OpenSource Connections

edrone

Haystack On Tour - Kraków

opensourceconnections.com

edrone, February 23rd 2023

Welcome to Haystack On Tour!

edrone , February 23rd 2023

- Who are we?
  - OpenSource Connections
  - edrone
- Why is search still so hard to get right?
- What is Haystack on Tour?
- Agenda for today

Who are OpenSource Connections ?

- A team of search engine specialists with deep knowledge & decades of experience
- Our approach is data-driven, scientific and focused on business needs
- We help organizations in the USA and EU build powerful, scalable, accurate and relevant search applications by **empowering their teams to succeed** with a mix of tools, processes, training and consultancy

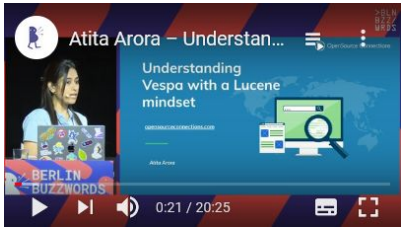
www.opensourceconnections.com

- We write books & reports

- We host leading search events…

HAYSTACK

www.haystackconf.com



…and present at many more

APACHECON NORTH AMERICA OCTOBER 3 - 6, 2022 NEW ORLEANS LOUISIANA

Enterprise Search & Discovery 2022

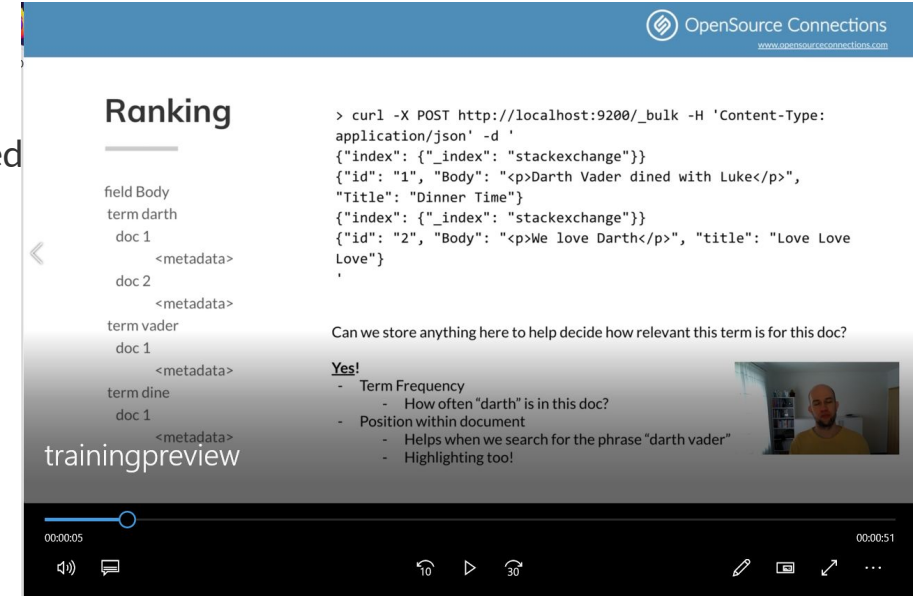
>BLN BZZ/ WRDS

- We build open source, free tools for tuning search engines

New trainings

Think Like a Relevance Engineer - now self-led on Moodle! Videos, slides, labs & quizzes - 5 week course with instructor sessions every Tuesday 930-11am US Eastern Time

- TLRE Elasticsearch - from 20th March
- TLRE Solr - from 1st May
- Hello Learning to Rank - dates to be announced
- www.opensourceconnections.com/training



Beyond Relevance - Search Result Quality

- In person at Haystack US
- Thursday 27th & morning of Friday 28th April
- Charlottesville, Virginia
- Tickets at www.haystackconf.com

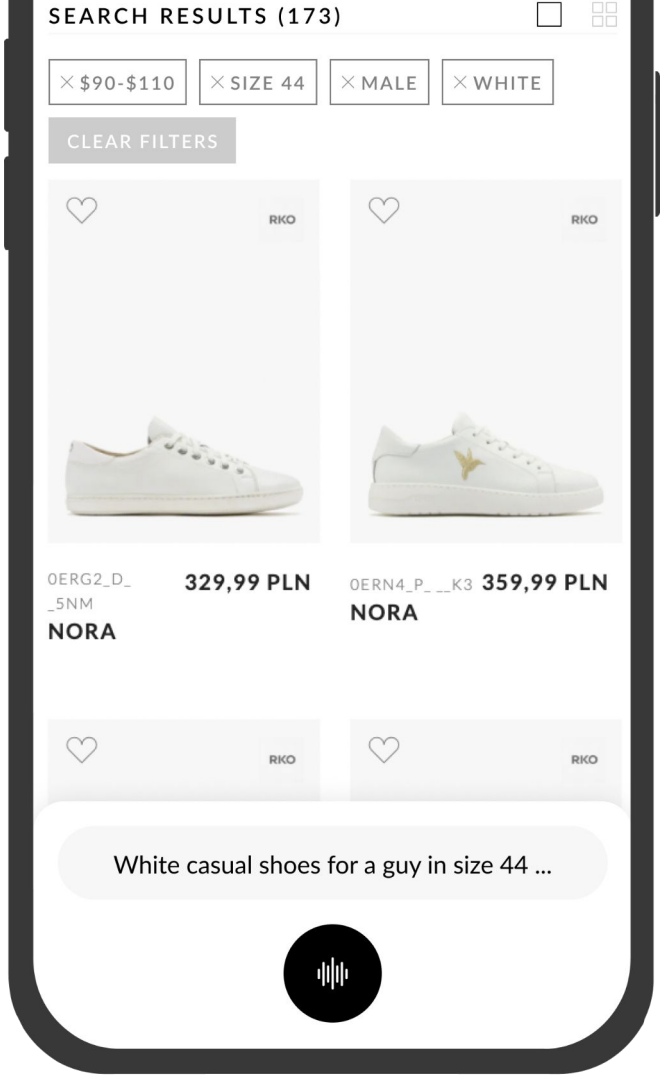
Who are edrone ?

- Advanced tools for e-commerce
- 1200+ shops
- 20+ R&D team working on Voice Search

Top 50 Commerce Products BEST SOFTWARE AWARDS 2022

BEST IN CLOUD

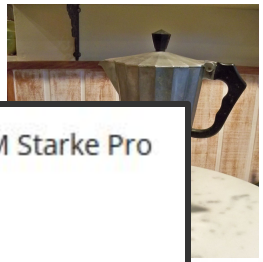
amazon web services | Well-Architected



Why is search still so hard to get right?

I don't speak your language!

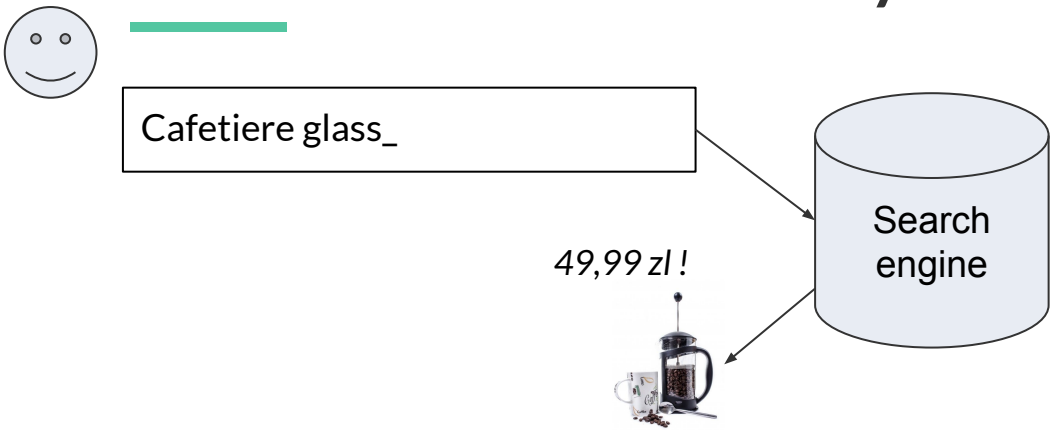
French press_



…
"dzbanek do kawy"

?? 0 results ??

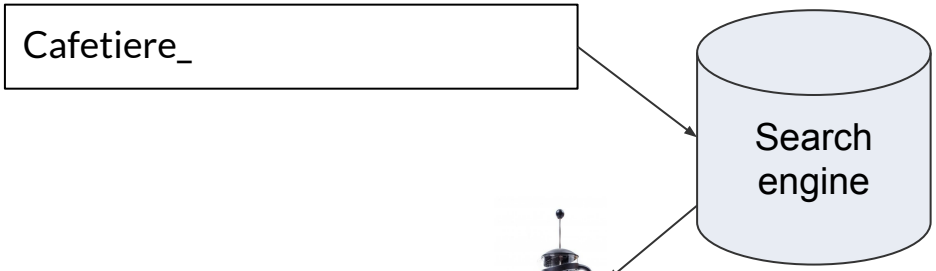
I don’t know what you want!



“Actually I wanted some new glass for my old broken cafetiere…..”

Intention is hard to guess.

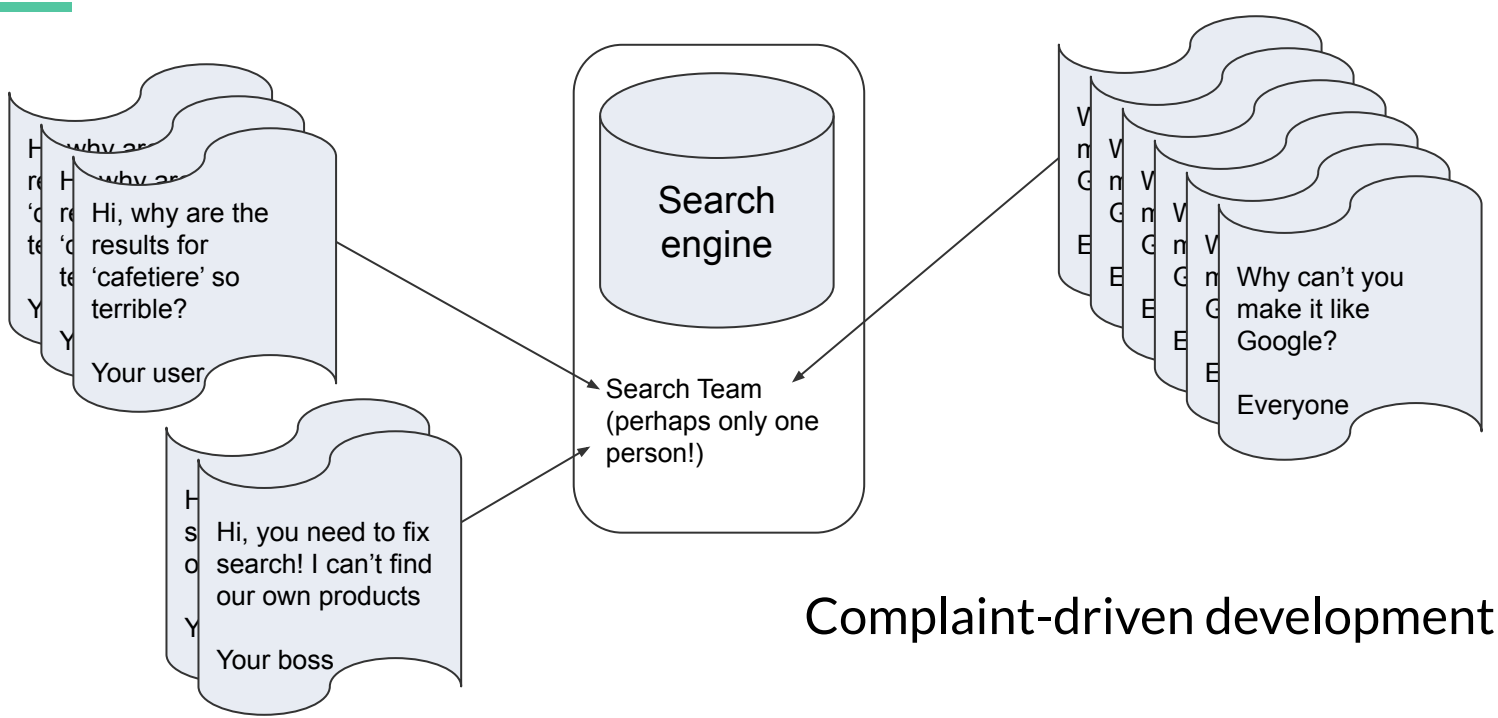
I don't know how good my search is!



Are these results relevant?
How can I measure this?
What metrics can I produce?

Relevance is subjective.

Everybody hates my search engine!



Complaint-driven development

Too many choices!

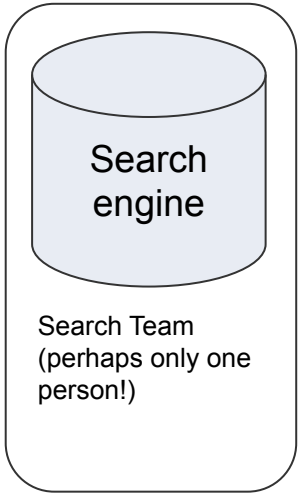
Commercial platforms

algolia Lucidworks coveo

On-prem?
SaaS?
PaaS?

Open source platforms

vespa Elasticsearch OpenSearch Solr



Vector search

Hugging Face jina Pinecone
GSI TECHNOLOGY qdrant vectara
deepset MILVUS 2.0 Weaviate Open Source

"All that glitters is not gold"

A wicked problem

*"This helps to explain why an approach which takes individual elements of a search solution and tries to solve them **will almost certainly fail.** The requirement is to work as a team across multiple elements of an Enterprise Search implementation with a leader with the experience to challenge and then work with a team to resolve an element. Even then **there is a high probability that not all the elements can be resolved**, which is why Enterprise Search applications need to be well supported with a search team post a nominal implementation."*

*a **wicked problem** is a problem that is difficult or impossible to solve because of incomplete, contradictory, and changing requirements that are often difficult to recognize. (Wikipedia)*

https://www.linkedin.com/pulse/why-enterprise-search-wicked-problem-martin-white/

A wicked problem

"This helps to explain why an approach which takes individual elements of a search solution and tries to solve them will almost certainly fail. The requirement is to work as a team across multiple elements of an Enterprise Search implementation with a leader with the experience to challenge and then work with a team to resolve an element. Even then there is a high probability that not all the elements can be resolved, which is why Enterprise Search applications need to be well supported with a search team post a nominal implementation."

*a **wicked problem** is a problem that is difficult or impossible to solve because of incomplete, contradictory, and changing requirements that are often difficult to recognize. (Wikipedia)*

https://www.linkedin.com/pulse/why-enterprise-search-wicked-problem-martin-white/

What is Haystack On Tour?

Haystack Events since 2018

- Haystack US - April, Charlottesville x 4
- Haystack EU - Sep/Oct, London x 1, Berlin x 2, ??
- Haystack LIVE! Meetup - every 4-6 weeks, Thursday 5pm CET
- First Haystack On Tour Paris, November 2023 - more planned
- Most talks available as recordings on Youtube

Haystack Events since 2018

- Haystack US - April, Charlottesville x 4
- Haystack EU - Sep/Oct, London x 1, Berlin x 2, ??
- Haystack LIVE! Meetup - every 4-6 weeks, Thursday 5pm CET
- First Haystack On Tour Paris, November 2023 - more planned
- Most talks available as recordings on Youtube
- *"Haystack is the conference for organizations where search, matching, and relevance really matters to the bottom line. For search managers, developers, relevance engineers & data scientists finding ways to innovate, see past the silver bullets, and share what actually has worked well for their unique problems. Please come share and learn!"*

Haystack this year

- Haystack US - April, Charlottesville x 4
  - Week of 24th April 2023
    - Start with a Meetup the evening of Monday 24th April
    - Main conference days Tuesday 25th & Wednesday 26th April
    - Stay in town for trainings, hacking & a free co-working space 27th April
    - Social event on Friday 28th April (we hope to organise a tour of local wineries)
  - Tickets on sale now at www.haystackconf.com

We want to....

- Help you with today's search problems

We want to....

- Help you with today's search problems
- Show you what's coming next in search

We want to....

- Help you with today's search problems
- Show you what's coming next in search
- Encourage mutual support:
  - Help you learn from each other
  - Give you a chance to make new contacts
  - Build the Polish search community

Building amazing search is hard.

Let’s try and make it easier, together!

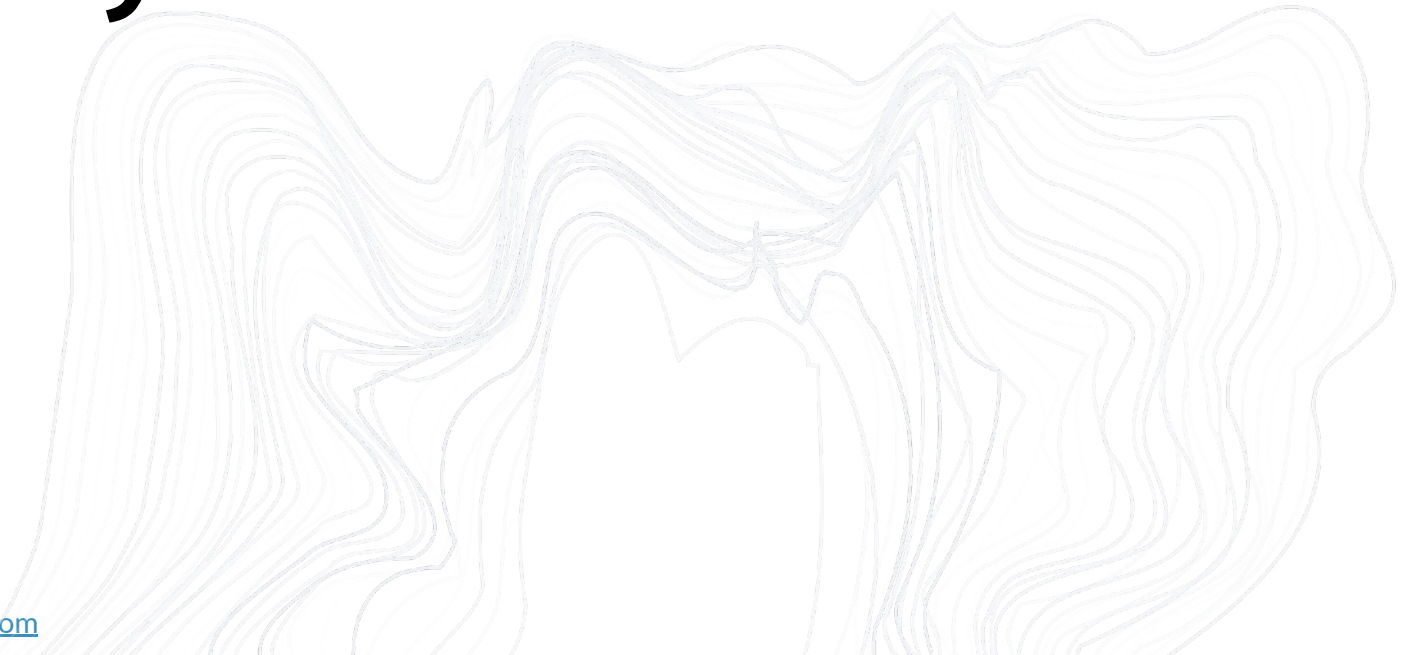
edrone

OpenSource Connections

Agenda

14:00 – 15:00	**Intro: Charlie Hull: “Why is search still so hard to get right?”** **Rene Kriegler: “Towards systematic search quality improvement”**
15:00 – 15:10	*Break / Generate Topics for Barcamp*
15:10 – 15:40	**Piotr Stachowicz: “Voice Search – challenges & opportunities”**
15:40 – 16:10	**Arek Flinik (Lekta.AI): “Searching in the real world: why people not always want what they say they want – UX lessons learned from implementing search functions in commercial voice applications”**
16:10 – 16:30	*Break for finger food / Generate Topics for Barcamp*
16:30 – 17:00	**Atita Arora: “Vectorise your E-commerce Search”**
17:00 – 17:30	**Zbyszko Papierski: “Content deduplication: vector vs keyword approaches”**
17:30 – 17:50	*Aggregate topics & vote for Barcamp topics*
17:50 – 19:30	Barcamp (Food available from 19:00)
19:30 – 22:00	Networking

Thank you.



Thanks for coming!

And thanks to our kind hosts and all our speakers!

Further Information

Events & groups

Haystack Conference (US week of April 24th, EU in September/October) www.haystackconf.com

- **Submit your talk and tickets now on sale!**

Haystack LIVE! Meetup (every 4-6 weeks online) https://www.meetup.com/Haystack-Search-Relevance-Online/

Relevance Slack www.opensourceconnections.com/slack

OSC Youtube (100+ talks) https://www.youtube.com/c/OpenSourceConnections

When is the next Polish search event?

Further reading (more on the OpenSource Connections website!)

Search Insights https://thesearchnetwork.com/publications/

Searching the Enterprise

https://opensourceconnections.com/wp-content/uploads/2020/08/ES_book_final_journal_version.pdf

Relevant Search https://opensourceconnections.com/blog/2015/04/09/osc-is-taming-search/ (with discount code!)

AI-Powered Search https://www.manning.com/books/ai-powered-search