Who am I?

Foundations and Trends® in Information
Retrieval
Vol. 11, No. 1 (2017) 1–142c●2017
U. Kruschwitz and C. Hull
DOI: 10.1561/1500000053

- Started at Muscat with Dr. Martin Porter
- Co-founder of Flax, the UK's leading open
  source search consultancy (2001-2019)
- Now a Managing Consultant at OpenSource
  Connections
- Open source advocate, event organiser, author
  & speaker

http://thesearchnetwork.com/

OpenSource Connections

- "Empowering The World's Search Teams"
- Elasticsearch & Solr experts, committers
- Consultancy & training - "Think Like a Relevance Engineer", LTR, NLP
- US & UK offices
- Clients include USPTO, NHS Wales, LexisNexis, Wikimedia Foundation, Snagajob, Under Armour
- Community initiatives include Meetups & conferences, Relevance Slack - join at www.opensourceconnections.com/slack

With examples using Elasticsearch and Solr

Relevant SEARCH

Doug Turnbull
John Berryman

MANNING
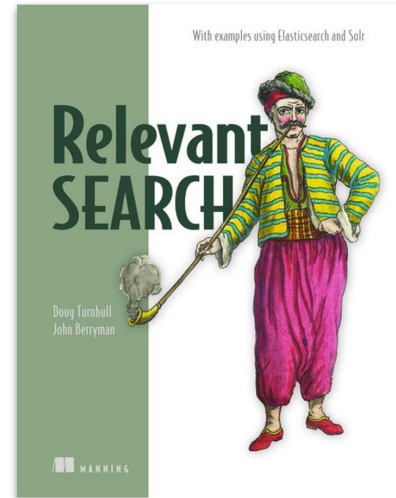
HAYSTACK

www.haystackconf.com
The Search Relevance Conference
(Haystack LIVE! Meetups every few weeks)

OpenSource Connections

Ecommerce search is broken

Let's prove it!

- Pick a site (Sainsbury, Gap…)
- Test queries
  - "white tshirt"
  - "Cheap" vs "Expensive"
  - Speeling mistooks
  - "Shirt without stripes"

- Really, in 2020?
- Again, really, when ecommerce is a vital service now?

Introduction

Unless you're running an online store with only a few products, chances are you have a site search engine of some kind. Just like Google or Bing, this will let your users type in a word or phrase and click a button (usually something that looks like a magnifying glass). The result should be a list of items, hopefully in the right order, relevant to that word or phrase.

Unfortunately it's not quite that simple - and worse, maybe your site search is broken altogether, which can cost you users, traffic and sales. Here's a list of things to check so you can make sure you're giving the best site search experience to your users. Score the results from 0 (worst) to 3 (best) and be honest! Fire up your site, click on that search box and let's get started.

Twelve point checklist

Task	Notes	Score 0-3
Find the search box	Is it easy to find that site search box? Is it obvious that it's a search feature? Does it have space for a few words? If your users can't identify it as a search engine they may not use it at all.	
Basic search	Type in something you know is on your website - 'orange' if you sell food, 'lawnmower' if you sell gardening supplies. Do the results look OK? Let's not analyze things too much yet, but this is a first pass to check that your search engine works at all.	
Search speed	How quickly do results come back? If it's anything more than a second then it might not be acceptable to users - it should ideally be no more than 250 milliseconds. Even Google with its billions of pages returns results pretty quickly.	
Autosuggest	If you type a few letters, does the site suggest some possible queries (autocomplete)? Does it also suggest some possible results (autosuggest)? These features are common today and users will expect them - they also give users a quick way to jump to a result they want.	
Spelling	Let's try something we know is a deliberate misspelling of something that is definitely on your website. If you sell shoes, type in 'bluu shoes', if your website offers books, enter 'teen fixtion' - you get the idea. Make sure the correctly spelled version gives some results, obviously! What happens? Do you get offered a correct spelling instead, a 'Did you mean….' prompt? If you just give up and say that there's no results for that search query, you're not doing so well.	

Test your site with A Guide to Better Ecommerce Site Search

Why is ecommerce search broken?

- Ecommerce platforms bundle & bungle search
  - Bad integration
  - Lack of control
  - Little or no measurement capability
  - Platforms care about inventory management and other non user oriented aspects.

Why is ecommerce search broken?

- Ecommerce platforms bundle & bungle search
  - Bad integration
  - Lack of control
  - Little or no measurement capability
  - Platforms care about inventory management and other non user oriented aspects.
- Disconnect between marketers & IT
  - No-one told marketers how search works, or gave them the right tools
  - IT aren't close to 'the business' but are often blamed
  - Prioritisation of issues is hard & HiPPOs roam

Why is ecommerce search broken?

- Ecommerce platforms bundle & bungle search
  - Bad integration
  - Lack of control
  - Little or no measurement capability
  - Platforms care about inventory management and other non user oriented aspects.
- Disconnect between marketers & IT
  - No-one told marketers how search works, or gave them the right tools
  - IT aren't close to 'the business' but are often blamed
  - Prioritisation of issues is hard & HiPPOs roam
- Little understanding of customers
  - Intent recognition is hard
  - Search analytics still patchy
  - Data scientists don't 'get' search

Why is this important?

- Better search means better conversion rates
  - Your customers aren't using the same language as you are
  - Zero results pages burn money
  - Great search helps retention

Why is this important?

- Better search means better conversion rates
  - Your customers aren't using the same language as you are
  - Zero results pages burn money
  - Great search helps retention
- Search knowledge informs strategy
  - What's in your query logs?
  - Understanding intent

Why is this important?

- Better search means better conversion rates
  - Your customers aren't using the same language as you are
  - Zero results pages burn money
  - Great search helps retention
- Search knowledge informs strategy
  - What's in your query logs?
  - Understanding intent
- Search is a gateway to personalisation
  - A 'virtual aisle' for each customer
  - ML, LTR, recommendations, chat bots...

"UK retail ecommerce sales will account for 27.5% of total retail sales this year, and that proportion will approach one-third by 2024. The post-pandemic hit to physical retail will be felt far into the future."
Emarketer

"Online sales are now expected to grow from initial predictions of 11% to 19% in 2020, reaching £78.9bn, this is up from £66.3bn in 2019....Amazon is expected to benefit the most from this surge in demand for online retail"
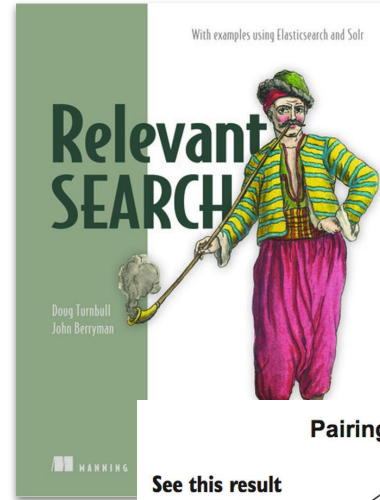*Edge by Ascential*

Measure, test, repeat

OpenSource Connections

5 Steps to Better Relevance

- **M**easurement
- **E**xperimentation
- **R**apidity
- **C**ollaboration
- **H**elp
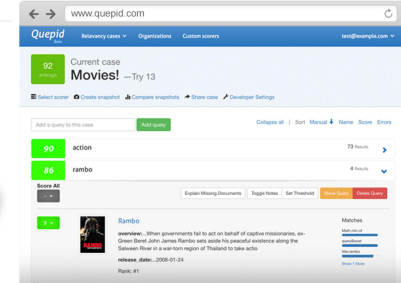
Relevancy Framework
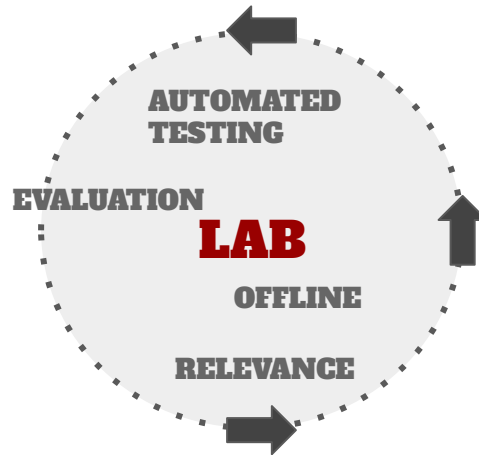
Search Quality: A Business-Friendly Perspective
Peter Fries - Haystack 2018

Relevancy Framework
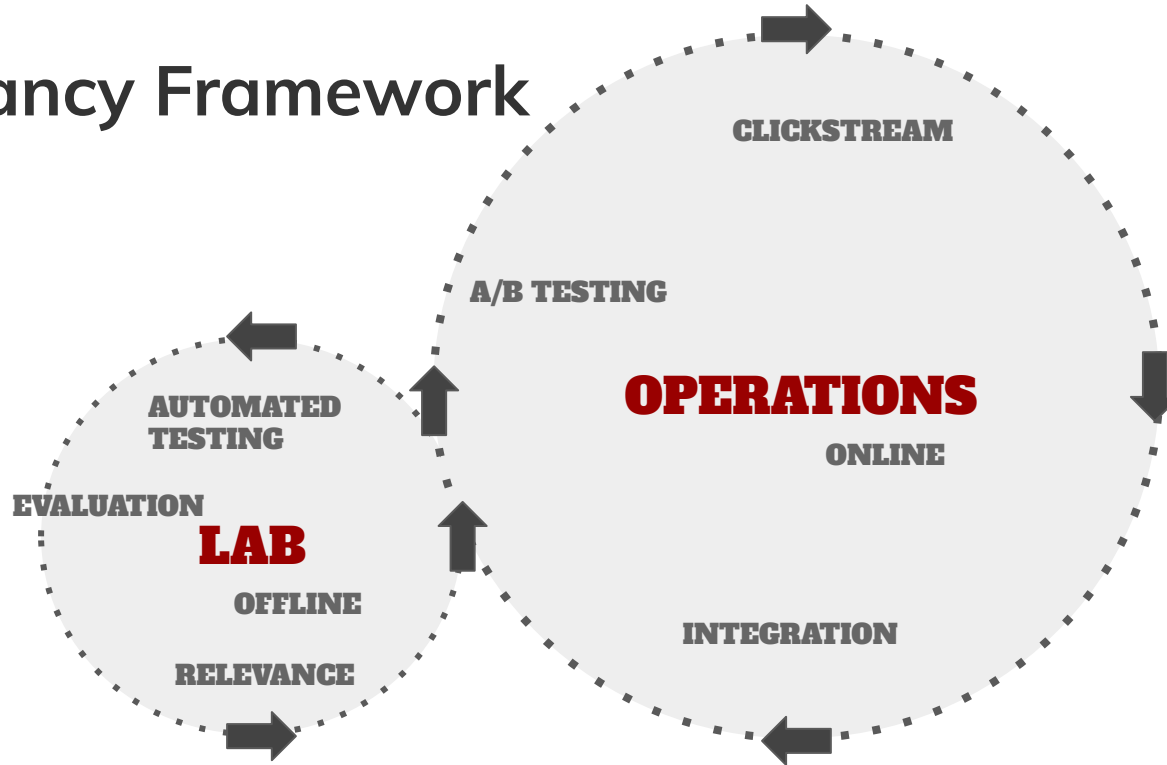
OpenSource Connections

CLICKSTREAM

A/B TESTING

OPERATIONS

ONLINE

INTEGRATION

AUTOMATED TESTING

EVALUATION

LAB

OFFLINE

RELEVANCE

Search Quality: A Business-Friendly Perspective
Peter Fries - Haystack 2018

www.opensourceconnections.com

Tools & Processes

EXP1: Learnings

Baseline to Experiment 1 Filters

Grouped ☐ Stacked

settings ▼

Generator Chainsaw Electric Fence Fence panel Gun safes

Baseline to Experiment 1 Product Categories

☐ Grouped ☐ Stacked

settings ▼

Generator Chainsaw Electric Fence Fence panel Gun safes

90 Day Relevance Canvas

Chorus Electronics

Feb 11
Iteration 1

Problem
- Lack of Trust in Results
- Lots of noise in the system. Too many false positive matches.
- Facet set is too diverse.
- Lacking data curation
- No holistic view of search journey (across recommendations, data quality, content, conversions)

Solution
- Fix the Facets.
- Navigational Search
- Auto Category Drilldown
- Offline Testing
- Active Search Management
- LTR Platform Live
- Attribute Search

QA (Search Nerds)
- LGTM for Facets (Categories)
- NDCG@4, 8, 12
- DCG@4,8,12 for LTR

Core 90 Day Focus

Increased Precision, reduced Recall and Noise.

Data Assets
- Real time Store Info to the Web
- Products Data
- User Behavior

User Population
- Mobile
- Desktop
- (Organic Only)
- (Internal Only)

Search Types
- Exploratory
- Attribute Search
-

...h Goes ↑
...research through search
...h then Click!) & Direct Conversion
...k re: "Findability"

External

OneRelevance Relevancy Planning Workbook

ID	Name	Problem Cluster ID	Query Population Type	Status	Experiment Story No.	Hypothesis We believe that <this capability>	Will result in <this outcome>	We will know we have succeeded when <we see a measurable signal>	Target Platform	Level of Effo	Time Requi red
EXP001			Precedent search	Refined	S-83404	optimising query parameters for Precedents	user's desired precedent in position 1	we have great p@k at position 1	Solr	H	
EXP002			Form search	Refined	S-83405	optimising query parameters for Forms	user's desired Form in position 1	we have great p@k at position 1	Solr	H	
EXP003				To be refined	S-46742	unrated document gets a rating of 1 (the lowest rating) and hence 0 dcg	having some degree of overlap btw solr and ML	we are able to see if there is overlap or not with the scoring change	Solr		
EXP004				To be refined	S-83406	unrated documents get a rating of 4 (the highest rating)	having some degree of overlap btw solr and ML	we are able to see if there is overlap or not with the scoring change	Solr		
EXP020		PC6	Boost by Pop	REMOVE		increasing the popularity boost in the query template;	more relevant documents moving up in the SERP; and less relevant documents not being over-promoted	we see [6%] of the sample testing feedback moving from 'worse' to 'same' or 'better' relevance.	MarkLogic		
EXP021				Refined	S-82760	increasing the threshold for popular query candidates; and	more relevant documents moving up in the SERP; and less relevant documents not being over-promoted	we see [6%] of the sample testing feedback moving from 'worse' to 'same' or 'better' relevance.	MarkLogic		
EXP025		PC6	Boost by Pop	Refined	S-82760	Updating the snapshot of data behind Boost by popularity (with the Forms included)	more relevant documents moving up in the SERP	we see an increase in hDCG score compared to the baseline	MarkLogic		
	Synonyms			Completed		Enabling synonyms	increase relevancy but may impact precision	we see an improved hDCG number	Solr		3 day
	Stemming			Completed		different stemming configurations	cause a shift in precision and recall	we see an improved hDCG or verify that the current configuration is suitable for the content.	Solr		3 day
	Minmatch			Completed		different mm thresholds	increase precision but may alter	we see an improved hDCG with	Solr		2 day

Why a tool stack for *ecommerce* search?

- Strong demand for onsite $€ARCH optimisation
  - Need tools to measure quality
  - Sophisticated search management requirements - "searchandizing"
  - Reduce ramp-up time and start optimising for your business earlier

- Reduce the gap: open source search engines are not built for ecommerce
  - Ranking models made for text documents vs highly structured data (fields!) in e-commerce
  - How do we deal with variants of a product? (matching, ranking, facets)
  - Boosting, redirects, synonyms often hard to manage with text files & code

The team

Combined 45 years of experience in search
Open Source enthusiasts
ASF member, Committers on: Solr, Querqy, SMUI, Quepid, Contributions: RRE, NiFi

@pbartusch @renekrie
@dep4b

>BLN
BZZ/
WRDS

Paul Maria Bartusch,

René Kriegler,

Johannes Peter &

Eric Pugh
Towards an open source tool stack for e-commerce search

Wednesday 10th June
19:30-21:30 CEST
MICESlive

A joint virtual conference with
HAYSTACK + **MICES**

OpenSource Connections

Solr

Blacklight

Querqy
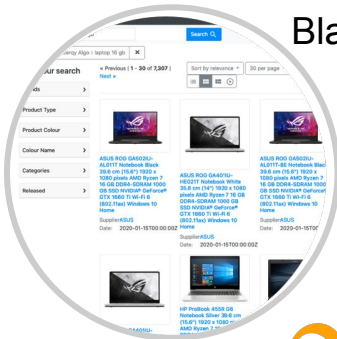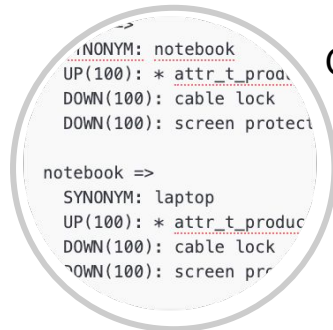
SMUI
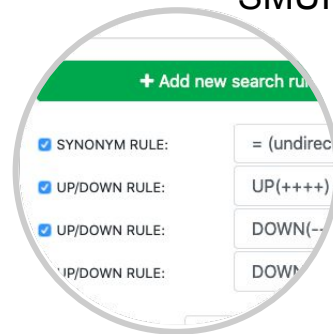
Docker

Chorus

RRE
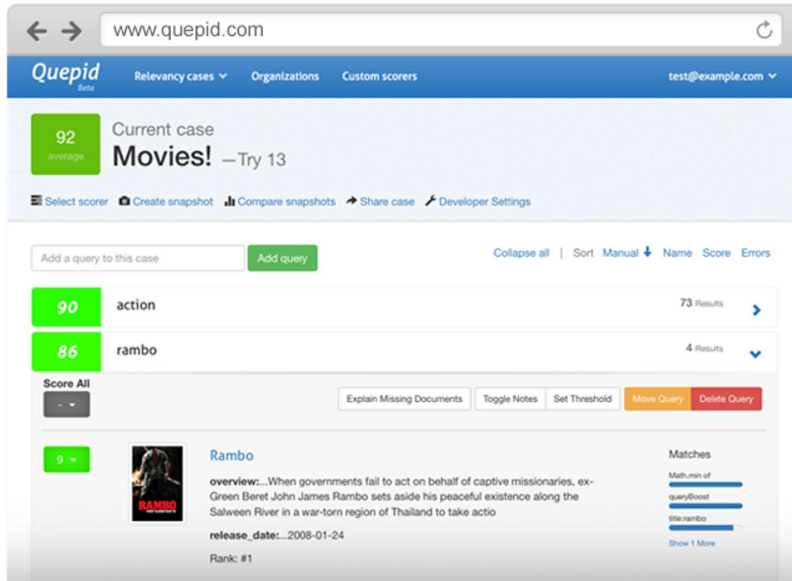
Quepid

An open source stack for e-commerce search

- Reduce time to get on par with commercial search engines
    - package as integrated toolset for Elasticsearch and Solr (Chorus bootstrap application)

- Supply solutions for typical e-commerce search:
    - Tools to support merchandiser/search manager (SMUI/Querqy)
    - Easy and extendible query building and rewriting (Querqy)
    - Manual judgment collection (Quepid)
    - Automated search relevance testing (RRE)
    - Parameter/configuration optimisation (??, maybe RRE or Quaerite)
    - Simple (Demo only) search UI (Blacklight)
    - Basic schema setup for typical ecommerce (inventory, pricing, store locations) use cases

Chorus

Quepid

- Search Relevance dashboard & testing
- Used by clients such as Advance Auto, O'Reilly Media, Historic England, Salesforce and others
- 1500+ registered users
- Free hosted version at www.quepid.com
- Open source at github.com/o19s/quepid
- iSpy - Prototype connectors to *any website frontend* (in development)

Querqy

Library and framework for
- query rewriting
- optimised query building with many parameters to tune search relevance

Plugins available for Solr (2014) and Elasticsearch (2019)
Apache 2 License
github.com/querqy & querqy.org

Querqy - Common Rules Rewriter

- synonyms

```
personal computer =>
    SYNONYM: pc


wall* =>
    SYNONYM: wall $1

(eg. wallmount => wall mount)
```

- boost/penalty

```
iphone =>
    UP(200): apple
    DOWN(10): * category:accessories
```

- filter

```
iphone =>
    FILTER: apple
    FILTER: -case
```

- delete

```
cheap iphone =>
    DELETE: cheap
```

Querqy - Common Rules Rewriter - Advanced

- RawQuery: Opening the door to combined Querqy-Lucene power

```
new =>
    DELETE: new
    UP(1.0): * release_date:[NOW/DAY-4DAYS TO NOW/DAY+1DAY]

special offer =>
    DELETE: special offer
    FILTER: * strike_price:[* TO *]
    UP(2.0): *
{!func}if(gte(rint(mul(div(sub(strike_price,price),strike_price),100)),20),20,0)
```

Rewriters that come with Querqy

- Common Rules Rewriter

- Replace Rewriter

- Number-Unit Rewriter

- Shingle Rewriter

- Word Break Rewriter

- Write your own - it's a framework!

```
mobile => (mobile OR smartphone)
ombile => mobile
laptop 15" => laptop AND screen_size:[13.5 TO 16.5]
i phone => (i phone OR iphone)
grainfree => (grainfree OR grain free)
voer voor honden => (voer voor honden OR hondenvoer)
```

Demo - Chorus Electronics Store

What's next?

Ugh.. Google Analytics?

Nothing good

RRE
AgentQ

Quepid
Supahands?

CLICKSTREAM

A/B TESTING

OPERATIONS

ONLINE

INTEGRATION

AUTOMATED
TESTING

EVALUATION

LAB

OFFLINE
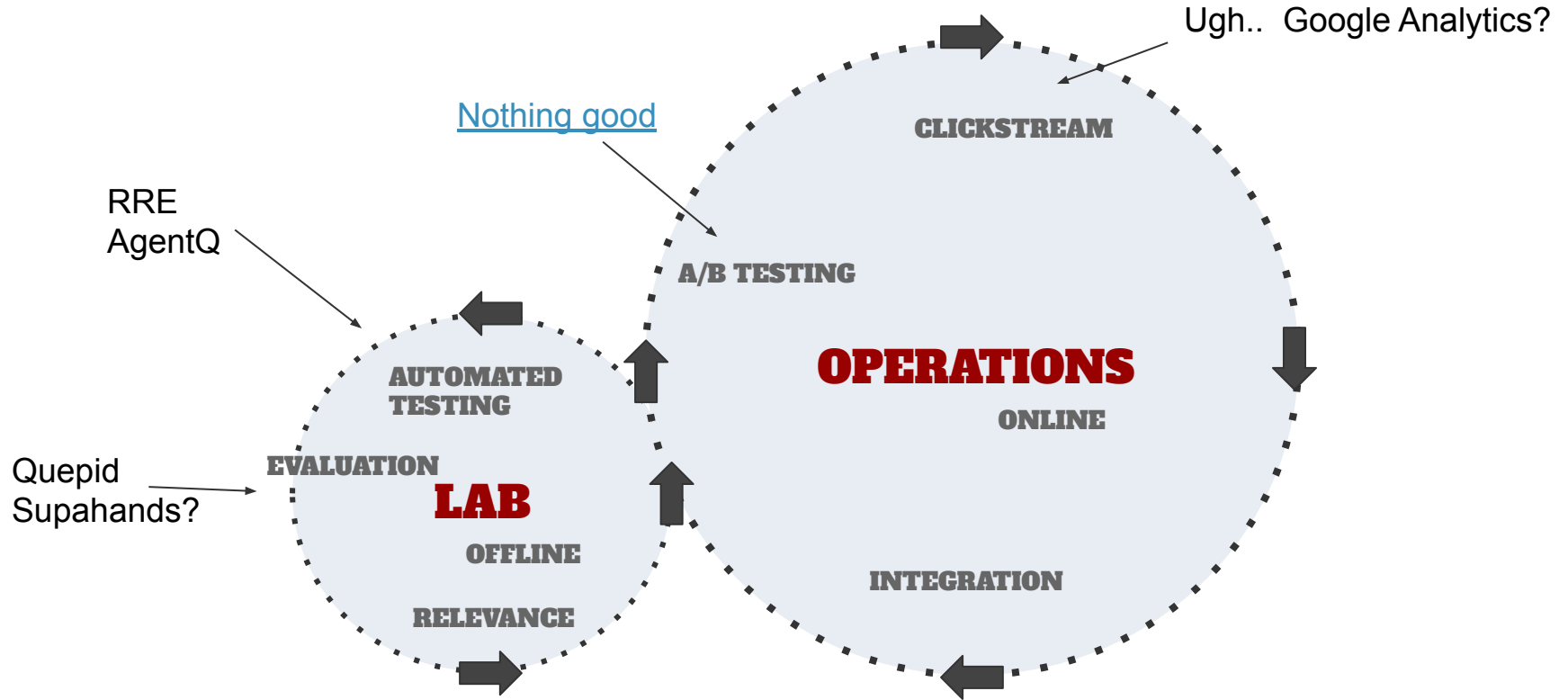
RELEVANCE

State of play

Join the Chorus!

- Meet Pete the Product Owner - blog & video series
- https://github.com/querqy/chorus for code
- www.querqy.org for documentation
- Free online workshops running at https://plainschwarz.com/ps-salon/
- Join Relevance Slack
- Let OSC help you build better ecommerce search!

Thank you.

Contact me at chull@opensourceconnections.com and follow @FlaxSearch