

# INTRODUCING GAR



GENERATIVEAI AUGMENTED (LEXICAL) RETRIEVAL

Sept, 2024

Guest Speaker:

*Systematically Improving RAG Applications*

# Obligatory Bio Slide

👋 Hi I'm Doug  
(@softwaredoug everywhere)

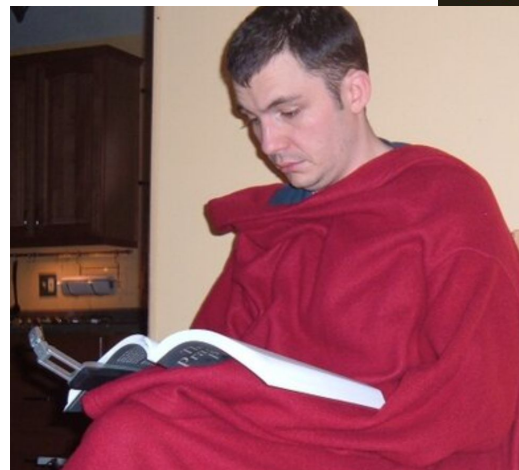
Long-time search enthusiast... Not  
yet (never?) an expert

I wrote some search books, did some open  
source

I work at Reddit

I worked at Shopify & OpenSource Connections  
in search

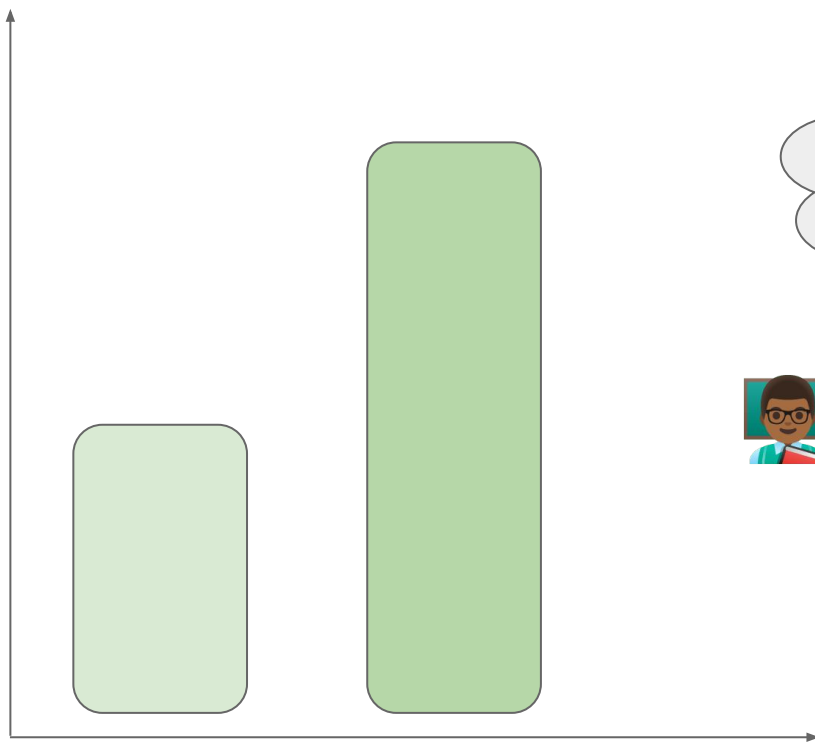
I blog here: <http://softwaredoug.com>



:itme:



# Search conference talks



BM25

MyVectorSearchTransformerRankerTwoTowerCrossEncoderThingy



As you can see, we  
have redefined  
search for all  
time

# Beautiful experimental datasets

Passage	Text
1234	Network switches, functions & role in networks. 1 1.2 Network Switches: Functions & Role in Networks A network switch or switching hub is a computer networking device that connects multiple computers together within one local area network (LAN).
5678	American Board of Integrative Holistic Medicine. Offers CME credits for The 14th Annual Science and Clinical Application of Integrative Holistic Medicine. This program is completed exclusively online and physicians can claim up to 60.5 AMA PRA Category 1 credits.

# ... VS search reality



PM

Why aren't these results coming back

Huh these products are missing a description!?



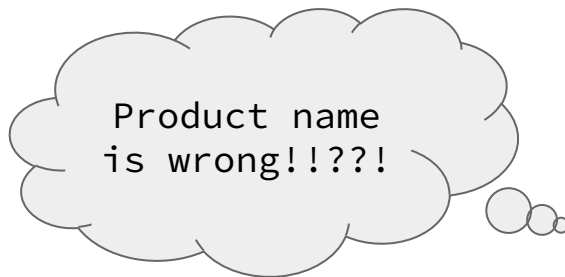
Dev/DS

	product_name	product_description	grade	product_id
37	betsy 31.5 " w writing desk and chair set	None	2.0	6974
38	biergh wooden 31 " writing desk and chair set	None	2.0	41138

# ... VS search reality



PM



Dev/DS

**product\_name**

**product\_description** **grade**

<b>2</b>	cordova kids desk	this <b>office chair</b> is perfect for the kids and ...	0.0
<b>3</b>	brister kids desk	if you are looking for a <b>proper chair set</b> for ...	0.0

# Conclusion: actually look at your data

1. **LABEL** - simple hand labels, to engagement based, to LLM based  
(<https://softwaredoug.com/blog/2021/02/21/what-is-a-judgment-list>)
2. **DEBUG** - Figure out why ranking is boring BM25 baselines
3. **FIX:** data problems lurk leading to easy lexical wins (LLMs can help!)

# GARTHODOLOGY

prototyping  
before betting  
the farm





# Start small

- Start with a **handful of queries** – for early debuggability/exploration

desk for kids		
jordanna solid wood rocking		
kari 2 piece		
...		



# Use (n)DCG



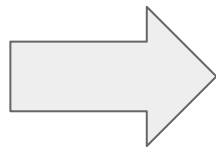
- You probably know what NDCG is by now :)

Query	Title	Grade	Rank	Gain
desk for kids	Kids desk	2	1	3.2
desk for kids	Taco stand	0	2	0
desk for kids	Kids study guide	1	3	0.5
...	...	...	...	...

$$\text{NDCG@3} = \frac{\text{DCG}}{\text{idCG}}$$
$$= \frac{3.2+0+0.5}{\text{idCG}}$$

# Use (n)DCG-f (for filter)

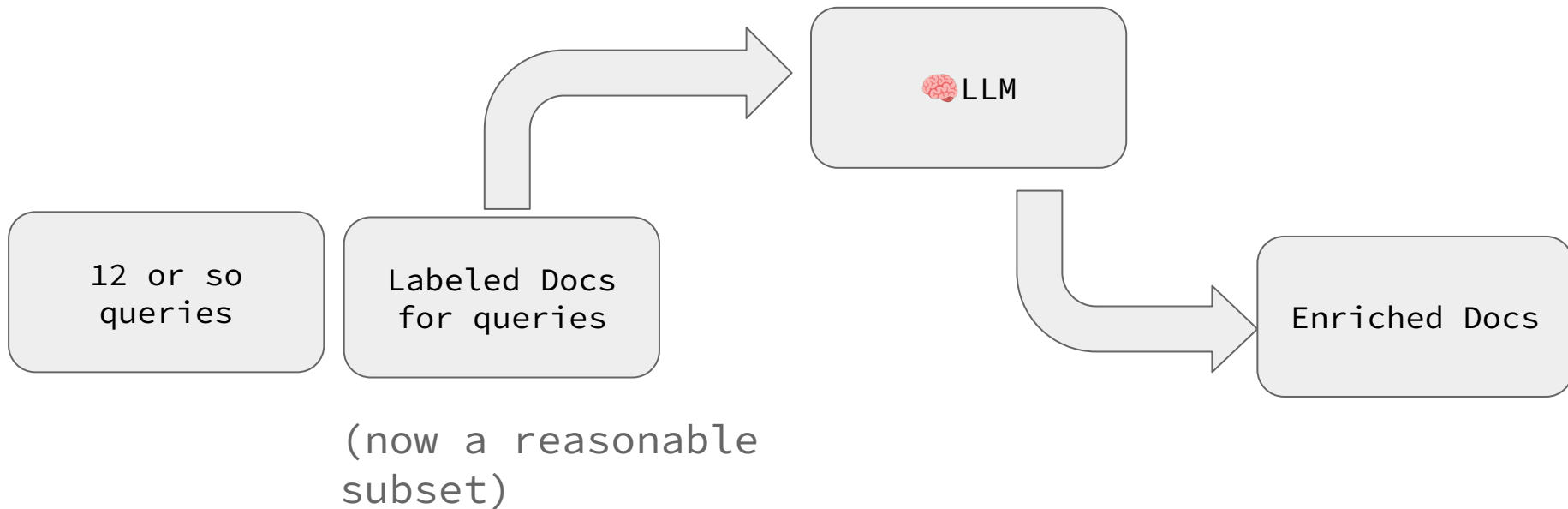
Query	Title	Grade	Rank
desk for kids	Kids desk	2	1
<del>desk for kids</del>	<del>Barbie dream desk</del>	??	2
desk for kids	Kids study guide	1	3
desk for kids	Desk chair	1	4



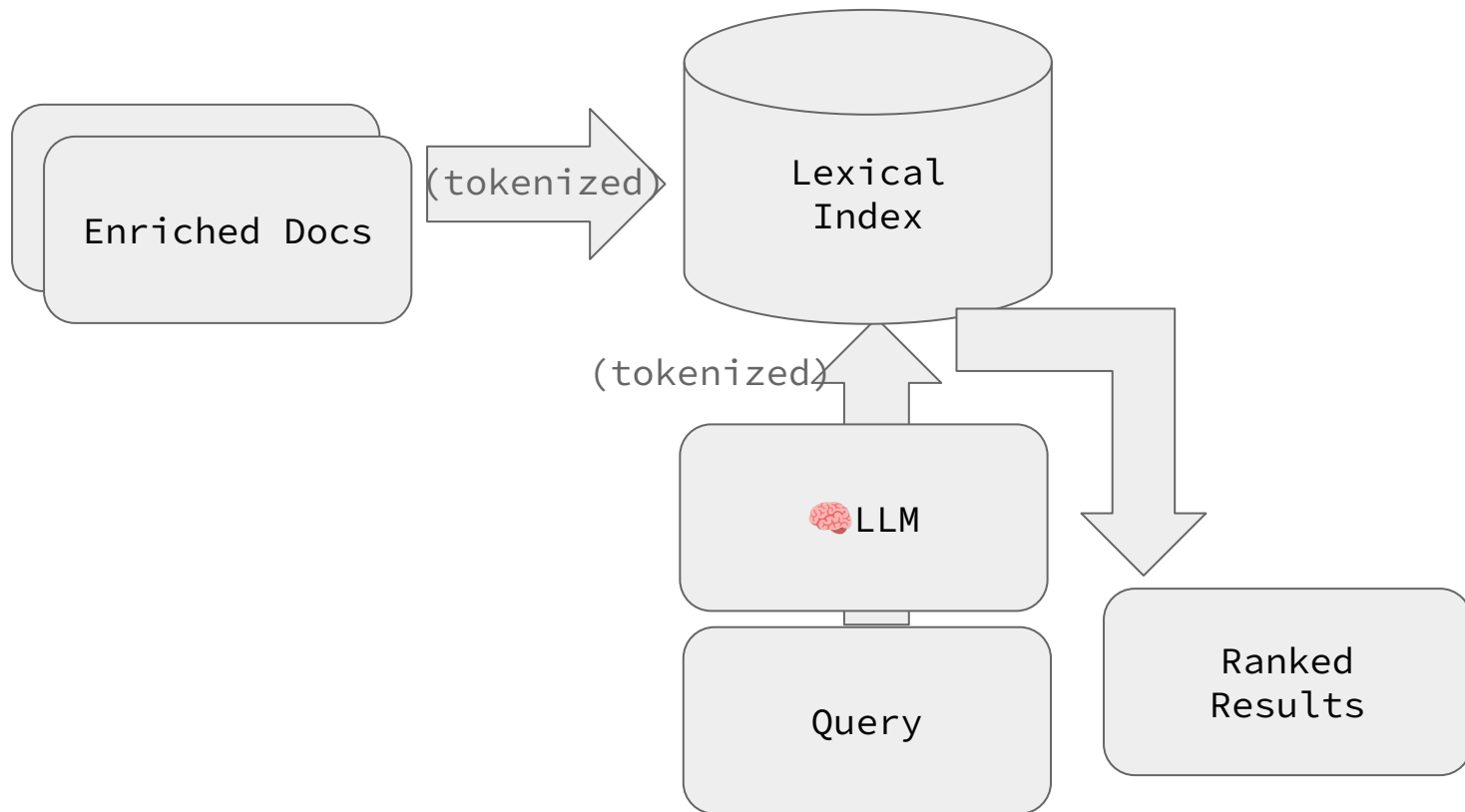
Query	Title	Grade	Rank	Gain
desk for kids	Kids desk	2	1	3.2
desk for kids	Kids study guide	1	2	1.7
desk for kids	Desk chair	1	4	0.5

Within the universe of labeled results: do we move relevant above irrelevant?

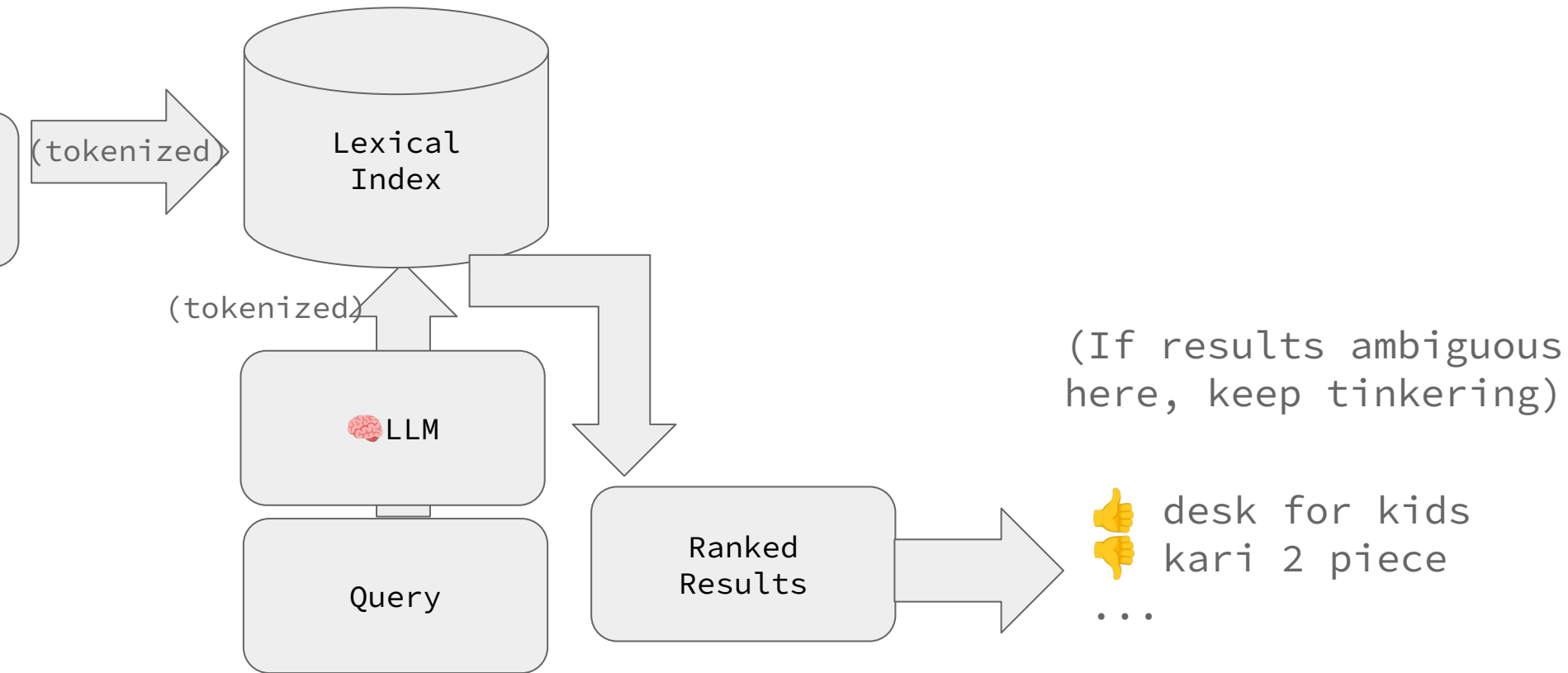
# Now with LLMs...



# Now search

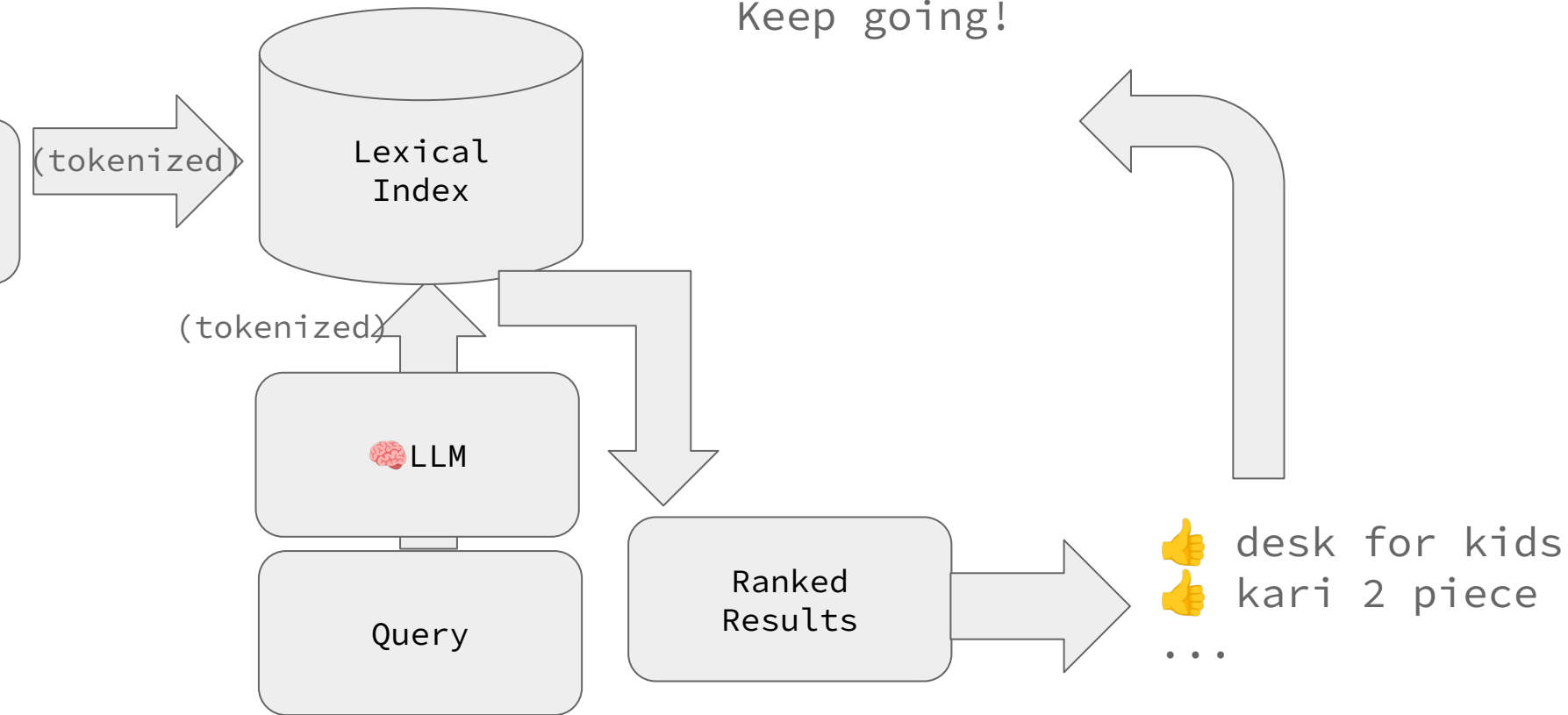


# Eval, rinse, repeat



# Scale up

Add more queries... + their docs!  
Keep going!



FIXING THE PROBLEM



# Boring BM25 baseline

for each tokenized term  
score += Best BM25 of title/description for term

```
▶ def ndcg_m(results):  
    max_dcg = np.sum((2**(np.ones(10) + 1)) / np.log(np.arange(10) + 2))  
    results['gain'] = (2**results['grade'] - 1) / np.log(results['rank'] + 1)  
    dcgs = results.groupby('query')['gain'].sum()  
    dcgs = dcgs.sort_values()  
    return dcgs / max_dcg
```

```
ndcg_ms_baseline = ndcg_m(results_baseline)  
ndcg_ms_baseline.mean(), ndcg_ms_baseline
```

```
↔ (0.3236988733393219,
```

[NDCG-m, see “flavors of NDCG”](#)

[WANDS Baseline Notebook](#)

WANDS - Wayfair furniture search dataset (<https://github.com/wayfair/WANDS>)

# Prompt to clean title



LLM

Create a very short (3-4 word) title for product descriptions. Prefix your response with "Title:"

Here are some examples

Product Description: simple student table kids desk white with drawers made of high-quality material , with drawers and storage functions , it is suitable for your kids . our table and chairs set are ideal for study , playing games , working on puzzles , and more .

Title: Kids' Study Desk with Storage

...

Constrain "title" output to simple "just the facts"

Use only description, not existing title

Examples of good / bad titles

# Cleaned title

	(Before)	(Cleaned)
	<b>product_name</b>	<b>llm_title_fixed</b>
<b>product_id</b>		
<b>17053</b>	kids desk	2-in-1 Booster Seat
<b>18018</b>	cordova kids desk	Adjustable Kids' Vanity Chair
<b>8434</b>	brister kids desk	Iron Kids' Chair Set

# Simple fix, nice NDCG bump

```
ndcg_ms_test.mean()
```

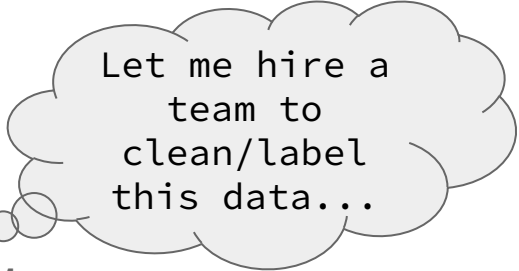


0.3669798345935841

	gain
query	
desk for kids	0.310413
jordanna solid wood rocking	0.000000
oriental vanity	0.000000
nectar queen mattress	0.000000
bedroom wall decor floral, multicolored with some teal (prints)	0.020583
48 in entry table with side by side drawer	-0.032749
alyse 8 light	0.000000
kari 2 piece	0.000000
tufted upholstered bed diamond	0.091282

# Today's search management

## Ye Olde Times

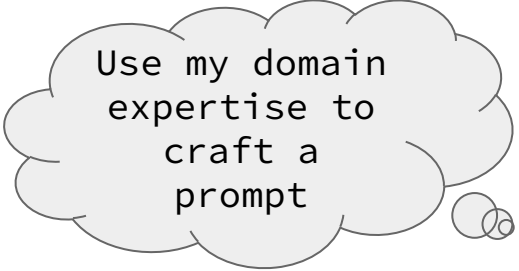


Let me hire a  
team to  
clean/label  
this data...



PM

## Today



Use my domain  
expertise to  
craft a  
prompt

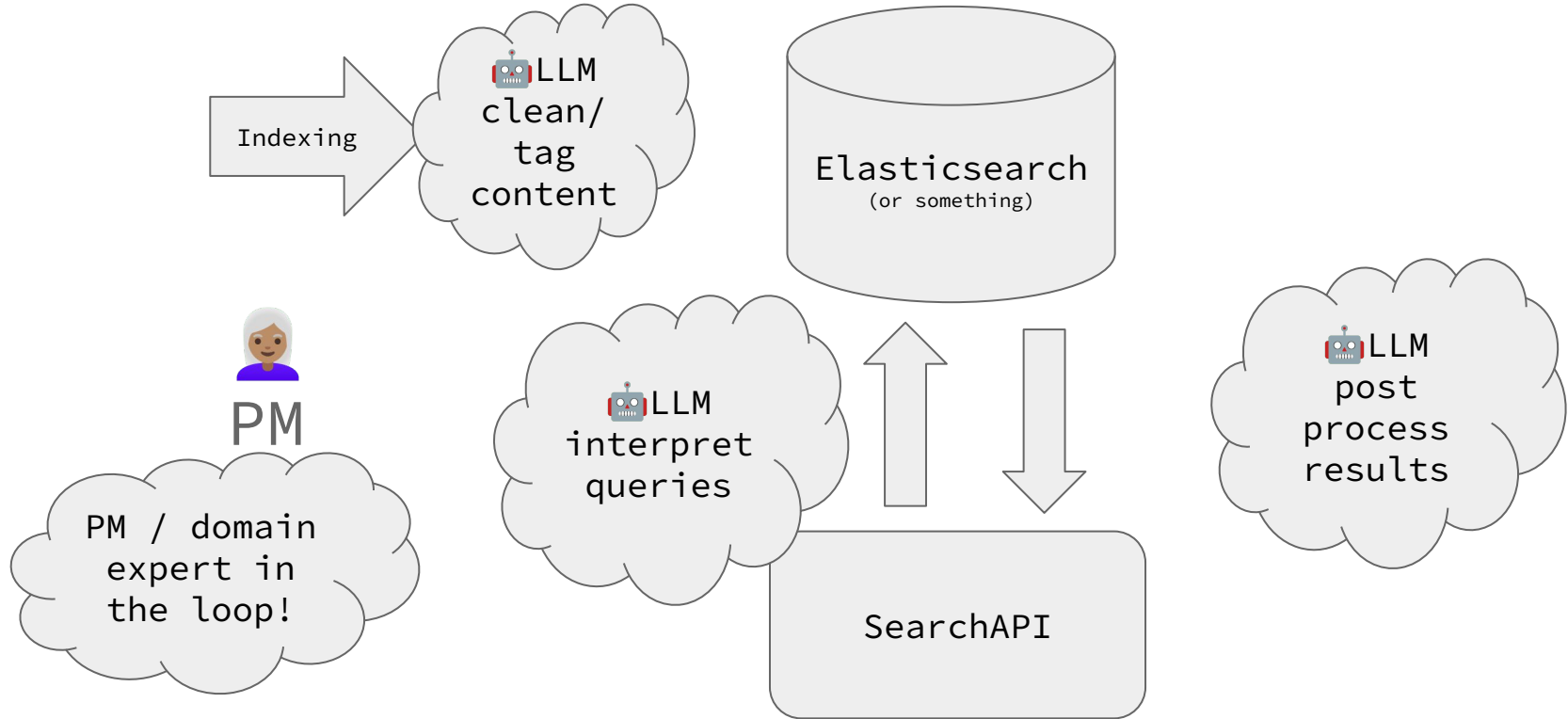


PM

# Challenge: fill in descriptions

- Can you craft a prompt to fill in empty descriptions?
- What would be a good way to do this?

# Hence GAR - GenerativeAI Augmented (lexical) Retrieval



SYNTHESIZING  
SIMILARITY



# Big problem: vocabulary gaps

queries

formal light fixture

traditional lighting

...

Product Name: alyse chandelier

Product Description: This beautiful, elegant chandelier



PM

Hmm, a vocab mismatch!

# Every search team: LLM to make synonyms



LLM

You are an e-commerce search product SEO assistant of a furniture store.

I want you to generate synonyms for query terms I provide. Specifically most important synonyms are

- \* style
- \* material
- \* color
- \* item type

Do not try to generate synonyms beyond these, ie don't try to interpret synonyms for product lines.

...

queries

formal light fixture

traditional lighting

traditional->elegant, formal  
light fixture->chandelier...



PM

# Blind synonyms: kind of mid

## ✓ Queries improved / harmed

```
▶ delta_ndcg = ndcg_ms_test - ndcg_ms_baseline[test_queries]  
delta_ndcg.sort_values()
```



	gain
query	
tufted upholstered bed diamond	-0.060297
48 in entry table with side by side drawer	-0.004823
bedroom wall decor floral, multicolored with some teal (prints)	-0.000703
jordanna solid wood rocking	0.000000
oriental vanity	0.000000
nectar queen mattress	0.000000
alyse 8 light	0.000000
kari 2 piece	0.000000
desk for kids	0.055494

$$\text{score} = 10 * \text{BM25}(\text{orig}) + \text{BM25}(\text{expanded})$$



## Results

Similar  
results for  
blind content  
spam  
(notebook)

# Dumb synonyms / Smart synonyms

```
[60] query_synonyms('tufted upholstered bed diamond')
```


```
API CALL!  
Expanded: tufted (buttoned, padded) upholstered bed (frame, furniture) diamond (pattern, design)  
'tufted (buttoned, padded) upholstered bed (frame, furniture) diamond (pattern, design)'
```

```
[62] query_synonyms('desk for kids')
```

```
API CALL!  
Expanded: desk (table, workstation) for kids (children, youngster)  
'desk (table, workstation) for kids (children, youngster)'
```

```
query_synonyms('kari 2 piece')
```

```
API CALL!  
Expanded: kari (modern, contemporary) 2 piece (set, pair)  
'kari (modern, contemporary) 2 piece (set, pair)'
```

 WTF this is a furniture store, everything is “furniture”

youngster!?

actually kinda smart?

Understands the product line



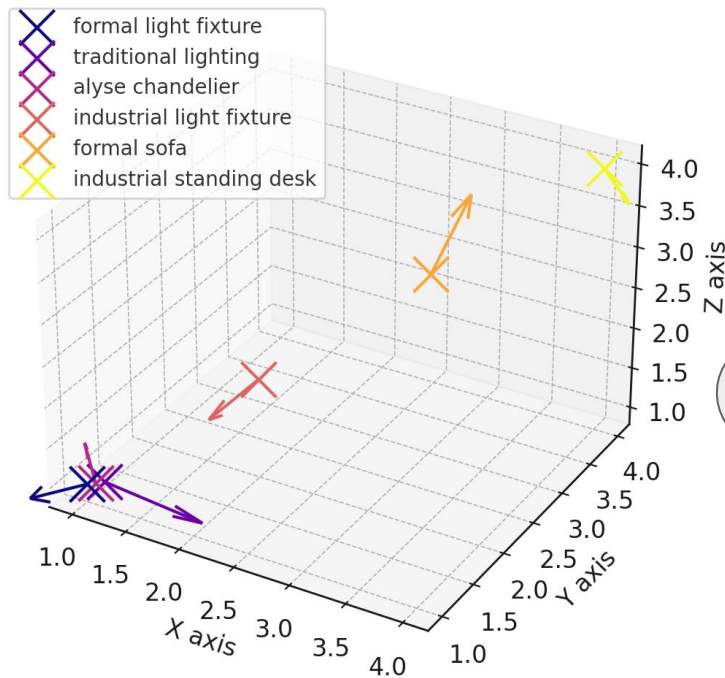
## Results

# Random synonyms, slop...

Queries

Corpus

3D Query-Document Similarity Representation with Varied Vectors



Not informed  
by actual  
query-document  
relationships

# Goal

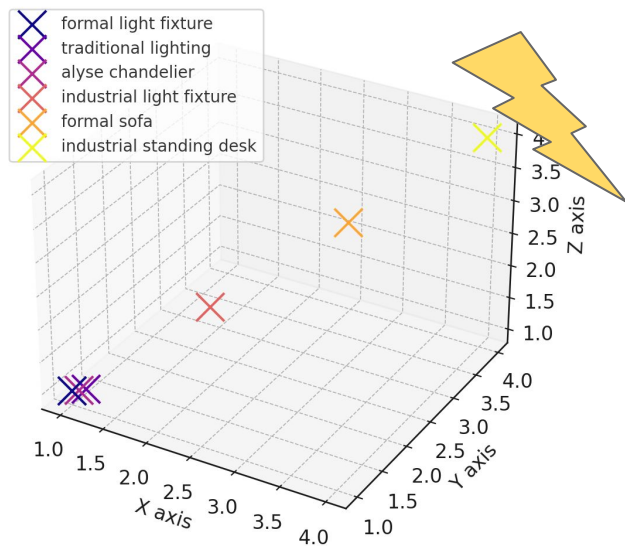
Queries



Corpus



3D Query-Document Similarity Representation



# Lexical Search Trick: Taxonomies

## Google Product Taxonomy

554 - Furniture > Baby & Toddler Furniture

6349 - Furniture > Baby & Toddler Furniture > Baby & Toddler Furniture Sets

7068 - Furniture > Baby & Toddler Furniture > Bassinet & Cradle Accessories

6393 - Furniture > Baby & Toddler Furniture > Bassinets & Cradles

558 - Furniture > Baby & Toddler Furniture > Changing Tables

7070 - Furniture > Baby & Toddler Furniture > Crib & Toddler Bed Accessories

7072 - Furniture > Baby & Toddler Furniture > Crib & Toddler Bed Accessories > Crib Bumpers & Liners

7071 - Furniture > Baby & Toddler Furniture > Crib & Toddler Bed Accessories > Crib Conversion Kits

6394 - Furniture > Baby & Toddler Furniture > Cribs & Toddler Beds

6969 - Furniture > Baby & Toddler Furniture > High Chair & Booster Seat Accessories

559 - Furniture > Baby & Toddler Furniture > High Chairs & Booster Seats

6433 - Furniture > Beds & Accessories

4437 - Furniture > Beds & Accessories > Bed & Bed Frame Accessories

505764 - Furniture > Beds & Accessories > Beds & Bed Frames

451 - Furniture > Beds & Accessories > Headboards & Footboards

2720 - Furniture > Beds & Accessories > Mattress Foundations

2696 - Furniture > Beds & Accessories > Mattresses

441 - Furniture > Benches

...

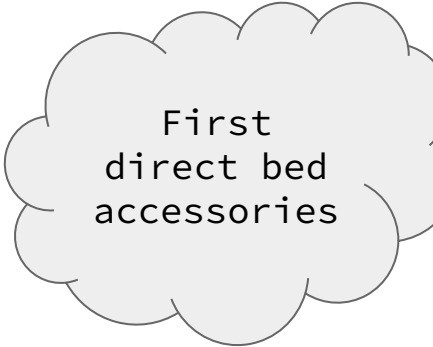
# Taxonomic similarity

Query: 7070 - Furniture > Baby & Toddler Furniture > Crib & Toddler Bed Accessories

## Most similar (children - hyponyms)

7072 - Furniture > Baby & Toddler Furniture > Crib & Toddler Bed Accessories > Crib Bumpers & Liners

7071 - Furniture > Baby & Toddler Furniture > Crib & Toddler Bed Accessories > Crib Conversion Kits



First  
direct bed  
accessories



# Taxonomic similarity

Query: 7070 - Furniture > Baby & Toddler Furniture > Crib & Toddler Bed Accessories

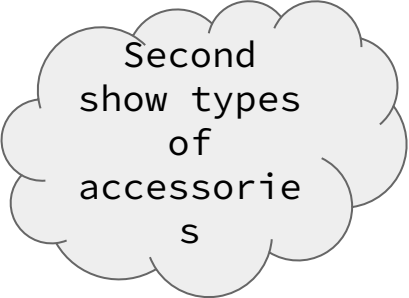
## Most similar (children - hyponyms)

7072 - Furniture > Baby & Toddler Furniture > Crib & Toddler Bed Accessories > Crib Bumpers & Liners

7071 - Furniture > Baby & Toddler Furniture > Crib & Toddler Bed Accessories > Crib Conversion Kits

## Next similar (parent - hypernyms)

554 - Furniture > Baby & Toddler Furniture



Second  
show types  
of  
accessorie  
s

# Taxonomic similarity

Query: 7070 - Furniture > Baby & Toddler Furniture > Crib & Toddler Bed Accessories

## Most similar (children - hyponyms)

7072 - Furniture > Baby & Toddler Furniture > Crib & Toddler Bed Accessories > Crib Bumpers & Liners

7071 - Furniture > Baby & Toddler Furniture > Crib & Toddler Bed Accessories > Crib Conversion Kits

## Next similar (parent - hypernyms)

554 - Furniture > Baby & Toddler Furniture

## Next most similar (siblings)

6349 - Furniture > Baby & Toddler Furniture > Baby & Toddler Furniture Sets

7068 - Furniture > Baby & Toddler Furniture > Bassinet & Cradle Accessories

6393 - Furniture > Baby & Toddler Furniture > Bassinets & Cradles

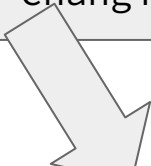
558 - Furniture > Baby & Toddler Furniture > Changing Tables

7070 - Furniture > Baby & Toddler Furniture > Crib & Toddler Bed Accessories



# Taxonomic Similarity

q=table for  
baby changing



LLM labels query

## Label with taxonomy:

558 - Furniture >  
Baby & Toddler Furniture >  
Changing Tables

Elegant  
baby diaper  
table




LLM labels product

## Label with taxonomy:

558 - Furniture >  
Baby & Toddler Furniture >  
Changing Tables

# Prompt



I'm going to give you a product description of furniture, you need to tell me the google product taxonomy the item most likely falls in, using the following subset of the taxonomy.

Prepend your response with "Taxonomy:"

If you are uncertain, please say "Taxonomy:No appropriate classification" if none found.

Here is the taxonomy to use:

{taxonomy}

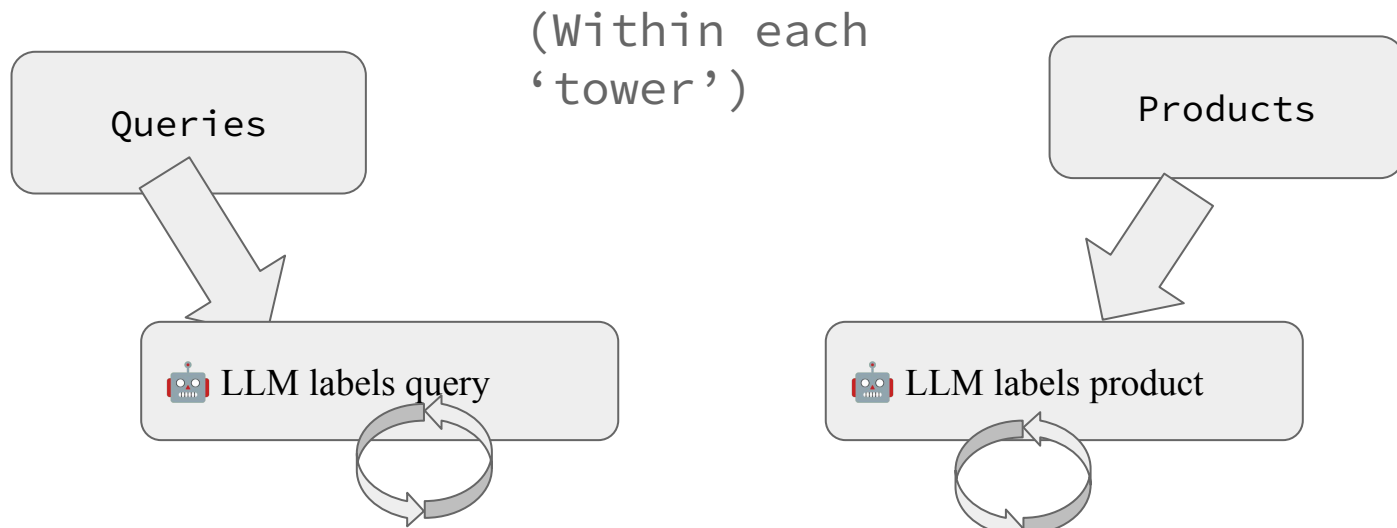
Some examples:

Insert furniture portion of google taxonomy

Allow for "I don't know" responses to avoid guessing

Examples of expected behavior / edge cases

# Measurement + Iteration



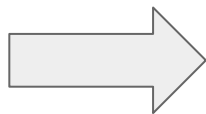
**LABEL** - create test cases, confirm labeled correctly

**DEBUG** - Monitor classifier output, update prompt with new edge cases

# Similarity - BM25 works!

Path tokenize, score with BM25

558 - Furniture >  
Baby & Toddler Furniture >  
Changing Tables



['558 - Furniture',  
'558 - Furniture > Baby & Toddler Furniture',  
'558 - Furniture > Baby & Toddler Furniture >  
Changing Tables']

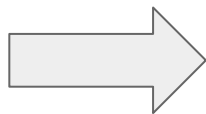
(Tokenize query / product into path  
tokens, score with BM25, get tax  
similarity :) )

Oldie but goodie talk

# Similarity - BM25 works!

Path tokenize, score with BM25

558 - Furniture >  
Baby & Toddler Furniture >  
Changing Tables



[‘558 - Furniture’,  
‘558 - Furniture > Baby & Toddler Furniture’,  
‘558 - Furniture > Baby & Toddler Furniture >  
Changing Tables’]

Lowest IDF

Highest IDF

BM25 wants to rank more  
specific nodes higher

Oldie but goodie talk

# Results

```
ndcg_ms_test2.mean()
```

```
0.3558915078832073
```

	gain
query	
48 in entry table with side by side drawer	0.000000
alyse 8 light	0.000000
bedroom wall decor floral, multicolored with some teal (prints)	0.113597
desk for kids	0.231159
jordanna solid wood rocking	-0.055023
kari 2 piece	0.000000
nectar queen mattress	0.000000
oriental vanity	0.000000
tufted upholstered bed diamond	0.000000



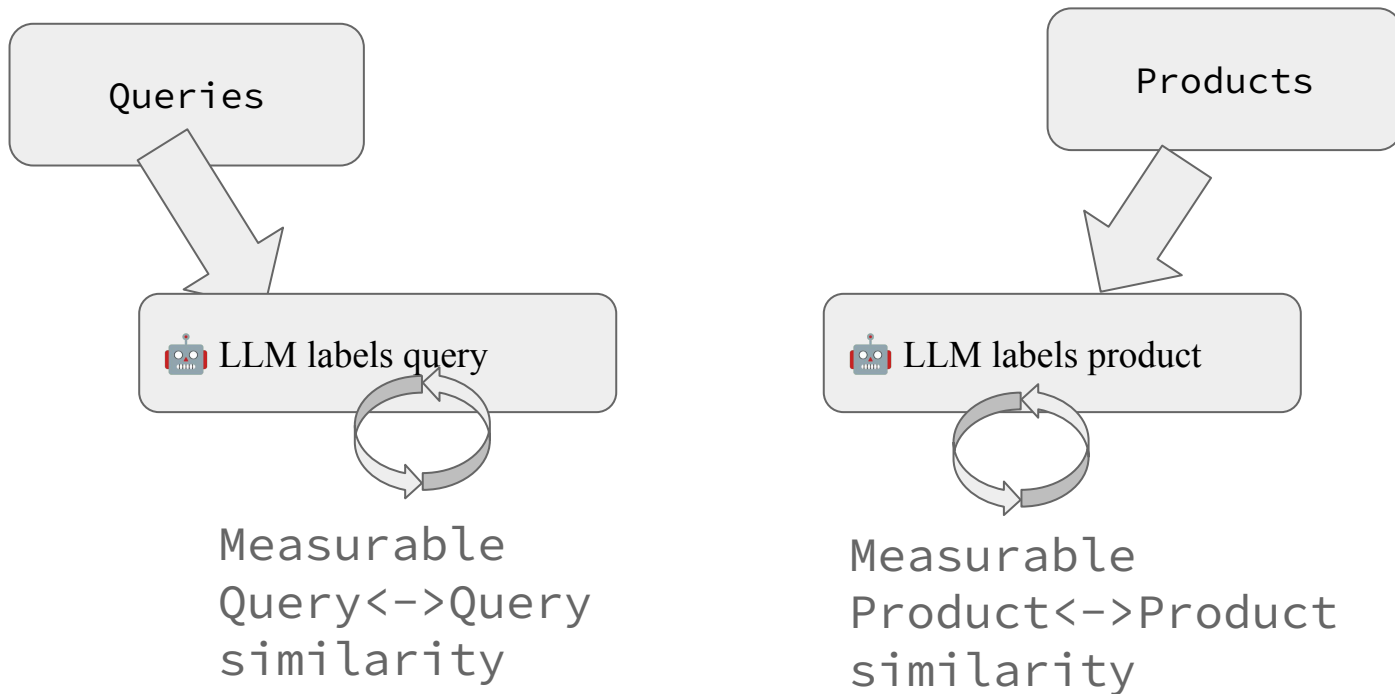
# Adding cleaned title?

```
ndcg_ms_test3 = ndcg_m(results_test3)
ndcg_ms_test3 = ndcg_ms_test3[test_queries]
ndcg_ms_test3.mean()
```

⇒ 0.3689556300982193

(Slightly additive, probably not statsig)

# “Two Tower” approaches work in lexical too



*Other dimensions: color, size, etc*

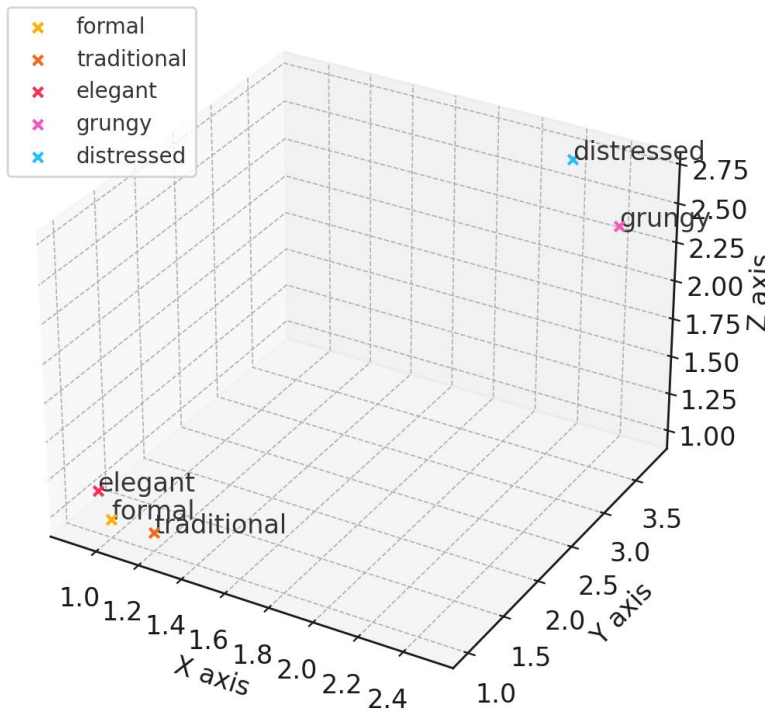
# What similarities are there?

- Direct text match (BM25, or other)
- Fuzzy / typo probability
- Many taxonomic classifications (color, etc)
- Numerical attributes (size, etc)
- User attributes

CONCLUSIONS

# Not a replacement for vector search

3D Word Embedding Representation



# Many teams aren't looking at their data

- No way to debug individual queries+look for patterns
- Yet there are often low hanging fruit in your dataset!
- IR research != your dataset - you need to find *in-domain* fixes

# Many teams can't prototype

- Do you have a kaggle-able version of your search?
- Can you prototype search in a colab notebook?
- Can you measure a handful of queries, with only a few docs labeled?

*Often experiment == start with A/B test right away, when it should mean gradual increase in queries from 10s->100s-> ... -> millions during development*

# LLMs augment old school search hacks

LLMs keep me in the  
loop with  
traditional search  
management!



PM

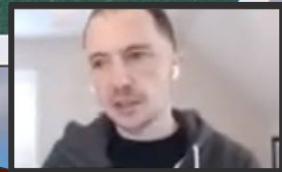


If you have questions i might answer them i might not haha

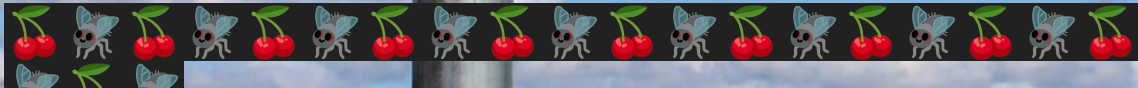


# Question

Slide by ian turnbull... again. (if you havent thrown rotten fruit and stuff at my dad please do it now)



Examples of rotten fruit



# Notebook links

## Baseline

- [WANDS Baseline](#)

## Naive (didn't work)

- [Naive content generation](#)
- [Naive \(query\) synonym generation](#)
- [“Blind” synonym generation](#)

## What worked

- [Cleanup title \(LLM generates new title\)](#)
- [Use Google Product Taxonomy](#)