

HAYSTACK

US 2022

Keynote - Charlie Hull, OSC

Welcome to Charlottesville!



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com

#haystackconf

We're **OpenSource** **Connections** and thank you for coming!



- 18 employees & growing
- Offices in Charlottesville, Berlin and Cambridge UK
- We empower search teams through our writing, events, training and consulting



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com
#haystackconf

Haystack Highlights

Wednesday 27th April

- Two tracks of talks from now until 4.45 pm - 15 minute breaks between
- Lunch from 11.45am-1pm - try the Downtown Mall, find some friends?
- Lightning Talks here at 3.15pm - suggest yours in Relevance Slack - use the **#haystack-conference** channel for this (and anything else!)
- Welcome to our Online participants!
- Check out our Hiring & Looking board in the foyer
- From 5.30 - Drinks reception & buffet dinner at Kardinal Hall (a short walk)
- Please wear masks in the theatres and whenever you can indoors to protect yourself & everyone else



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com
#haystackconf

Haystack Highlights

Thursday 28th April

- Two tracks of talks from 9am until 4.30 pm
- Lunch from 11.45am-1pm - try the Downtown Mall
- Search Radar here at 11am
- Women in Search panel at 1.15pm
- Please continue to wear masks in the theatres and whenever you can indoors to protect yourself & everyone else



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com
#haystackconf

Welcome to Haystack!

Thankyou to all our speakers, the anonymous team of reviewers who picked our talks this year and to Jeff Zemerick of OSC for running the process - we really couldn't do it without you!

10:00-10:45am EDT	<p>Learning a Joint Embedding Representation for Image Search using Self Supervised Means</p> <p>Image search interfaces either prompt the searcher to provide a search image (image-to-image search) or a text description of the image (text-to-image search). Image to image search is generally implemented as a nearest neighbor search in a dense image embedding space, where the embedding is derived from Neural Networks pre-trained on a large image corpus such as ImageNet. Text to image search can be implemented via traditional (TF/IDF or BM25 based) text search against image captions or image tags.</p> <p>In this presentation, we describe how we fine-tuned the OpenAI CLIP model (available from Hugging Face) to learn a joint image/text embedding representation from naturally occurring image-caption pairs in literature, using contrastive learning. We then show this model in action against a dataset of medical image-caption pairs, using the Vespa search engine to support text based (BM25), vector based (ANN) and hybrid text-to-image and image-to-image search.</p> <p>Sujit Pal Location: Theater 4</p>	<p>An approach to modelling implicit user feedback for optimising e-commerce search</p> <p>More than other domains, e-commerce search depends on implicit user feedback to optimise search result ranking as buying decision criteria such as 'an attractive price' and 'brand sympathy' are very hard to make explicit. On the other hand, this decision making can be observed implicitly in web tracking.</p> <p>E-commerce search cannot just use more generally known approaches to click modelling. For example, the common assumption that users would view search results sequentially doesn't hold for grid layouts and our model will have to deal with further contextual biases such as the device type or even time of the day.</p> <p>In this talk, I shall introduce an approach to using implicit user feedback based on Bayesian hierarchical modelling. It will provide a solution for dealing with contextual biases more generally. The model will cope with varying quantities of observations and it allows to incorporate different types of events, such as clicks and checkouts.</p> <p>René Kriegler Location: Theater 5</p>
11:00am- 11:45pm EDT	<p>Beyond precision and recall - ensuring 'aboutness' in topical classification using confidence scores</p> <p>Taxonomies play an important role in many LexisNexis products, allowing our customers to run searches using predefined topics either as pre or post-filters. Because our topical classification is automated, there can be a wide array of relevance in results, from the very relevant to the more marginal. We need to ensure that documents containing a heavy breadth and depth of discussion of a particular topic surface at the top of the results list. This</p>	<p>Scalable Semantic Search for Online Learning Applications</p> <p>Semantic search is one of the Course Hero's key products where a student can type her course's question and get an answer from hundreds of millions answered questions. The sentence embedding model sits at the core of the semantic search by which we generate a vector index for questions. One challenge is to pick the best embedding model in terms of accuracy & running-time. We have designed an evaluation framework where we use Quora duplicate questions and Faiss</p>



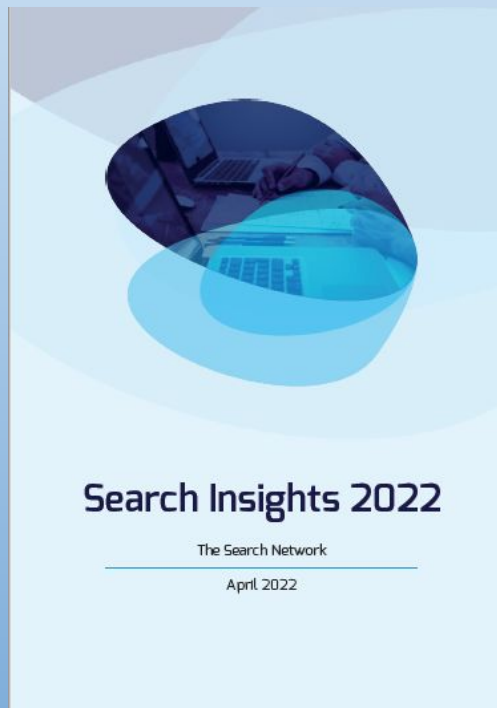
The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com
#haystackconf

Search Insights 2022

Free 78-page PDF report!

- 78 pages of articles, glossary, vendor list
- By Eric Pugh & Charlie Hull of OSC, Doug Turnbull, Tim Allison, Martin White and many other search professionals
- All Creative Commons licensed so please share
- Special early edition for Haystack - get yours on Relevance Slack in #general



And now for something
completely different...



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com
#haystackconf

Two sides of my life




Charlie Hull
UK DIRECTOR

[Email](#) | [Twitter](#) | [Github](#)

+44 (0)8700 118334

[LEARN MORE](#)



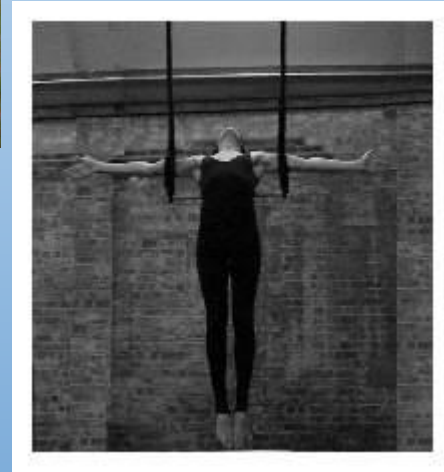
Relevancy, Solr, Elasticsearch, Search Strategy, Open Source,



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com
#haystackconf

Two sides of my life



grip
circus
theatre

HAYSTACK

The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com
#haystackconf

1. Magicians and Jugglers: Closed and Open



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

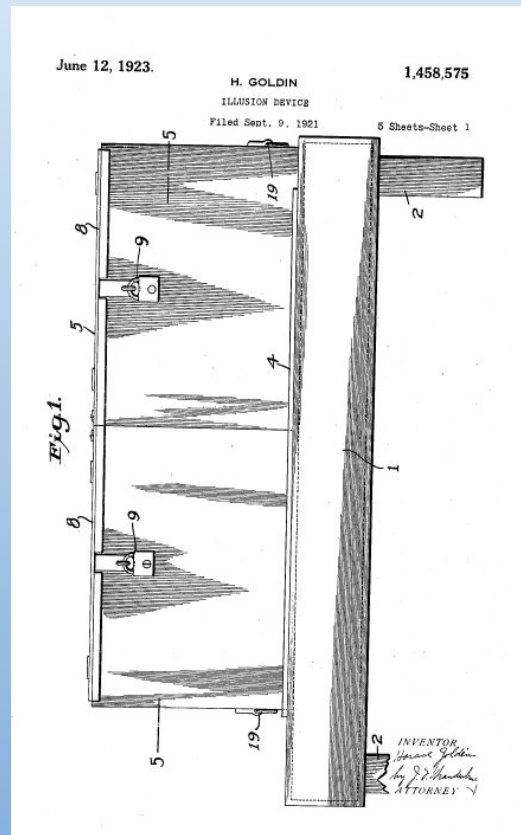
www.haystackconf.com

#haystackconf

Magicians protect their secrets



P.T. Selbit performing the 'saw a person in half' trick - Jasper Maskelyne - Maskelyne's Book of Magic. David McKay Company, 1937.



U.S. patent number 1,458,575 - Horace Goldin, 12 June 1923

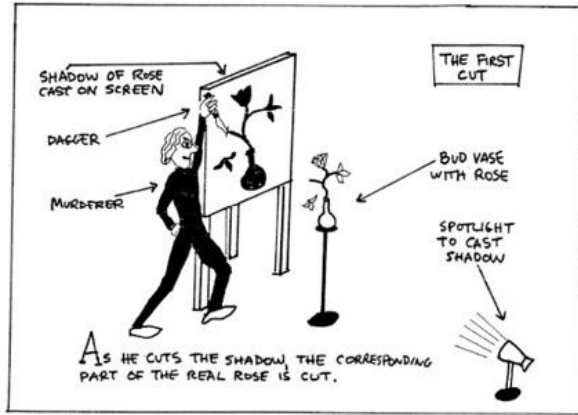
Magicians protect their secrets

D. Except as noted, members shall not reveal the modus operandi of magic effects or principles to the lay public. Exceptions are (1) where the intent is clearly to teach, so that the recipient may learn how to make and/or perform the effect or effects, rather than simply to satisfy curiosity; (2) where the method of an effect is integral to the plot, as in a detective novel or play; (3) where the method is important to the history of the performer or inventor or the evolution of the effect under discussion, as in a magic history or biography; and (4) where an effect is performed and then the simple method is revealed as a gag or stunt. However, the Ethics Committee, with the approval of the Board of Trustees, may grant a member the privilege of a specific exception if they decide that the Art of Magic is not injured thereby.



Teller sues Gerard Dogge

He looks down at the (stage right) branch of the rose. He looks back at the shadow of that branch. He raises the dagger.



He places the tip of the dagger on the paper screen, just at the point where the shadow of the branch joins the shadow of the leaves. He stabs gently through the paper.

“Teller’s partner, **Penn Jillette**, has publicly stated, “No one knows how *Shadows* is done, and no one will ever figure it out.”

Well, another magician tried.

Belgian entertainer **Gerard Dogge** posted a YouTube video of an illusion called *The Rose & Her Shadow* and offered to reveal the secrets for \$3,050”

<https://www.hollywoodreporter.com/business/business-news/teller-wins-la-wsuit-copied-magic-690347/> (2012)

“LAS VEGAS (CN) - A federal judge ordered a Belgian magician to stop selling a magic prop that infringes on a trick created by Raymond Teller, of Penn & Teller, and awarded Teller \$545,000 in damages and costs.”

<https://www.courthousenews.com/teller-gets-545000-in-magic-fight/> (2014)

Teller’s patent for the *Shadows* trick

So is magic ...
closed source?



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com
#haystackconf

2. Juggling is different



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

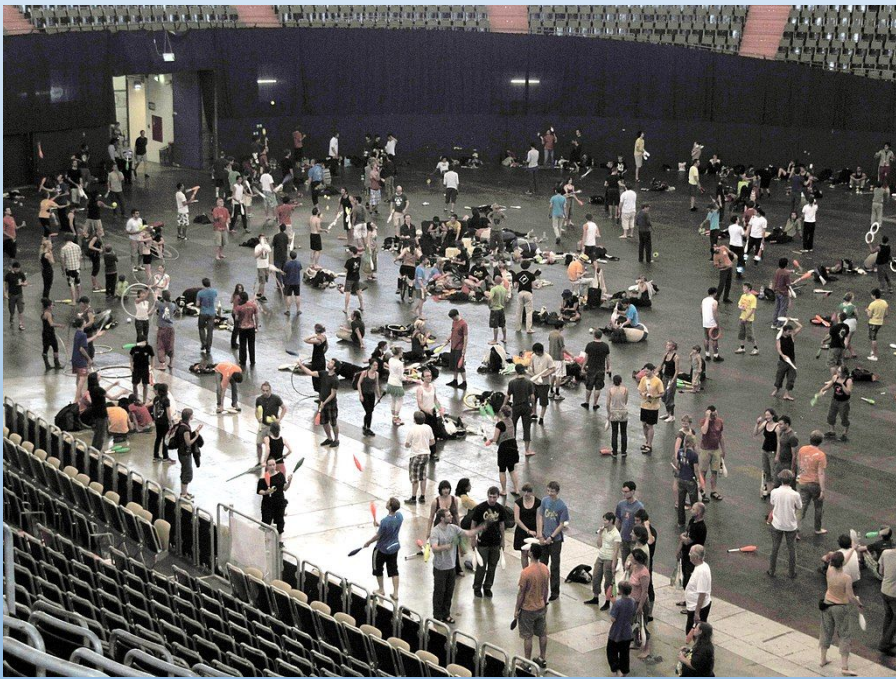
www.haystackconf.com

#haystackconf



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com
#haystackconf



HAYSTACK

The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com
#haystackconf

Equipment	Record	Juggler	Year
14 beanbags	14 catches	Alex Barron	2017
13 beanbags	15 catches	Alex Barron	2013
12 beanbags	20 catches	Alex Barron	2017
11 beanbags	33 catches	Alex Barron	2017
10 beanbags	40 catches	Tom Whitfield	2021
9 balls	55 s	Anthony Gatto	2006
8 balls	1 min 13 s	Anthony Gatto	2006
7 balls	16 min 25 s	Adolfo Esteban Cardenas	2019
6 balls	25 min 17 s	Adolfo Esteban Cardenas	2019
5 balls	2 h 41 min 27 s	Ofek Snir	2016
4 balls	2 h 46 min 48 s	Zdeněk Bradáč	2010
3 balls	12 h 5 min	David Slick	2009

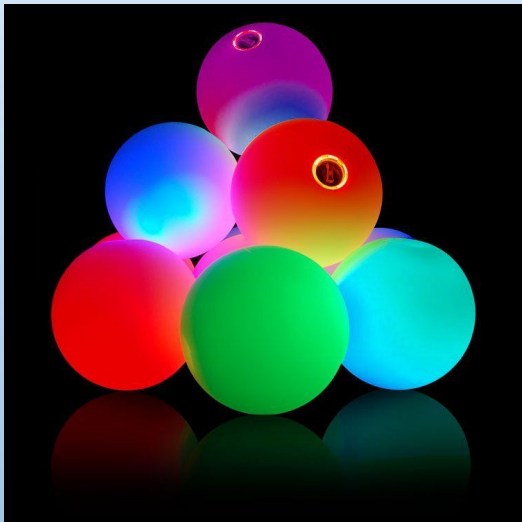
<https://www.youtube.com/watch?v=WK6AlhTjUbY>

From Wikipedia https://en.wikipedia.org/wiki/Juggling_world_records



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com
#haystackconf



HAYSTACK

The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com

#haystackconf

Generosity... & allowing failure



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com

#haystackconf

3. Ghost of Haystacks Past



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com

#haystackconf

We need to commoditize repeat/busy work

*The first
Haystack in
2018*

*Doug
Turnbull's
Keynote*

Plunger Invented
(1850-1900)

Plunging, a
plumber's
method

Anyone can
plunge their toilet



Ideation

Industry

Commodity

#haystackconf



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com
#haystackconf

Our perspective: turn silos into plungers



Protect Sea Turtles

What wheels do we see reinvented?

Who can we partner with to fund/build open source?

How can we market/community build?

How can we add product vision for broader adoption?

#haystackconf

*The first
Haystack in
2018*

*Doug
Turnbull's
Keynote*



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com
#haystackconf

4. How far we've come



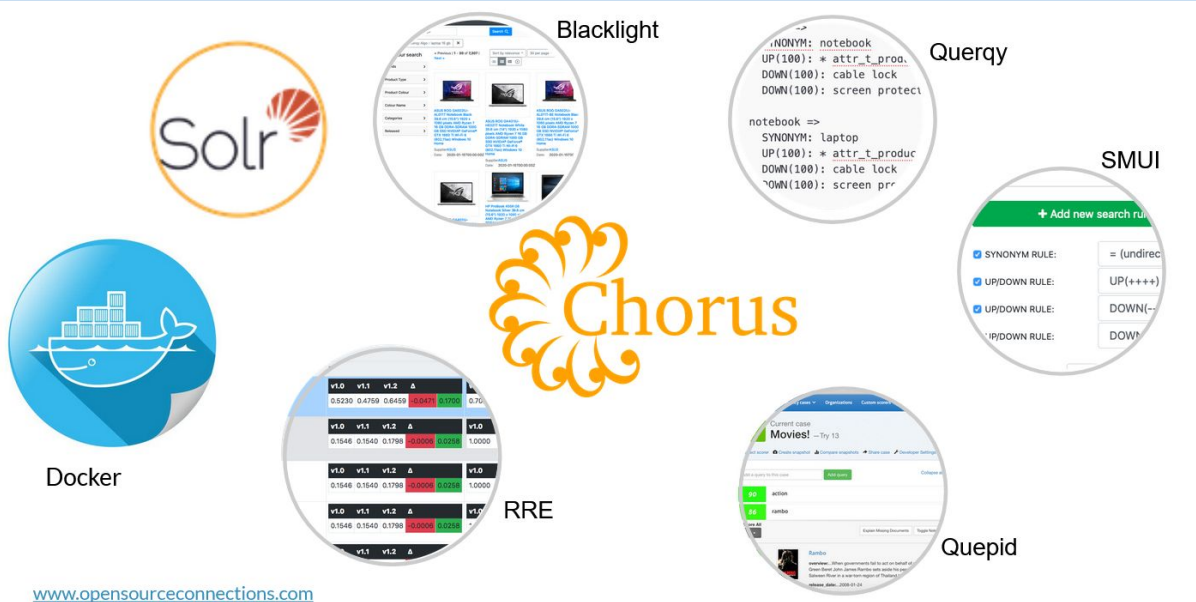
The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com

#haystackconf

Quepid *show your queries some* ❤️

2018: closed source, paid-for SaaS version



2022: Open source reference implementation for e-commerce search (Elasticsearch and OpenSearch versions on the way!)
SaaS version of Quepid is free (and much improved since then)



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com
#haystackconf

Search Relevance & Matching Tech

#general Information retrieval, relevance, recommendations, traditional search, voice search, bots, and all the cool stuff one can build!

1 Pinned + Add a bookmark

Friday, March 25th

Kelsey Juraschka 7:00 PM
joined #general along with 5 others.

Wednesday, March 30th

Meghan Boyd 4:46 PM
Hi Everyone- curious if anyone could share with me a good example of hit highlighting for an e-commerce site? It doesn't seem to be that common but if anyone knows of a good reference implementation.
4 replies Last reply 9 days ago

Charles Martin 4:55 PM
A nice review of Transformers for ranking (for those who have not seen it already)
<https://arxiv.org/pdf/2010.06467.pdf>
3

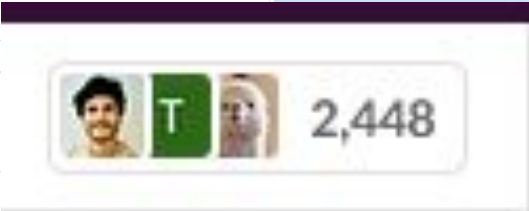
Robyne Gallacher 5:33 PM
joined #general along with Wilson Fearn.

Lisa Biella 7:11 PM
We are proceeding to the second talk of the London Information Retrieval Meetup, right now!
"How to Cache Your Searches: an Open source Implementation" from @Daniele Antuzi, R&D Software Engineer at Sease Ltd
Register here and join us on zoom: <https://www.meetup.com/London-Information-Retrieval-Meetup-Group/events/284508442/>

Meetup
London Information Retrieval Meetup March 2022 (HYBRID), Wed, Mar 30, 2022, 6:15 PM | Meetup
Wed, Mar 30, 6:15 PM BST: We are so happy to announce the twelfth London Information Retrieval Meetup, a free evening meetup aimed to Information Retrieval passionate and professionals who are curi (25 kB)

LONDON INFORMATION

Message #general



A few hundred in 2018?



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com
#haystackconf



Haystack LIVE! Search Relevance Talks Online

📍 Charlottesville, VA
👤 1,026 members · Public group
👤 Organized by **Charlie Hull** and 1 other

[Edit group info](#)

Share: [f](#) [t](#) [in](#) [✉](#)

[About](#) [Events](#) [Members](#) [Photos](#) [Discussions](#) [More](#)

Create event

Manage group

What we're about

This is a group for anyone interested in search relevance, relevance engineering, natural language search, Learning to Rank, machine learning fo...

[Read more](#)

Upcoming events (1)

[See all](#)

THU, APR 14, 2022, 4:00 PM BST

Elasticsearch & AI powered Cisco Enterprise Search- Sujith Joseph & Prem Malhotra

Online event

We built an Enterprise Search platform using GCP services and Elasticsearch and

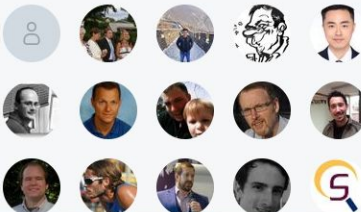
Organizers



Charlie Hull and 1 other
[Message](#)

Members (1,026)

[See all](#)



Members (1,026)

25 past events
All available on Youtube
Some attracted 120+ people



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com

#haystackconf

5. New magic



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com

#haystackconf

 deepset

 MILVUS 2.0


Weaviate
Open Source

 Elasticsearch

 Hugging Face

 jina

 Pinecone

 OpenSearch

 SI TECHNOLOGY

 drant

 vespa

...

“Smaller vector DB players: 71% are open source”
(Dmitry Kan)

5. What can we do?



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com

#haystackconf

Publish

Show

Let's share!

Give

And don't be afraid to drop, every
mistake teaches you something!

Be more like a juggler than a magician!



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com

#haystackconf

Thank you for listening!

Twitter: @FlaxSearch



The Search Relevance Conference
27/28 April 2022 - Charlottesville, VA
Join Relevance Slack at www.opensourceconnections.com/slack

www.haystackconf.com

#haystackconf