren
- Über die Reaktion des Users neue Einsichten in seine Vorlieben gewinnen

Messbarkeit

- Im offline-Training des Modells gut messbar:
  - Zeitliches Splitten der Sales
  - Zusätzliches Belohnen der Empfehlung von Produkten mit Neuigkeitswert


serendipity
(n) finding something good
without looking for it

Serendipity > Novelty


TAJ-MAHAL
TAJ-MAHAL
TAJ-MAHAL
TAJ-MAHAL
TAJ-MAHAL
TIREZ

Novelty

Serendipity

Vorteile

- Kombiniert
  - Catalog Coverage
  - Novelty
  - Nützlichkeit

Risiko

- Sehr subjektiv
- Schlechte Empfehlungen können Vetrauensverlust beim User verursachen

Messbarkeit

- Einfaches Baseline Modell (offensichtliche Empfehlungen)
- Welche Empfehlungen unseres Systems sind nicht offensichtlich?
- Wieviele davon sind nützlich?
- (P. Cremonesi et al., 2010)


Trust

Recommendations funktionieren dann am besten, wenn der User Vertrauen darin hat, dass sie ihm Mehrwert bieten.

Customer Lifecycle

- Neue User: Play it safe!
- Stammkunden: Optimize the Long Tail
- Merke: Unterschiedliche Phasen des Customer Lifecycle haben unterschiedliche Anforderungen

Messbarkeit

- Von allen Faktoren am schlechtesten offline zu trainieren
- Online-Feedback messen:
  - Wie wirkt sich das Einführen von "neuartigen Empfehlungen" auf die Interaktion der User mit den Empfehlungen aus?

Vertrauen aufbauen

- Transparenz der Empfehlungen (Bob Wielinga et al., 2007)
  - Wie sicher ist sich der Algorithmus?
  - Was ist die Berechnungsgrundlage?

Trust Interface (Li Chen et al., 2007)

The most popular product

	Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard drive capacity	Display size	Weight
	-	$2'095.00	1.67 GHz	4.5 hour(s)	512 MB	80 GB	38.6 cm	2.54 kg

We also recommend the following products because

they are cheaper and lighter, but have lower processor speed

	Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard drive capacity	Display size	Weight
	-	$1'499.00	1.5 GHz	5 hour(s)	512 MB	80 GB	33.8 cm	1.91 kg
	-	$1'739.99	1.5 GHz	4.5 hour(s)	512 MB	80 GB	38.6 cm	2.49 kg
	-	$1'625.99	1.5 GHz	5 hour(s)	512 MB	80 GB	30.7 cm	2.09 kg
	-	$1'426.99	1.5 GHz	5 hour(s)	512 MB	60 GB	30.7 cm	2.09 kg
	-	$1'929.00	1.2 GHz	4 hour(s)	512 MB	60 GB	26.9 cm	1.41 kg
	-	$1'595.00	1 GHz	5.5 hour(s)	512 MB	40 GB	26.9 cm	1.41 kg

they have higher processor speed and bigger hard drive capacity, but are heavier

	Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard drive capacity	Display size	Weight
	-	$1'220.49	1.8 GHz	5 hour(s)	1 GB	100 GB	38.1 cm	2.95 kg
	-	$2'148.99	2 GHz	4 hour(s)	1 GB	100 GB	39.1 cm	2.9 kg
	-	$1'379.00	3.3 GHz	2 hour(s)	512 MB	100 GB	43.2 cm	4.31 kg
	-	$2'235.00	1.8 GHz	2.5 hour(s)	1 GB	100 GB	43.2 cm	3.99 kg
	-	$2'319.00	1.7 GHz	4.5 hour(s)	512 MB	100 GB	43.2 cm	3.13 kg
	-	$2'075.00	1.8 GHz	1.67 hour(s)	512 MB	100 GB	43.2 cm	4.4 kg

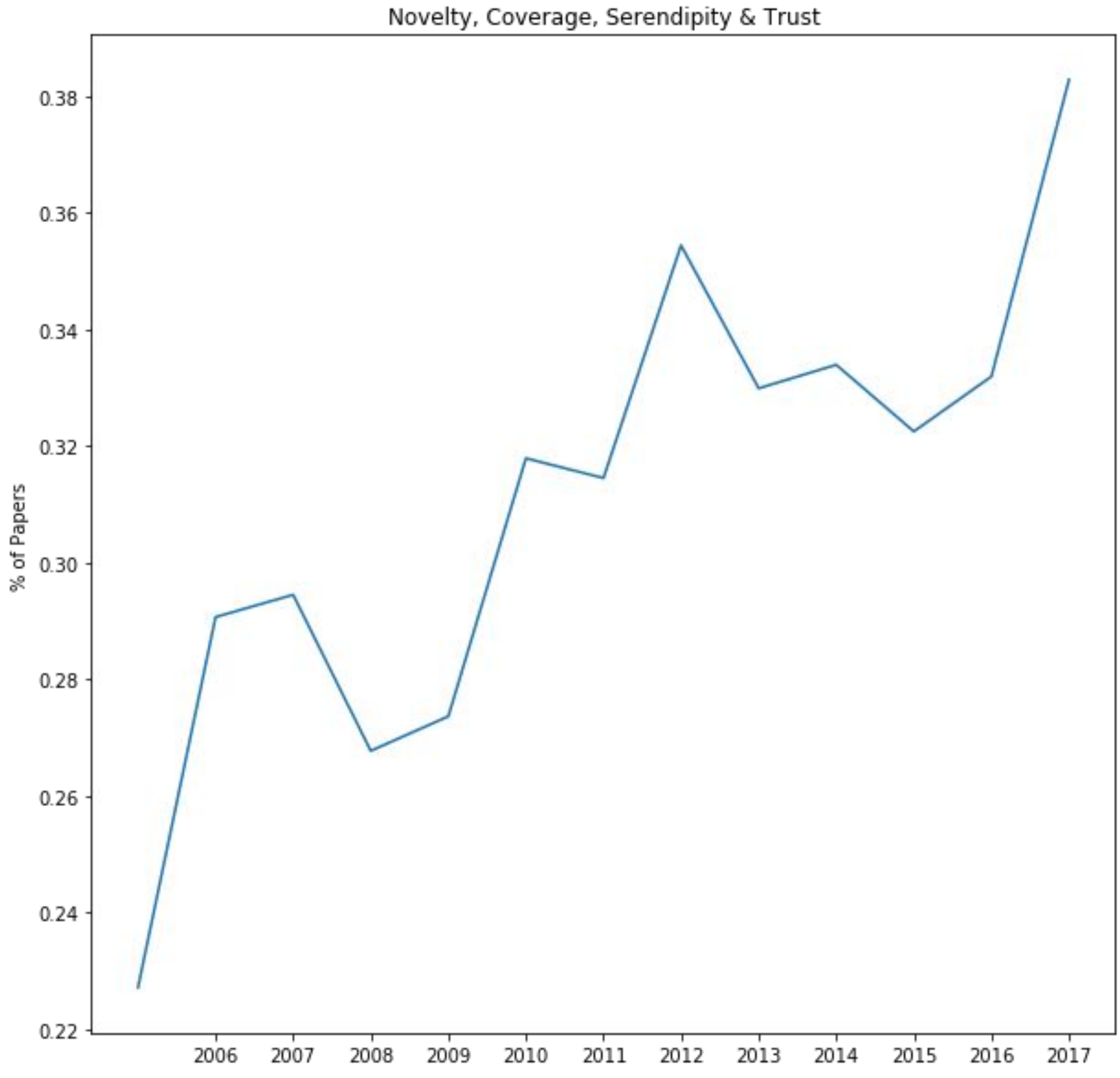
they are lighter and have longer battery life, but smaller display size

	Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard drive capacity	Display size	Weight
	-	$1'529.00	1.7 GHz	6.5 hour(s)	512 MB	80 GB	33.8 cm	1.77 kg
	-	$1'599.00	1.7 GHz	6.5 hour(s)	512 MB	80 GB	33.8 cm	1.91 kg
	-	$1'125.00	1.5 GHz	6 hour(s)	512 MB	80 GB	30.7 cm	2 kg
	-	$2'099.99	1.2 GHz	9 hour(s)	512 MB	60 GB	26.9 cm	1.41 kg
	-	$1'649.00	1.1 GHz	8.5 hour(s)	512 MB	40 GB	26.9 cm	1.36 kg
	-	$969.00	1.2 GHz	6 hour(s)	256 MB	39 GB	30.7 cm	2.22 kg

they are cheaper, but heavier

	Manufacturer	Price	Processor speed	Battery life	Installed memory	Hard drive capacity	Display size	Weight
	-	$1'179.00	3.2 GHz	2 hour(s)	512 MB	80 GB	39.1 cm	3.62 kg
	-	$1'425.00	1.6 GHz	5.5 hour(s)	512 MB	80 GB	39.1 cm	2.86 kg

Fig. 4. The organization interface used in the user evaluation.



Novelty, Coverage, Serendipity & Trust
% of Papers
0.22
0.24
0.26
0.28
0.30
0.32
0.34
0.36
0.38
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017


Okay.

Optimierung von Empfehlungssystemen

Forschung

- Ziel: Allgemein einsetzbare Systeme
- Tiefe Erforschung von generellen Fragestellungen
- Trainieren von Modellen mit "offline"-Datensätzen
- Feedback über Studien mit realen Usern

Praxis

- Spezielle Herausforderungen: Jeder Shop ist anders
- Systeme müssen produktiv Ergebnisse liefern
- Aber: Echtes User Feedback
- Branchenkenntnis
- Iterative Verbesserungen

Best Practice

- Schwacher Zusammenhang zwischen offline und online Performance (Garcin et. al, 2014)
- Balance:
  - Vorselektieren der Modelle offline
  - Optimieren online

Techniken

- Gute KPIs
- Experiment-Design
- A/B-Testing
- Bandit-Algorithmen

GOOD
IS NOT
GOOD
ENOUGH

Literatur

- Paolo Cremonesi, Yehuda Koren, Roberto Turrin: "Performance of recommender algorithms on top-n recommendation tasks", RecSys '10 Proceedings of the fourth ACM conference on Recommender systems Pages 39-46
-
- Mouzhi Ge , Carla Delgado-battenfeld , Dietmar Jannach: "Beyond accuracy: evaluating recommender systems by coverage and serendipity (2010)", RecSys '10 Proceedings of the fourth ACM conference on Recommender systems Pages 257-260

- Bob Wielinga et al: "The effects of transparency on trust in and acceptance of a content-based art recommender", User Modeling and User-Adapted Interaction archive Volume 18 Issue 5, November 2008 Pages 455-496
-
- Pearl Pu, Li Chen: "Trust building with explanation interfaces", IUI '06 Proceedings of the 11th international conference on Intelligent user interfaces Pages 93-100

- Florent Garcin et al: "Offline and Online Evaluation of News Recommender Systems at swissinfo.ch", Proceeding RecSys '14 Proceedings of the 8th ACM Conference on Recommender systems Pages 169-176
-
- Aggarwal, Charu C.: "Recommender Systems", Springer, Erstausgabe 2016