

OFFLINE EVALUATION IN E-COMMERCE SEARCH:

APPLICATIONS AND REQUIREMENTS

About OTTO and otto.de



- Founded in **1949**
- Number of **employees 4,900**
- Revenue in 2018/19 **3.2 billion Euro**
- On average **1.6 million visits** on otto.de per day
- Up to **10 orders per second**
- More than **3 million items** on otto.de
- More than **400 OTTO market partners**
- Approx. **6,800 brands** on otto.de
- Expansion of the business model towards becoming a **marketplace**



OTTO's headquarter in Hamburg

About Us



Jens Kürsten
Tech Lead Search @otto.de



Andreas Wagenmann
Software Developer Search @otto.de

OTTO

~0.9 million

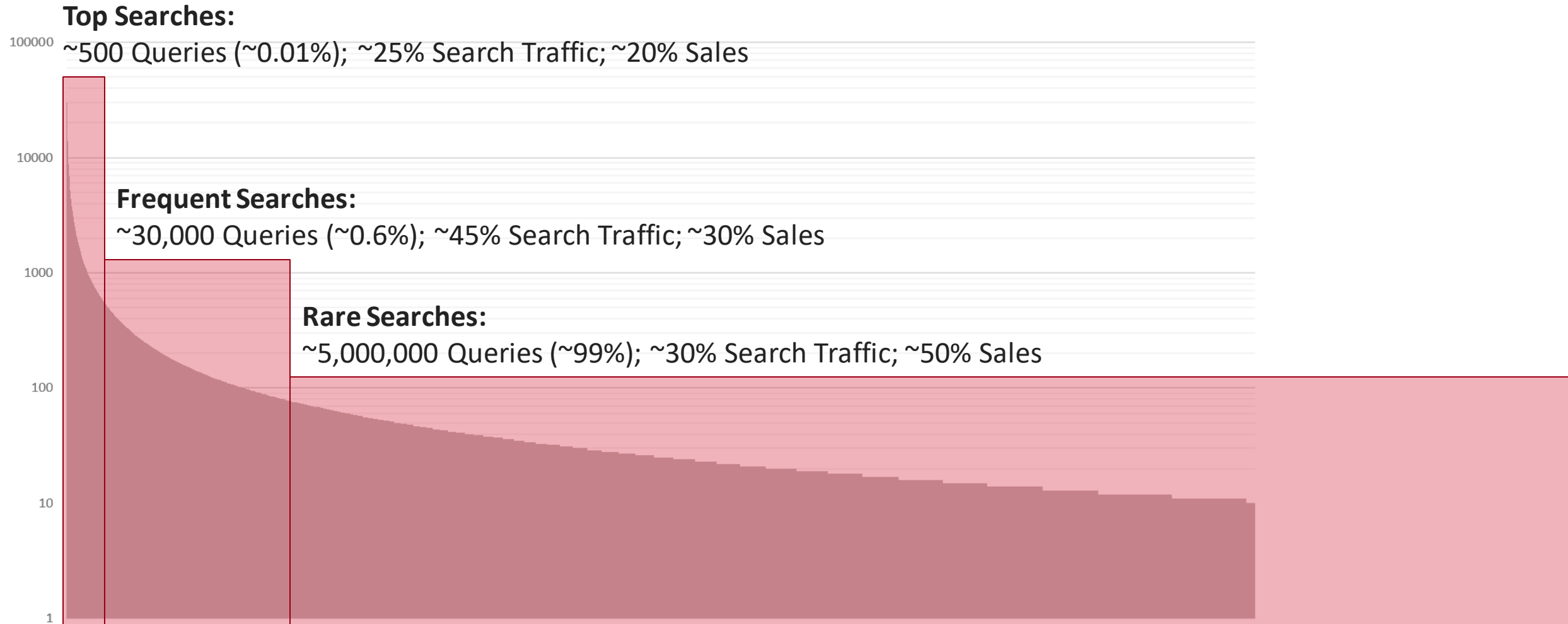
~3 million

~320 million

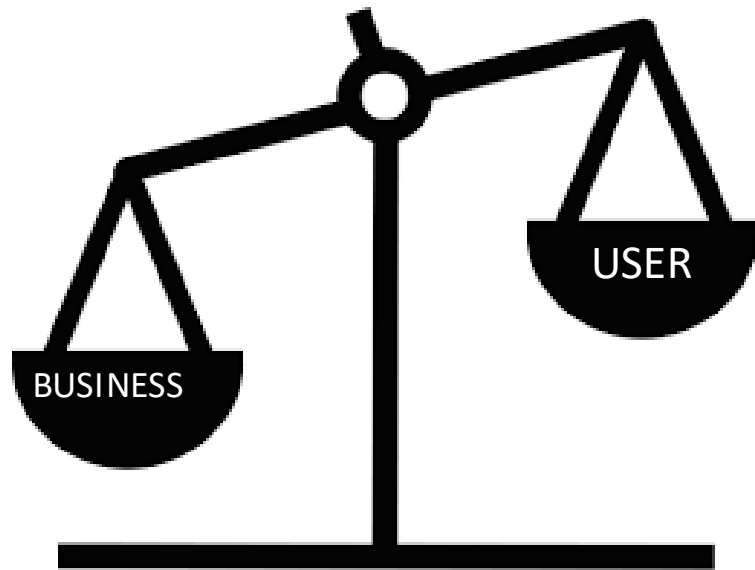
~40 million



What Is the Business Impact of Our Queries?



Key Requirements for Search @otto.de



Relevance @otto.de is determined by

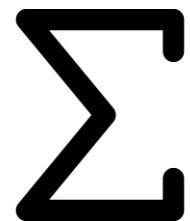
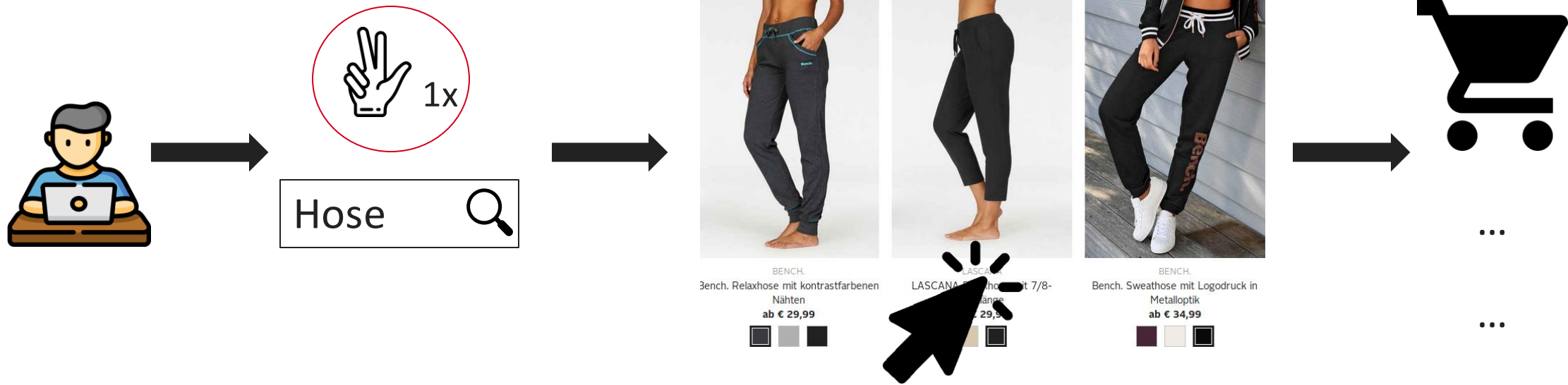
- user queries & product data (quality)
- performance indicators of our products
- category-specific business goals
- user interaction data

Finding the balance between the user intent and the business perspective is a challenge.

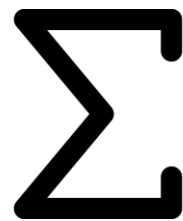


OFFLINE EVALUATION SETUP

Collecting Data



queries
(with hits)

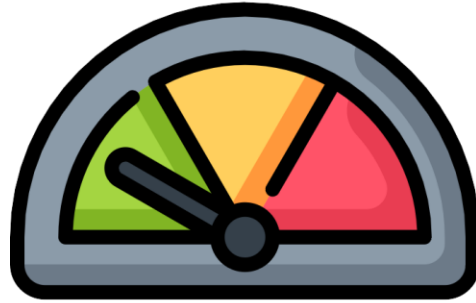
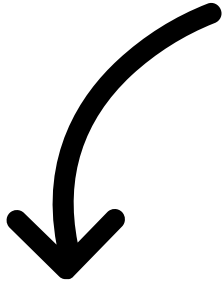


clicks after search
(with positions)



judgements:
query/product/
score

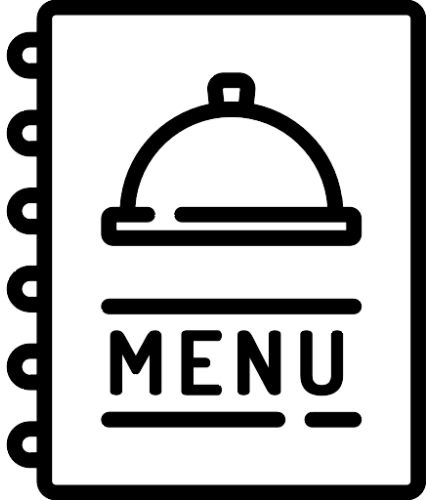
Use Cases & Configurations



ranking evaluation



selection diff



queries



frequencies

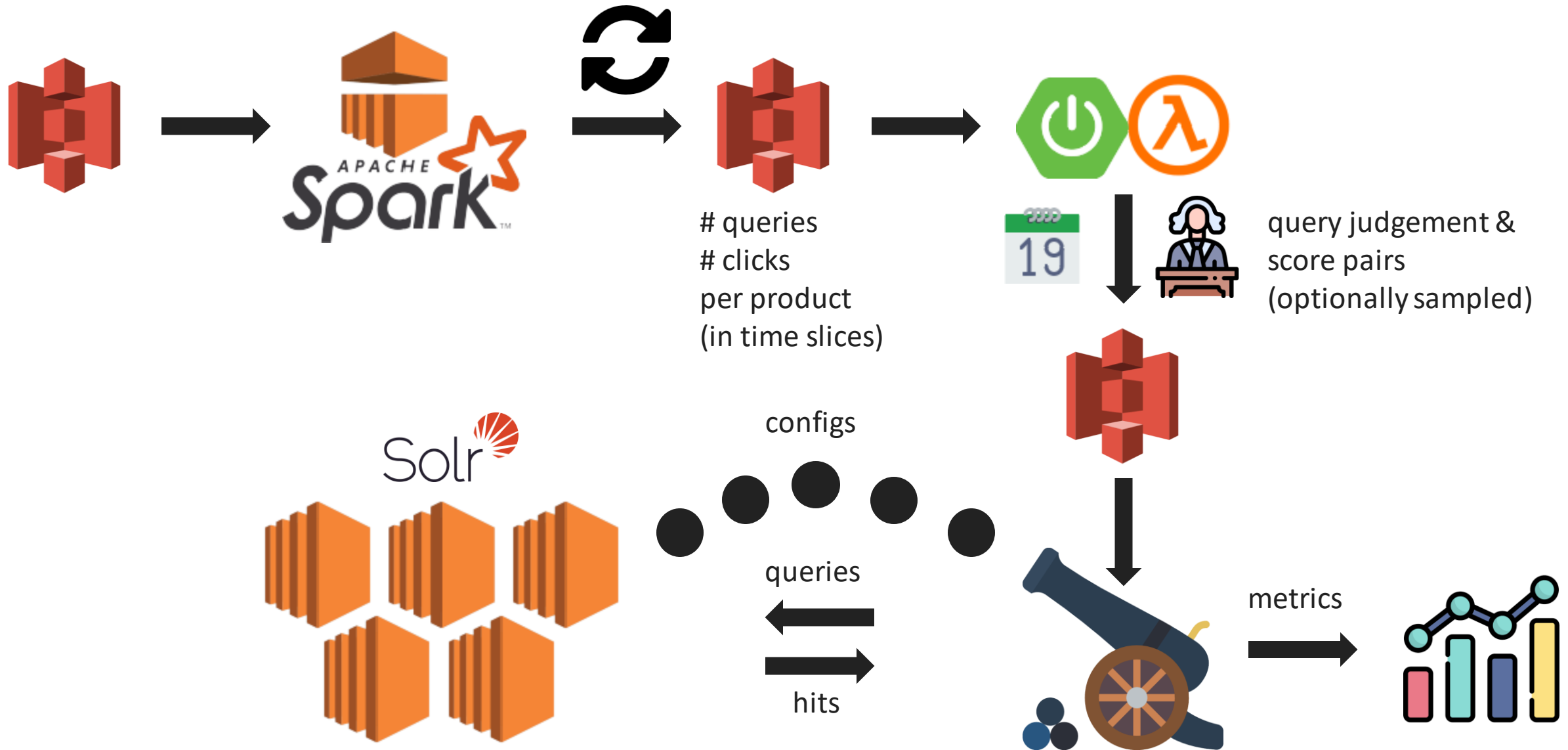


product clusters



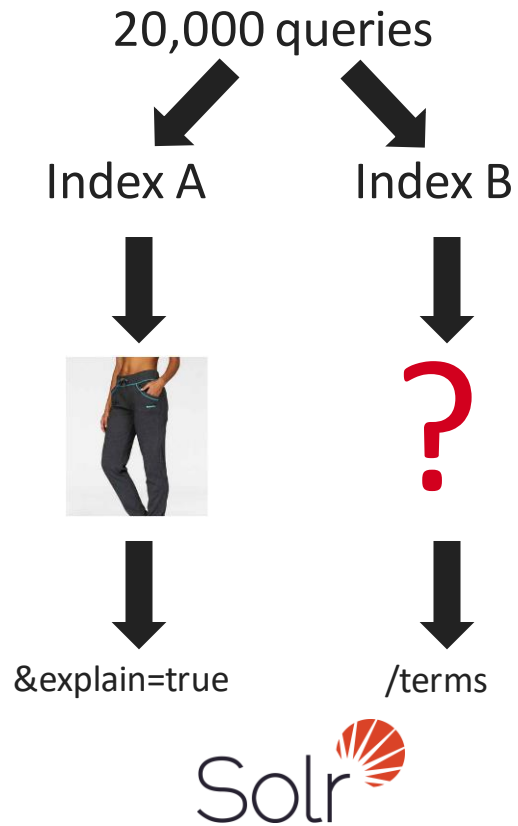
some mo' tail

Offline Evaluation Architecture



USE CASE 1: PRODUCT DATA CHANGES

Structural Changes in Indexed Data

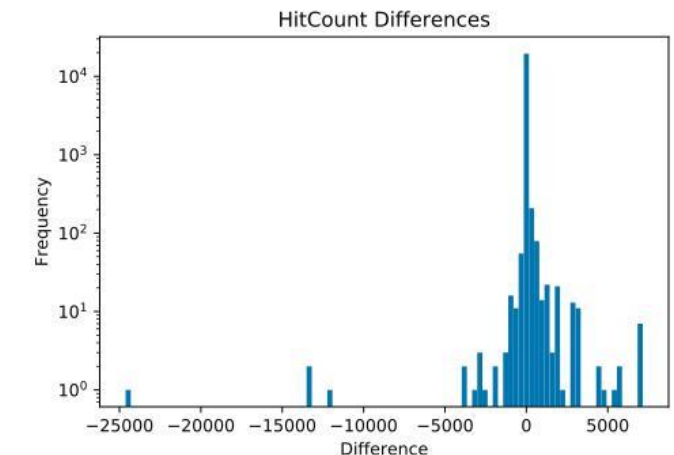


Subsamples

Different ranking	5191	25.95 %
Selection (more)	2339	11.69 %
Selection (smaller)	654	3.27 %
Equal results	11816	59.08 %

HitCount differences

samples	20000
mean	27.44
std	325.40
max	24625



Quality Metrics: Overview & Examples

Relative Metric Changes

Metric	Relative Diff
Precision@4	- 0.2 %
NDCG@10	-0.092 %
Precision@All	- 1.18 %
Recall	- 0.31 %
Avg. Precision	- 0.86 %

*“Schlafzimmmer komplett mit boxspringbett”
(missing: “komplettschlafzimm”):
→ 1 → 0 hits*

*“popsocket” (missing: “popsocket”):
→ 88 → 39 hits,
→ P@4 +33%, P@All +56 %, NDCG@10: +11 %, Avg.Precision: +17 %*

*“oberteil damen” (missing: “satinblus”):
→ 27558 → 27323 hits
→ P@4: +/- 0%, P@All: +1.9 %, NDCG: +/- 0 %, Recall: +1.9 %, Avg.Precision: - 2.9 %*

Quality Metrics: Subsamples

More/less hits refers to the new configuration

Relative Metric Changes
(more hits; n=2339)

Metric	Relative Diff
Precision@4	- 3.37 %
NDCG@10	- 2.18 %
Precision@All	- 13.66 %
Recall	- 0.75 %
Avg. Precision	- 6.34 %

Relative Metric Changes
(less hits; n=654)

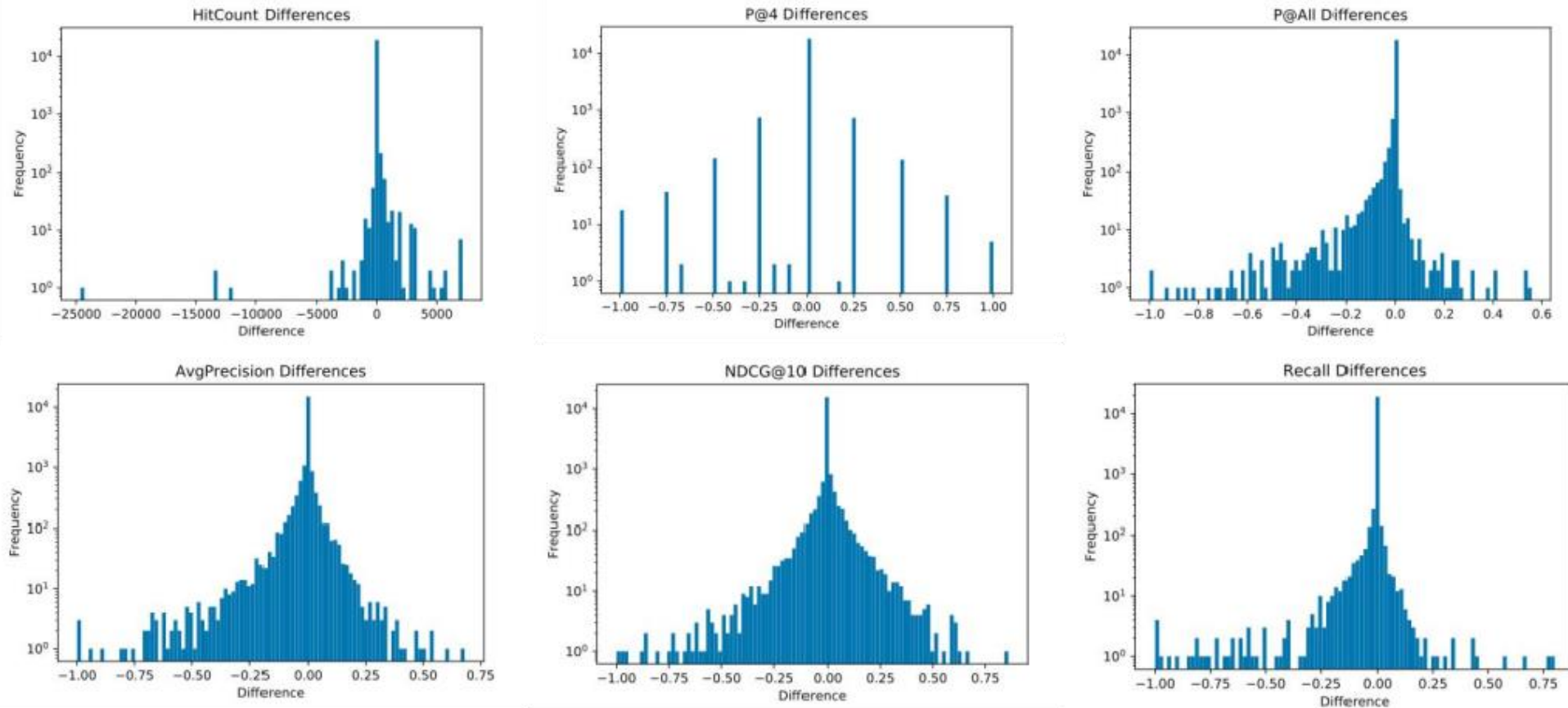
Metric	Relative Diff
Precision@4	- 6.1 %
NDCG@10	- 0.24 %
Precision@All	+ 5.01 %
Recall	- 4.18 %
Avg. Precision	- 2.43 %

Relative Metric Changes
(same hits; n=5191)

Metric	Relative Diff
Precision@4	- 4.00 %
NDCG@10	- 4.11 %
Precision@All	- 12.21 %
Recall	- 1.44 %
Avg. Precision	- 6.78 %

Result Visualization & Interpretation

Metric deltas: negative values → new configuration is better



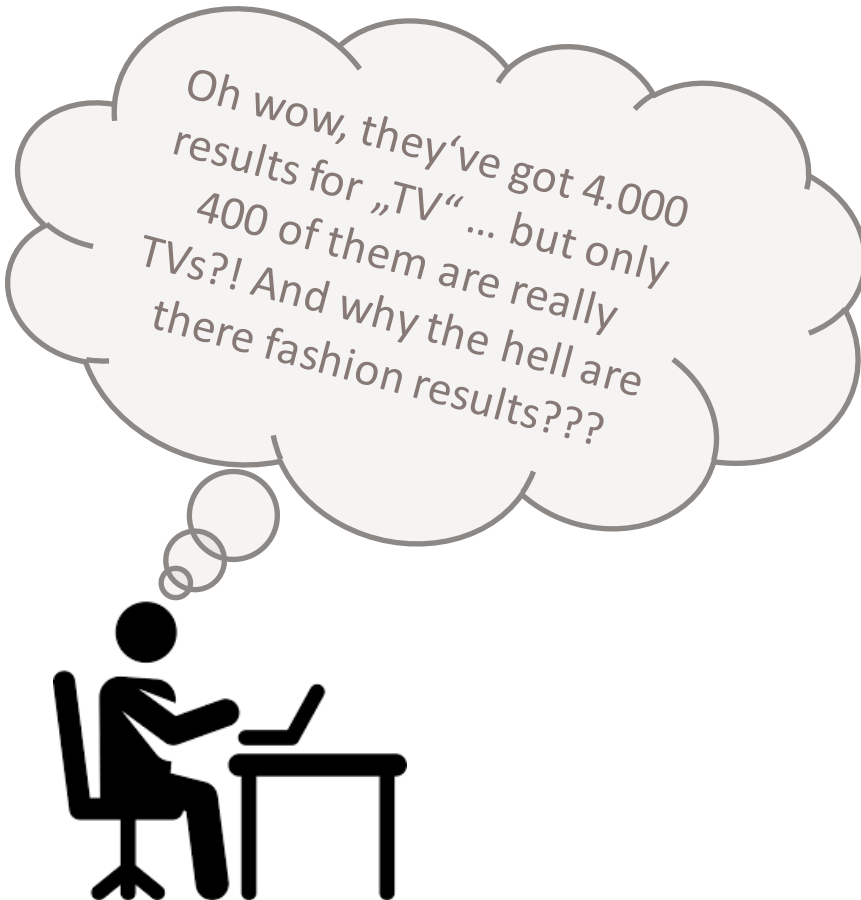
It looks like a draw

→ confirmed by on-site A/B test

USE CASE 2: THE PURSUIT OF PRECISION

Perpetual Challenge: Precision vs. Recall

OTTO



OTTO tv Service Mein Konto Merkzettel Warenkorb

Inspiration · Damen · Herren · Kinder · Wäsche/Bademode · Sport · Schuhe · Große Größen
Multimedia · Haushalt · Küche · Heimtextilien · Möbel · Baumarkt · Spielzeug · Marken · %Sale%

Startseite | **Suchergebnis für tv (4.011)** Seite 1 von 134

Sortieren nach **Topseller**

Sortiment

- ☐ Medien (1.714)
- ☐ Möbel (1.257)
- ☐ Technik (1.056)
- ☐ Ausrüstung (32)
- ☐ **Bekleidung (31)**
- ☐ Heimtextilien (10)
- ☐ Dekoration (14)

Kategorie

- ☐ Filme (1.692)
- ☐ Kommoden (975)
- ☐ Halterungen (179)
- ☐ **Fernseher (414)**
- ☐ Zubehör für Möbel (181)

THOMSON

Thomson 55UD6696 Curved-LED-Fernseher (139 cm/55 Zoll, 4K Ultra HD, Smart-TV)

UVP € 749,00
€ 479,99

A+

★★★★☆ (227)

- 139 cm (55"), 4K Ultra HD, Curved-LED
- Smart-TV, HbbTV, Netflix, 10-Bit-Unterstützung
- Integrierter Triple-Tuner (DVB-T2 HD/C/S2), PPI 1500 (HDR)
- WLAN, 3x HDMI, 2x USB, CI+ Modul-Schacht

[zum Artikel](#)

SAMSUNG

Topical Relevance vs. Business Value



Suchseite | Suchergebnisse für tv (72 von 3270)

96.88 Filter_assortmentsearch

81.56 Filter_basecolor

85.93 Filter_brand

86.06 Filter_brandcode

99.66 Filter_category

30.70 Filter_customerreview

12.45 Filter_energyefficiencycategory

13.49 Filter_functionandequipment

10.31 Filter_kind

SAMSUNG Samsung UE40NU7199UXZG LED-Fernseher (100 cm/40 Zoll, 4K Ultra HD, Smart-TV)	SAMSUNG Samsung UE65NU8509T Curved-LED-Fernseher (163 cm/65 Zoll, 4K Ultra HD, Smart-TV)	SAMSUNG Samsung UE65NU7179U LED-Fernseher (163 cm/65 Zoll, 4K Ultra HD, Smart-TV)	LG LG 70UK6950PLA LED-Fernseher (178 cm/70 Zoll, 4K Ultra HD, Smart-TV)	SAMSUNG Samsung UE49RU7379 Curved-LED-Fernseher (123 cm/49 Zoll, 4K Ultra HD)
369.99 €	1069.99 €	778.00 €	969.99 €	569.99 €
17.206223	17.067753	17.013895	16.849737	16.827995
SAMSUNG Samsung UE55NU8509T Curved-LED-Fernseher (138 cm/55 Zoll, 4K Ultra HD, Smart-TV)	TELEFUNKEN Telefunken OS-32H300 LED-Fernseher (81 cm/32 Zoll, HD-ready, Smart-TV)	TELEFUNKEN Telefunken D55U700M4CWH LED-Fernseher (140 cm/55 Zoll, 4K Ultra HD, Smart-TV)	BLAUPUNKT Blaupunkt BLA-50/405V-GB-11B4-UEGBQPX-EU LED-Fernseher (127 cm/50 Zoll, 4K Ultra HD, Smart-TV)	LG LG 43UK6300LLB LED-Fernseher (108 cm/43 Zoll, 4K Ultra HD, Smart-TV)
899.99 €	169.99 €	419.99 €	346.99 €	339.90 €
16.684359	16.51862	16.513584	16.491585	16.46508



Business Value vs. Relevance – Query „TV“



Topical Relevance vs. Business Value

OTTO



Suchseite | Suchergebnisse für gürtel (72 von 4234)

99.88 Filter_assortmentsearch

97.76 Filter_basecolor

48.61 Filter_beltsize

95.98 Filter_brand

96.36 Filter_brandcode

100.00 Filter_category

16.79 Filter_customerreview

24.26 Filter_fashionsizeeu

20.69 Filter_fashionsizeus

BUFFALO Buffalo Strandhose aus Leinen-Mix 45.00 € 11.993399

J.JAYZ J.Jayz Ledergürtel mit glänzender Schließe, auch in Plus Size Größen 19.99 € 11.918791

TOM TAILOR POLO TEAM TOM TAILOR Polo Team Shorts (Set, mit Gürtel) 34.99 € 11.862791

JOHN DEVIN John Devin Cargo Bermudas (Set, mit Gürtel) 33.99 € 11.571348

JOHN DEVIN John Devin Cargo Bermudas (Set, mit Gürtel) 29.99 € 11.382267

BENCH. Bench. Bermudas (Set, 2-tlg., mit Gürtel) mit vielen Taschen und Gürtel 29.99 € 10.967842

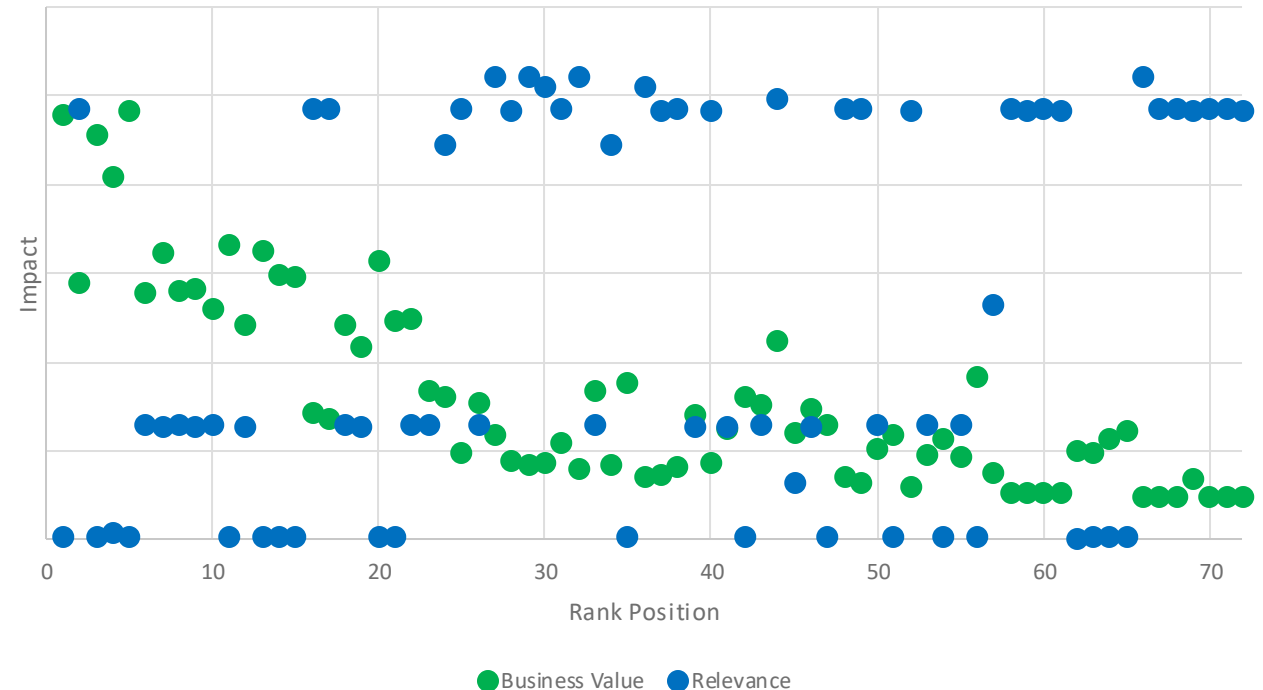
PETROL INDUSTRIES Petrol Industries Cargoshorts (Set, mit abnehmbarem Gürtel) 38.99 € 10.818434

RHODE ISLAND Rhode Island Cargo Bermudas (Set, 2-tlg., mit Gürtel) incl. Gürtel 28.99 € 10.781582

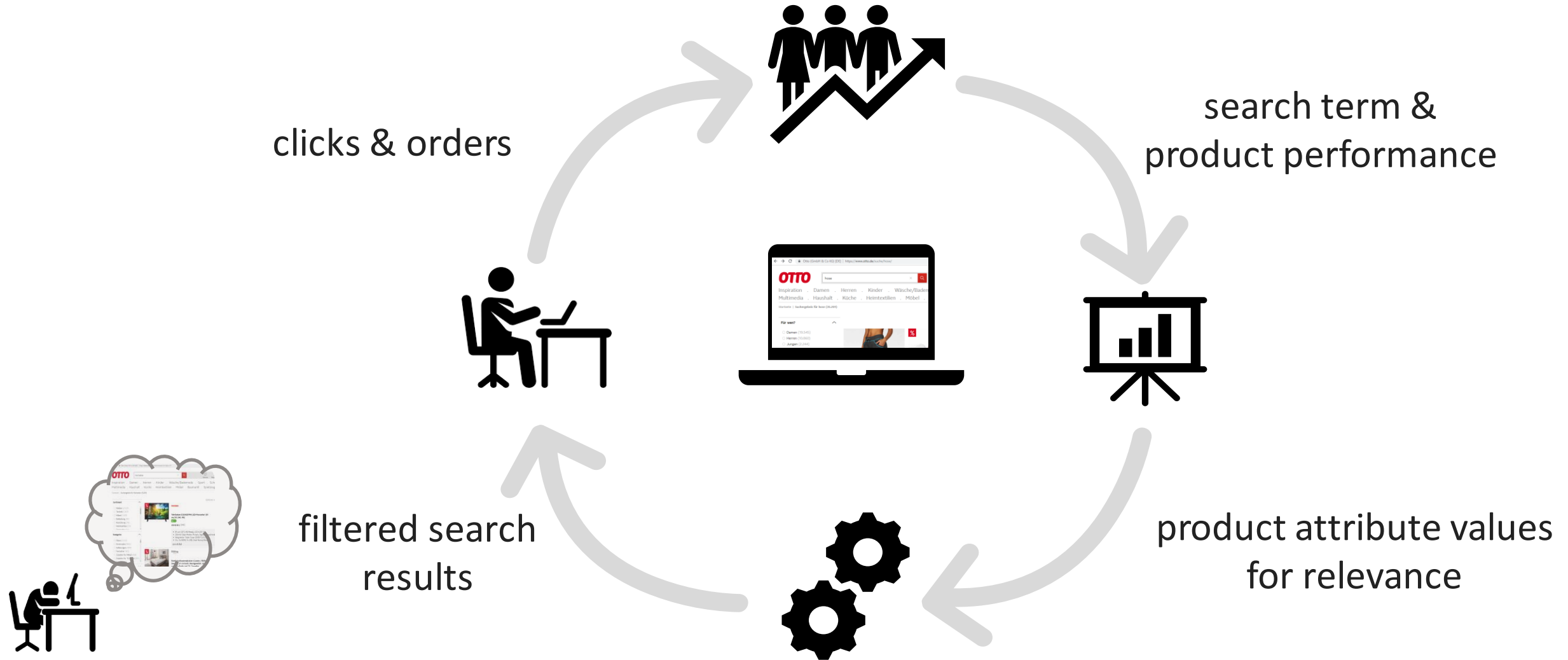
BRUNO BANANI Unisex-Bademantel »Robby«, Bruno Banani, mit »bb« Logostickerei 34.99 € 10.676108

BRUNO BANANI Bruno Banani ChinosHORTS inklusive Gürtel 39.99 € 10.592767

Business Value vs. Relevance - Query "belt"



Interaction-based Precision Improvement



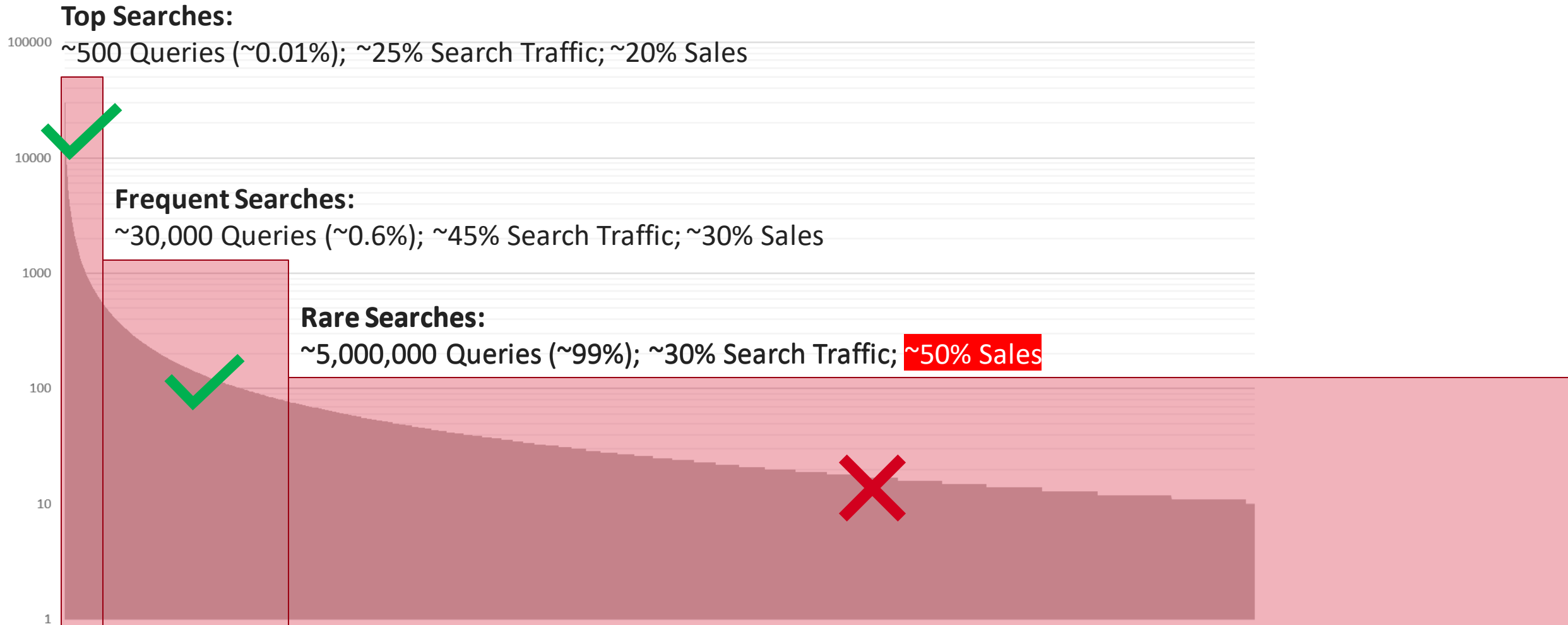
Precision Improvement PoC: Offline Results



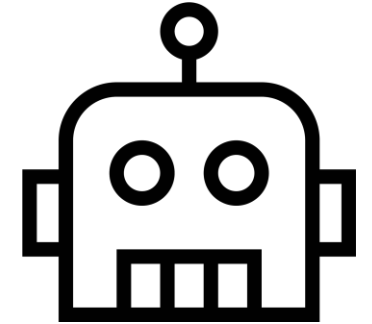
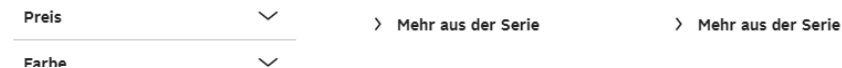
Configuration for Offline Evaluation	Uplift P@4	Uplift P@30	Uplift P@100	Uplift AP@30	Uplift AP@100	% of changed traffic	% of changed queries	Avg. difference of hits
product_ci_producttype-clicks-cov90	4,22%	5,76%	13,33%	5,76%	10,14%	86,61%	63,37%	31,19%
product_ci_producttype-clicks-cov95	4,19%	5,58%	12,87%	5,64%	9,89%	86,65%	63,38%	30,56%
product_ci_category-clicks-cov90	2,66%	4,21%	8,76%	3,79%	6,48%	88,35%	84,12%	14,35%
product_ci_category-clicks-cov95	2,62%	4,12%	8,50%	3,72%	6,31%	88,37%	84,14%	13,82%
product_ci_producttype-a2b-cov90	1,84%	2,82%	7,40%	2,81%	5,45%	74,54%	19,69%	32,56%
product_ci_producttype-a2b-cov95	1,84%	2,78%	7,33%	2,78%	5,41%	74,86%	19,69%	32,11%
product_ci_category-a2b-cov90	1,13%	1,84%	4,74%	1,69%	3,34%	79,86%	29,84%	14,96%
product_ci_category-a2b-cov95	1,16%	1,84%	4,72%	1,69%	3,33%	79,86%	29,84%	14,70%
product_ci_assortmentsearch-clicks-cov90	0,86%	1,67%	4,20%	1,53%	3,12%	88,97%	87,63%	8,02%
product_ci_assortmentsearch-clicks-cov95	0,85%	1,63%	4,06%	1,50%	3,02%	88,98%	87,63%	7,76%
product_ci_assortmentsearch-a2b-cov90	0,47%	0,90%	2,44%	0,82%	1,78%	80,71%	31,63%	7,54%
product_ci_assortmentsearch-a2b-cov95	0,48%	0,89%	2,43%	0,81%	1,77%	80,71%	31,63%	7,43%

OUR LEARNINGS FROM CONTINUOUS OFFLINE EVALUATION

Limitations of Click-based Judgements



OTTO



OTTO

SUMMARY

Connect with us.



jens.kuersten@otto.de



@faultfinder80



andreas.wagenmann@otto.de



@andiwagen

We are hiring.