



Lazarina Stoy

SEO & Data Science
Consultant

Panel: Content Marketing & Semantic SEO

How to automate your content audit with ML models



content audit process

(how it typically goes)

Organize your database – crawl,
content, metrics



Understand the data – Categorize your
inventory



Evaluate performance and assign
action items to underperforming content



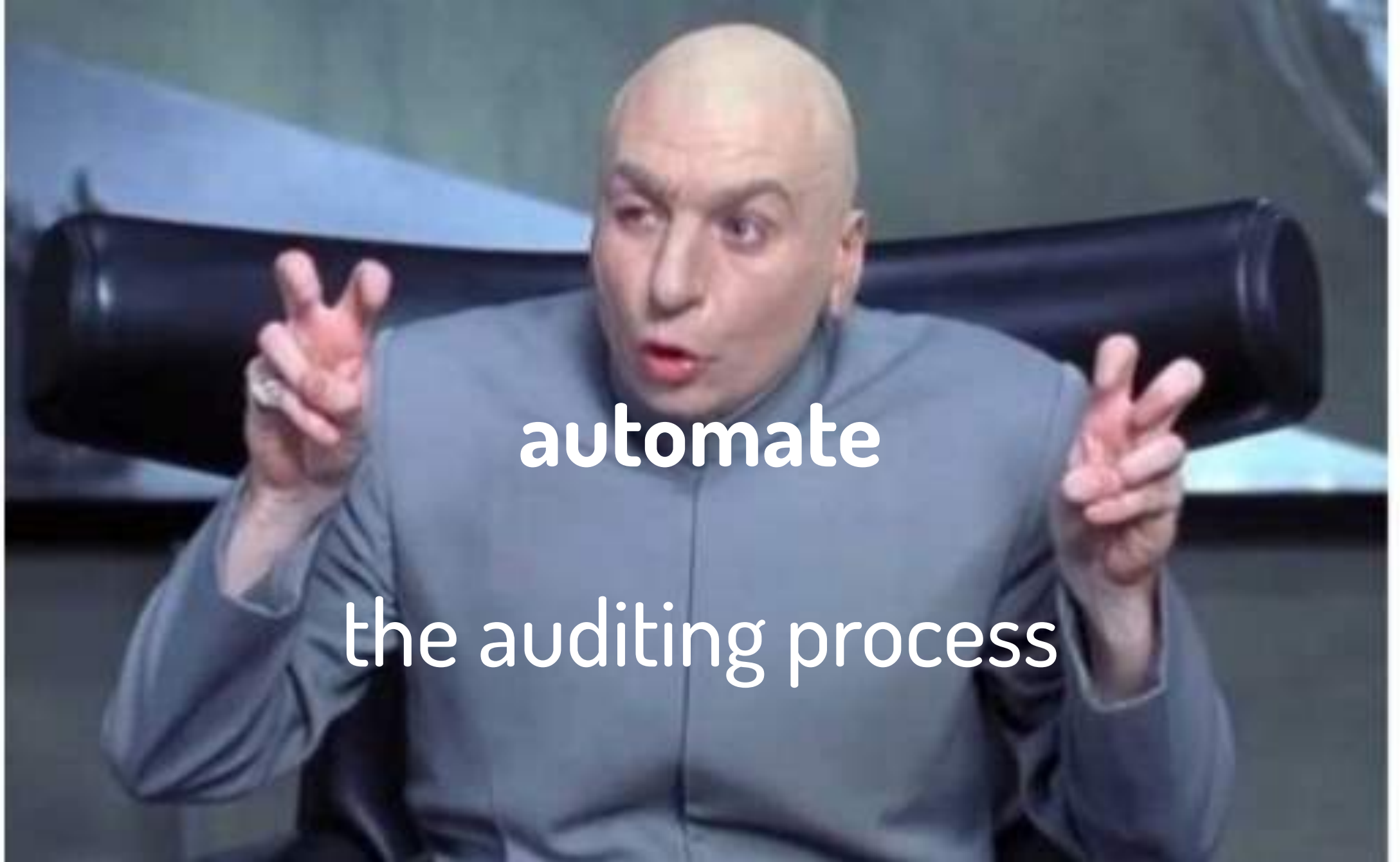
Implement & Review impact – Prioritize
what and when to implement and repeat



so, we'll go through the same steps

build better **systems**

implement ML APIs



automate

the auditing process

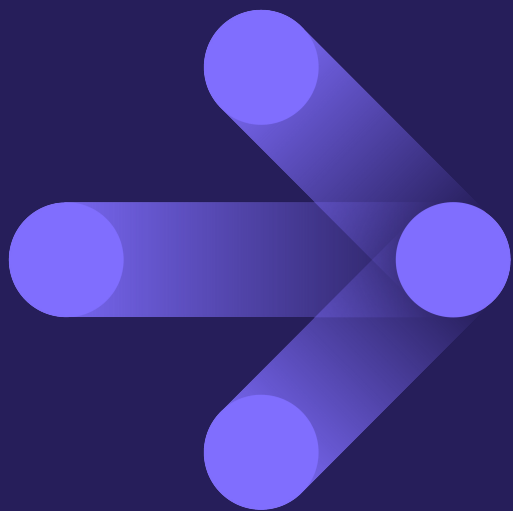
ultimately...
to create a **better** audit, **quicker**

Organize your database – crawl, content, metrics

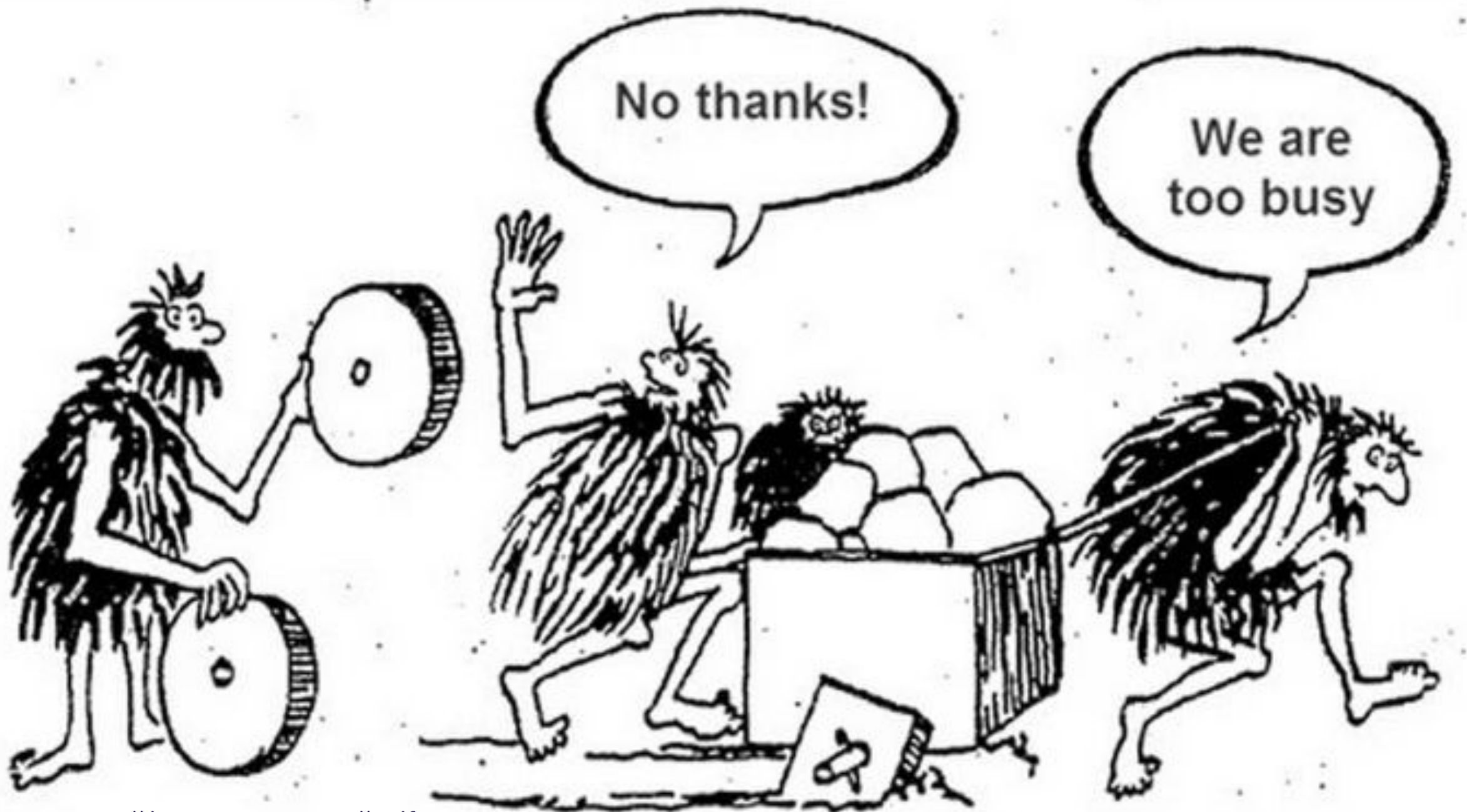
Understand the data – Categorize your inventory

Evaluate performance and assign action items to underperforming content

Implement & Review impact – Prioritize what and when to implement and repeat



Crawling and scrapping content



normal sites



sites with > 500K pages



highly recommended if
you're going route two



brighton
SEO. April 2022

How to build your own crawler, and why you should give it a try

Jess Peck | Local SEO Guide

 @jessthebp

 slideshare.net/JessPeck2/

goals at this stage are pretty simple

01

consistency

you need consistent data scraping across your entire website

02

accuracy

you need the elements you select to be accurately pulling data throughout the entire website

03

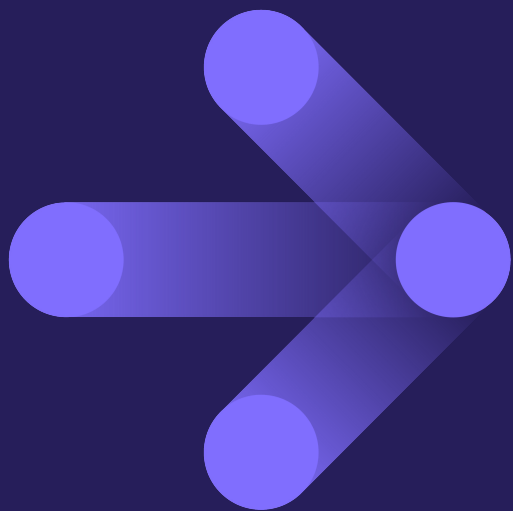
thoroughness

you need the process to be thorough in pulling all the relevant data you need for further analysis

04

useful data breakdown

you need your data to be broken down in a way that will facilitate better understanding of content quality later



Mapping important performance metrics



User engagement

- Average Session Duration
- Sessions
- Engaged Sessions
- Bounce rate
- Exit rate
- Total Users
- Active Users
- Average engagement time
- Engagement time per session
- Engagement rate
- Event interactions (custom)
- Conversions



Traffic sources

- # views/ event count from referral traffic
- # of views/ event count from social traffic
- # of views/ event count from organic traffic
- # of views/ event count from email traffic



Google Traffic data

- clicks
- impressions
- average position
- CTR
- # queries ranked per page
- queries ranked per page
- # of pages ranked
- # of pages with traffic

Repeat for
discover/news/web/image,
depending on type of content.

we can go a step further



Internal Links

- # internal links incoming
- # internal links outgoing
- # external links outgoing



Content readability

- Flesch Reading Ease Score
- Readability Score
- Word Count
- Sentence Count
- Average words per sentence



Backlinks

- # backlinks
- # follow backlinks
- # nofollow backlinks
- # text links
- # referring domains

putting this data together is a matter
of preference

Most likely



Google Sheets

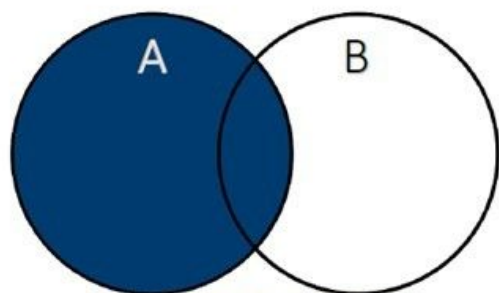


BigQuery

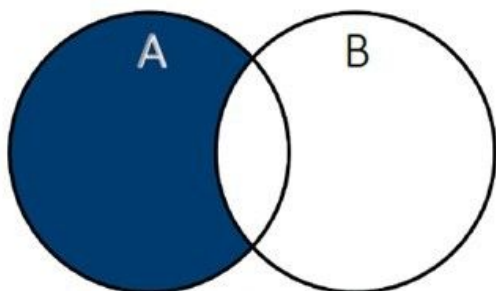
**Optional but highly recommended for
quick and pretty visualisations**



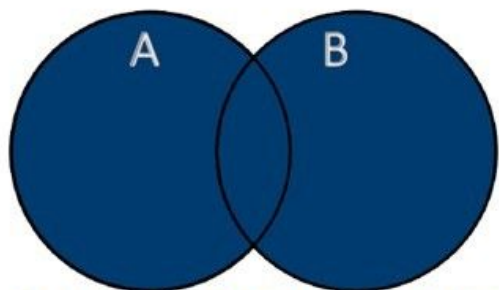
Looker Studio



LEFT INCLUSIVE

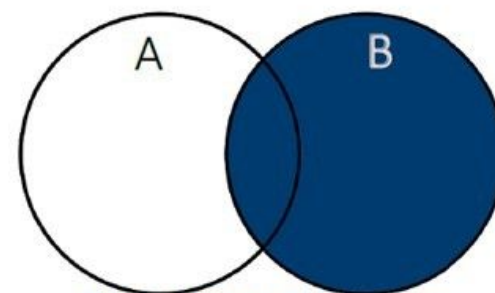


LEFT EXCLUSIVE

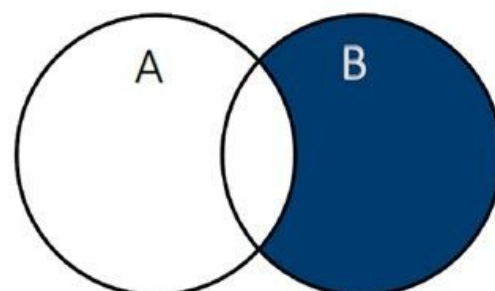


FULL OUTER INCLUSIVE

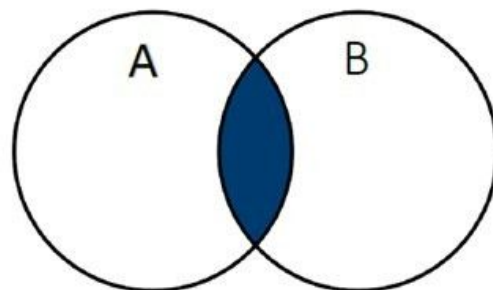
| SQL JOINS | |
|---|--|
| LEFT INCLUSIVE SELECT [Select List] FROM TableA A LEFT OUTER JOIN TableB B ON A.Key = B.Key | RIGHT INCLUSIVE SELECT [Select List] FROM TableA A RIGHT OUTER JOIN TableB B ON A.Key = B.Key |
| LEFT EXCLUSIVE SELECT [Select List] FROM TableA A LEFT OUTER JOIN TableB B ON A.Key = B.Key WHERE B.Key IS NULL | RIGHT EXCLUSIVE SELECT [Select List] FROM TableA A LEFT OUTER JOIN TableB B ON A.Key = B.Key WHERE A.Key IS NULL |
| FULL OUTER INCLUSIVE SELECT [Select List] FROM TableA A FULL OUTER JOIN TableB B ON A.Key = B.Key | FULL OUTER EXCLUSIVE SELECT [Select List] FROM TableA A FULL OUTER JOIN TableB B ON A.Key = B.Key WHERE A.Key IS NULL OR B.Key IS NULL |
| INNER JOIN SELECT [Select List] FROM TableA A INNER JOIN TableB B ON A.Key = B.Key | |



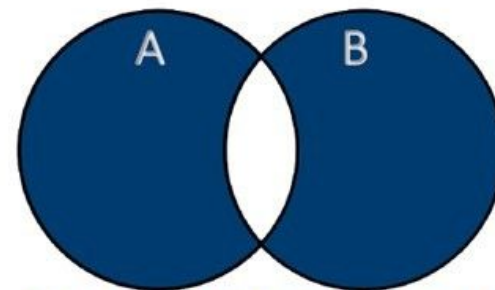
RIGHT INCLUSIVE



RIGHT EXCLUSIVE



INNER JOIN



FULL OUTER EXCLUSIVE

Organize your database – crawl, content, metrics

Understand the data – Categorize your inventory

Evaluate performance and assign action items to underperforming content

Implement & Review impact – Prioritize what and when to implement and repeat

in this stage, our aims are

01

label

label as much as possible

02

what works

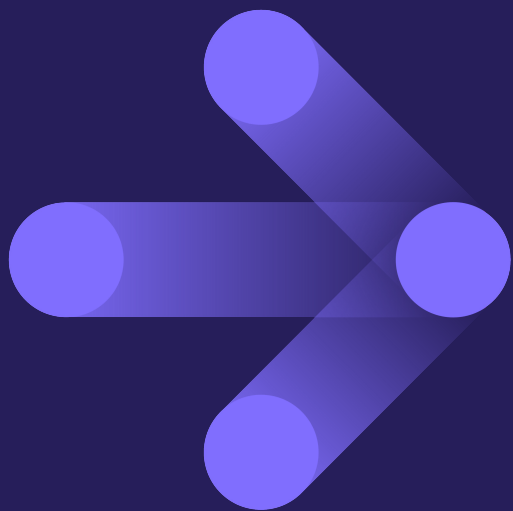
understand what works/doesn't

03

what fails

understand where the pitfalls are

let's start with the basics



site structure and authorship

| | A | B | C | D | E | F | G | H |
|----|------|-----------|-------------------|-------------------|---|----------------|-------------|------------------|
| 1 | | | Group → Ownership | | | | | |
| 2 | ID | Theme | Section | Sub-section | URL | Owner | Last Update | Update Frequency |
| 3 | 1 | home | | | https://www.distilled.net/ | Tim Allen | Jan '18 | - |
| 4 | 2 | services | | | https://www.distilled.net/services/ | Tim Allen | Nov '19 | Yearly |
| 5 | 2.1 | services | creative | | https://www.distilled.net/services/creative/ | Tim Allen | Nov '19 | Yearly |
| 6 | 3 | events | | | https://www.distilled.net/events/ | Tim Allen | Nov '19 | - |
| 7 | 3.1 | events | searchlove | london | https://www.distilled.net/events/searchlove-london/ | Tim Allen | Nov '19 | Quarterly |
| 8 | 3.2 | events | searchlove | san diego | https://www.distilled.net/events/searchlove-san-diego/ | Tim Allen | Nov '19 | Quarterly |
| 9 | 3.3 | events | searchlove | boston | https://www.distilled.net/events/searchlove-boston/ | Tim Allen | Nov '19 | Quarterly |
| 10 | 4 | resources | | | https://www.distilled.net/resources/ | Tim Allen | Nov '19 | - |
| 11 | 4.1 | resources | blog | | https://www.distilled.net/resources/blog-posts/ | Tim Allen | Nov '19 | - |
| 12 | 4.2 | resources | blog | regex in google s | https://www.distilled.net/how-to-use-regex-in-google-search/ | Paola Didone | Aug '18 | Yearly |
| 13 | 4.3 | resources | blog | google data studi | https://www.distilled.net/google-data-studio/ | Benjamin Estes | Mar '18 | Yearly |
| 14 | 4.4 | resources | blog | technical audi | https://www.distilled.net/technical-audio-seo-audio/ | Benjamin Estes | Sep '18 | Yearly |
| 15 | 4.5 | resources | guides | | https://www.distilled.net/resources/features/ | Tim Allen | Nov '19 | - |
| 16 | 4.6 | resources | guides | content guide | https://www.distilled.net/content-guide/ | Benjamin Estes | Jan '18 | Yearly |
| 17 | 4.7 | resources | guides | searchscape | https://www.distilled.net/searchscape/ | Tim Allen | Jan '18 | Yearly |
| 18 | 4.8 | resources | guides | tone of voice | https://www.distilled.net/tone-of-voice/ | Tim Allen | Jan '18 | Yearly |
| 19 | 4.9 | resources | guides | future of tv | https://www.distilled.net/future-of-tv/ | Tim Allen | Jan '18 | Yearly |
| 20 | 4.10 | resources | guides | brandopolis | https://www.distilled.net/brandopolis/ | Tim Allen | Jan '18 | Yearly |
| 21 | 4.11 | resources | guides | mobile seo guide | https://www.distilled.net/training/mobile-seo/ | Tim Allen | Jan '18 | Yearly |
| 22 | 4.12 | resources | guides | excel for seo | https://www.distilled.net/excel-for-seo/ | Tim Allen | Jan '18 | Yearly |
| 23 | 4.13 | resources | guides | linkbait guide | https://www.distilled.net/linkbait-guide/ | Tim Allen | Jan '18 | Yearly |
| 24 | 4.14 | resources | guides | video marketing | https://www.distilled.net/training/video-marketing/ | Tim Allen | Jan '18 | Yearly |
| 25 | 4.15 | resources | guides | facebook | https://www.distilled.net/training/facebook/ | Tim Allen | Jan '18 | Yearly |
| 26 | 4.16 | resources | distilledu | | https://www.distilled.net/u/ | Tim Allen | Nov '19 | - |

Theme/ Page template

| | A | B | C | D | E | F | G | H |
|----|------|-----------|------------|-------------------|---|----------------|-------------|------------------|
| 1 | | | | | Group → | Ownership | | |
| 2 | ID | Theme | Section | Sub-section | URL | Owner | Last Update | Update Frequency |
| 3 | 1 | home | | | https://www.distilled.net/ | Tim Allen | Jan '18 | - |
| 4 | 2 | services | | | https://www.distilled.net/services/ | Tim Allen | Nov '19 | Yearly |
| 5 | 2.1 | services | creative | | https://www.distilled.net/services/creative/ | Tim Allen | Nov '19 | Yearly |
| 6 | 3 | events | | | https://www.distilled.net/events/ | Tim Allen | Nov '19 | - |
| 7 | 3.1 | events | searchlove | london | https://www.distilled.net/events/searchlove-london/ | Tim Allen | Nov '19 | Quarterly |
| 8 | 3.2 | events | searchlove | san diego | https://www.distilled.net/events/searchlove-san-diego/ | Tim Allen | Nov '19 | Quarterly |
| 9 | 3.3 | events | searchlove | boston | https://www.distilled.net/events/searchlove-boston/ | Tim Allen | Nov '19 | Quarterly |
| 10 | 4 | resources | | | https://www.distilled.net/resources/ | Tim Allen | Nov '19 | - |
| 11 | 4.1 | resources | blog | | https://www.distilled.net/resources/posts/ | Tim Allen | Nov '19 | - |
| 12 | 4.2 | resources | blog | regex in google s | https://www.distilled.net/how-to-use-regex-in-google-search/ | Paola Didone | Aug '18 | Yearly |
| 13 | 4.3 | resources | blog | google data studi | https://www.distilled.net/google-data-studio/ | Benjamin Estes | Mar '18 | Yearly |
| 14 | 4.4 | resources | blog | technical audit c | https://www.distilled.net/technical-seo-audit-checklist/ | Benjamin Estes | Sep '18 | Yearly |
| 15 | 4.5 | resources | guides | | https://www.distilled.net/resources/features/ | Tim Allen | Nov '19 | - |
| 16 | 4.6 | resources | guides | content guide | https://www.distilled.net/content-guide/ | Benjamin Estes | Jan '18 | Yearly |
| 17 | 4.7 | resources | guides | searchscape | https://www.distilled.net/searchscape/ | Tim Allen | Jan '18 | Yearly |
| 18 | 4.8 | resources | guides | tone of voice | https://www.distilled.net/tone-of-voice/ | Tim Allen | Jan '18 | Yearly |
| 19 | 4.9 | resources | guides | future of tv | https://www.distilled.net/future-of-tv/ | Tim Allen | Jan '18 | Yearly |
| 20 | 4.10 | resources | guides | brandopolis | https://www.distilled.net/brandopolis/ | Tim Allen | Jan '18 | Yearly |
| 21 | 4.11 | resources | guides | mobile seo guide | https://www.distilled.net/training/mobile-seo-guide/ | Tim Allen | Jan '18 | Yearly |
| 22 | 4.12 | resources | guides | excel for seo | https://www.distilled.net/excel-for-seo/ | Tim Allen | Jan '18 | Yearly |
| 23 | 4.13 | resources | guides | linkbait guide | https://www.distilled.net/linkbait-guide/ | Tim Allen | Jan '18 | Yearly |
| 24 | 4.14 | resources | guides | video marketing | https://www.distilled.net/training/video-marketing/ | Tim Allen | Jan '18 | Yearly |
| 25 | 4.15 | resources | guides | facebook | https://www.distilled.net/training/facebook/ | Tim Allen | Jan '18 | Yearly |
| 26 | 4.16 | resources | distilledu | | https://www.distilled.net/u/ | Tim Allen | Nov '19 | - |

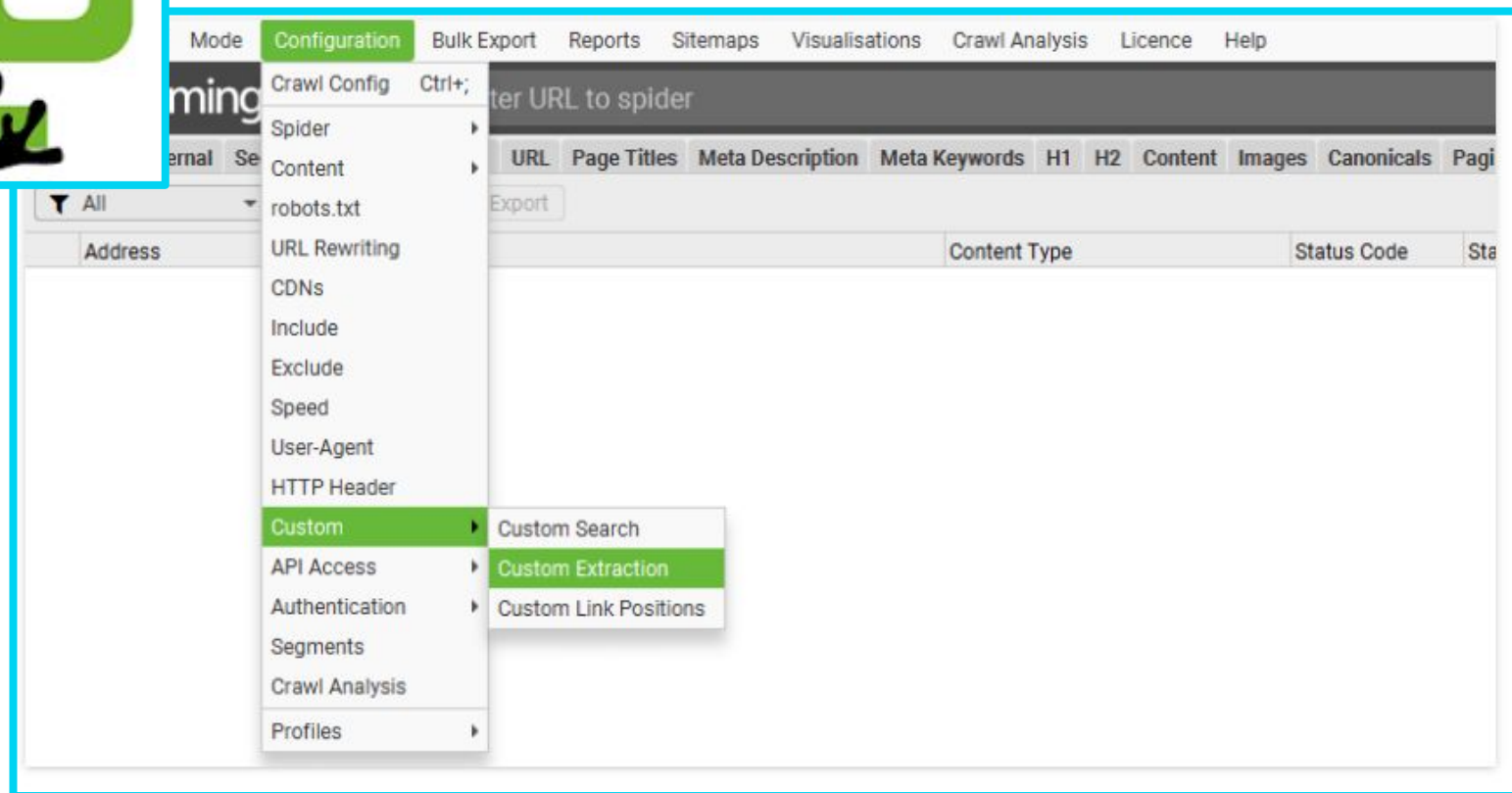
**Site Section & Subsection
(can be extracted from URL)**

| | A | B | C | D | E | F | G | H |
|----|------|-----------|------------|--------------------|---|-------------------|-------------|------------------|
| 1 | | | | | | Group → Ownership | | |
| 2 | ID | Theme | Section | Sub-section | URL | Owner | Last Update | Update Frequency |
| 3 | 1 | home | | | https://www.distilled.net/ | Tim Allen | Jan '18 | - |
| 4 | 2 | services | | | https://www.distilled.net/services/ | Tim Allen | Nov '19 | Yearly |
| 5 | 2.1 | services | creative | | https://www.distilled.net/services/creative/ | Tim Allen | Nov '19 | Yearly |
| 6 | 3 | events | | | https://www.distilled.net/events/ | Tim Allen | Nov '19 | - |
| 7 | 3.1 | events | searchlove | london | https://www.distilled.net/events/searchlove/ | Tim Allen | Nov '19 | Quarterly |
| 8 | 3.2 | events | searchlove | san diego | https://www.distilled.net/events/searchlove/ | Tim Allen | Nov '19 | Quarterly |
| 9 | 3.3 | events | searchlove | boston | https://www.distilled.net/events/searchlove/ | Tim Allen | Nov '19 | Quarterly |
| 10 | 4 | resources | | | https://www.distilled.net/resources/ | Tim Allen | Nov '19 | - |
| 11 | 4.1 | resources | blog | | https://www.distilled.net/resources/posts/ | Tim Allen | Nov '19 | - |
| 12 | 4.2 | resources | blog | regex in google s | https://www.distilled.net/how-to-use-regex/ | Paola Didone | Aug '18 | Yearly |
| 13 | 4.3 | resources | blog | google data studi | https://www.distilled.net/google-data-studio/ | Benjamin Estes | Mar '18 | Yearly |
| 14 | 4.4 | resources | blog | technical audit cl | https://www.distilled.net/technical-seo-audit/ | Benjamin Estes | Sep '18 | Yearly |
| 15 | 4.5 | resources | guides | | https://www.distilled.net/resources/technical-seo-audit/ | Tim Allen | Nov '19 | - |
| 16 | 4.6 | resources | guides | content guide | https://www.distilled.net/content-guide/ | Benjamin Estes | Jan '18 | Yearly |
| 17 | 4.7 | resources | guides | searchscape | https://www.distilled.net/searchscape/ | Tim Allen | Jan '18 | Yearly |
| 18 | 4.8 | resources | guides | tone of voice | https://www.distilled.net/tone-of-voice/ | Tim Allen | Jan '18 | Yearly |
| 19 | 4.9 | resources | guides | future of tv | https://www.distilled.net/future-of-tv/ | Tim Allen | Jan '18 | Yearly |
| 20 | 4.10 | resources | guides | brandopolis | https://www.distilled.net/brandopolis/ | Tim Allen | Jan '18 | Yearly |
| 21 | 4.11 | resources | guides | mobile seo guide | https://www.distilled.net/training/mobile-seo-guide/ | Tim Allen | Jan '18 | Yearly |
| 22 | 4.12 | resources | guides | excel for seo | https://www.distilled.net/excel-for-seo/ | Tim Allen | Jan '18 | Yearly |
| 23 | 4.13 | resources | guides | linkbait guide | https://www.distilled.net/linkbait-guide/ | Tim Allen | Jan '18 | Yearly |
| 24 | 4.14 | resources | guides | video marketing | https://www.distilled.net/training/video-marketing/ | Tim Allen | Jan '18 | Yearly |
| 25 | 4.15 | resources | guides | facebook | https://www.distilled.net/training/facebook/ | Tim Allen | Jan '18 | Yearly |
| 26 | 4.16 | resources | distilledu | | https://www.distilled.net/u/ | Tim Allen | Nov '19 | - |

Author

| | A | B | C | D | E | F | G | H |
|----|------|-----------|------------|--------------------|---|----------------|-------------|------------------|
| 1 | | Group → | | | | Ownership | | |
| 2 | ID | Theme | Section | Sub-section | URL | Owner | Last Update | Update Frequency |
| 3 | 1 | home | | | https://www.distilled.net/ | Tim Allen | Jan '18 | - |
| 4 | 2 | services | | | https://www.distilled.net/services/ | Tim Allen | Nov '19 | Yearly |
| 5 | 2.1 | services | creative | | https://www.distilled.net/services/creative/ | Tim Allen | Nov '19 | Yearly |
| 6 | 3 | events | | | https://www.distilled.net/events/ | Tim Allen | Nov '19 | - |
| 7 | 3.1 | events | searchlove | london | https://www.distilled.net/events/searchlove-london/ | Tim Allen | Nov '19 | Quarterly |
| 8 | 3.2 | events | searchlove | san diego | https://www.distilled.net/events/searchlove-san-diego/ | Tim Allen | Nov '19 | Quarterly |
| 9 | 3.3 | events | searchlove | boston | https://www.distilled.net/events/searchlove-boston/ | Tim Allen | Nov '19 | Quarterly |
| 10 | 4 | resources | | | https://www.distilled.net/resources/ | Tim Allen | Nov '19 | - |
| 11 | 4.1 | resources | blog | | https://www.distilled.net/resources/posts/ | Tim Allen | Nov '19 | - |
| 12 | 4.2 | resources | blog | regex in google s | https://www.distilled.net/how-to-use-regex-in-google-search/ | Paola Didone | Aug '18 | Yearly |
| 13 | 4.3 | resources | blog | google data studi | https://www.distilled.net/google-data-studio/ | Benjamin Estes | Mar '18 | Yearly |
| 14 | 4.4 | resources | blog | technical audit cl | https://www.distilled.net/technical-seo-audit-checklist/ | Benjamin Estes | Sep '18 | Yearly |
| 15 | 4.5 | resources | guides | | https://www.distilled.net/resources/features/ | Tim Allen | Nov '19 | - |
| 16 | 4.6 | resources | guides | content guide | https://www.distilled.net/content-guide/ | Benjamin Estes | Jan '18 | Yearly |
| 17 | 4.7 | resources | guides | searchscape | https://www.distilled.net/searchscape/ | Tim Allen | Jan '18 | Yearly |
| 18 | 4.8 | resources | guides | tone of voice | https://www.distilled.net/tone-of-voice/ | Tim Allen | Jan '18 | Yearly |
| 19 | 4.9 | resources | guides | future of tv | https://www.distilled.net/future-of-tv/ | Tim Allen | Jan '18 | Yearly |
| 20 | 4.10 | resources | guides | brandopolis | https://www.distilled.net/brandopolis/ | Tim Allen | Jan '18 | Yearly |
| 21 | 4.11 | resources | guides | mobile seo guide | https://www.distilled.net/training/mobile-seo-guide/ | Tim Allen | Jan '18 | Yearly |
| 22 | 4.12 | resources | guides | excel for seo | https://www.distilled.net/excel-for-seo/ | Tim Allen | Jan '18 | Yearly |
| 23 | 4.13 | resources | guides | linkbait guide | https://www.distilled.net/linkbait-guide/ | Tim Allen | Jan '18 | Yearly |
| 24 | 4.14 | resources | guides | video marketing | https://www.distilled.net/training/video-marketing/ | Tim Allen | Jan '18 | Yearly |
| 25 | 4.15 | resources | guides | facebook | https://www.distilled.net/training/facebook/ | Tim Allen | Jan '18 | Yearly |
| 26 | 4.16 | resources | distilledu | | https://www.distilled.net/u/ | Tim Allen | Nov '19 | - |

Publish date,
Last updated date,
Update frequency



AutoML

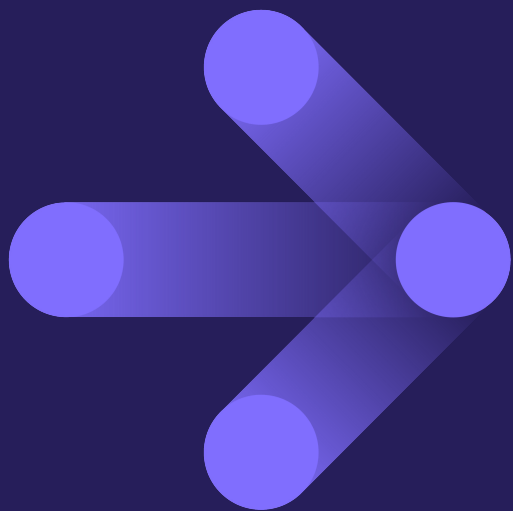
```
[ ] from google.colab.patches import cv2_imshow  
cv2_imshow(img_with_boxes)
```

The screenshot shows an Amazon product page for a toilet. Several areas are highlighted with bounding boxes and labeled with confidence scores:

- main product image**: 90.14%
- product name**: 94.47%
- product description**: 98.80%
- additional product images**: 99.40%

The product details visible on the page include:

- Product Name**: 11-0 Santa Rosa Comfort Elongated 1.6 GPF Toilet with On Flush Technology and Left-Hand Trip Lever, White
- Brand**: Kohler
- Ratings**: 24 ratings
- Available from these sellers**:
 - none
 - Imported
 - One-piece toilets integrate the tank and bowl into a seamless, easy-to-clean design
 - Comfort Height offers chair-height seating that makes sitting down and standing up easier
 - Compact elongated bowl offers added comfort while occupying the same space as a round-front bowl
 - AquaPiston canister allows water to flow into the bowl from all sides, creating a more powerful & effective flush
 - Durable canister design has 90% less exposed seal material than a 3-inch flapper, for leak-free performance



**target intent and
update needs**

Use rule-based classification to map out
target intent, personas, content freshness

The screenshot shows a Google Sheet with a formula in cell D2 and a dropdown menu for 'Target intent'.

Formula in D2:

```
=ifna(ifs(REGEXMATCH(lower(B2), "how|why|does|which|when|who|guide|tutorial|learn|examples|resource|ideas|tips|difference  
between|strategies|ways|example|directory|news|basics|concepts|troubleshoot|resolve|risks|beginners|upgrade|improve|optimize|case study|best practices|templates|pros and  
cons|analysis|articles|audit|benefits|blogs|build|calculate|campaign|choose|common questions|consider|cost|create|definition|definitive guide|learn|easy steps|essential|events|explainer  
video|factors|find|fix|grow|hacks|how to|ideal|importance of|increase|infographic|information|meaning|measure|metrics|mistakes|myths|opportunities|popular|practices|questions|remove|set  
up|shortcuts|stats|strategy|study|successful|tactics|techniques|terms|things to know|time|trends|tricks|types|updates|well  
explained|tutorials|framework|checklist|process|plan|test|glossary|sample|survey|exercises|report|certification|webinar|calendar|quiz|directories|to do list|design basics"), "Informational",  
REGEXMATCH(lower(B2), "buy|purchase|order|price|quote|for sale|sale|deal|offers|discount|coupon|promo code|voucher|bargain|cost|pricing|cheap|affordable|expensive|premium|near me|online|e-  
commerce|shopping cart|checkout|shop|store|retailer|dealer|supplier|marketplace|auction|bid|buy now|pay|payment|finance|leasing|rent|rental|subscribe|subscription|free  
trial|book|reserve|enquire|inquiry|quote|estimate|cost calculator|shipping|delivery|available|in stock|out of stock|pre-order|backorder|special offer|limited time offer|flash  
sale|clearance|wholesale|bulk buy|group buy|membership|reward|loyalty|gift card|cashback|trade-in|upgrade|renew|checkout|secure checkout"),  
"Transactional", REGEXMATCH(lower(B2), "best|top|vs|review|compare|comparison|ratings|recommendation|buyer's guide|alternative*|feature*|benefit*|cost-  
effective|affordable|premium|quality|deal|offers|savings|discount|value|choice|selection|rankings|choice|user
```

Target intent dropdown menu:

| Target intent |
|---------------|
| Unclassified |
| Informational |
| Commercial |
| Unclassified |
| Unclassified |
| Unclassified |
| Unclassified |
| Informational |
| Informational |
| Informational |
| Informational |
| Informational |
| Unclassified |
| Unclassified |

Create custom target intent categories for titles

even better ✨
classify ranked queries (GSC)
with custom intent filters
and assign the category based on that

F2

```
=IF(AND(ITS(REGEXMATCH(A2, "how|why|does|which|when|who|guide|tutorial|learn|examples|resource|ideas|tips|difference  
between|strategies|ways|example|directory|news|basics|concepts|troubleshoot|resolve|risks|beginners|upgrade|improve|optimize|case study|best  
practices|templates|pros and cons|analysis|articles|audit|benefits|blogs|build|calculate|campaign|choose|common  
questions|consider|cost|create|definition|definitive guide|earn|easy steps|essential|events|explainer video|factors|find|fix|grow|hacks|how  
to|ideal|importance  
of|increase|infographic|information|meaning|measure|metrics|mistakes|myths|opportunities|popular|practices|questions|remove|set  
up|shortcuts|stats|strategy|study|successful|tactics|techniques|terms|things to know|time|trends|tricks|types|updates|well  
explained|tutorials|framework|checklist|process|plan|test|glossary|sample|survey|exercises|report|certification|webinar|calendar|quiz|directo  
ries|to do list|design basics  
"), "Informational", REGEXMATCH(A2, "buy|purchase|order|price|quote|for sale|sale|deal|offers|discount|coupon|promo  
code|voucher|bargain|cost|pricing|cheap|affordable|expensive|premium|near me|online|e-|shopping  
cart|checkout|shop|store|retailer|dealer|supplier|marketplace|auction|bid|buy  
now|pay|payment|finance|leasing|rent|rental|subscribe|subscription|free trial|book|reserve|inquire|inquiry|quote|estimate|cost
```

A

B

C

D

E

F

1

Query

Landing Page

Average Position

Url Clicks

Impressions

Search Intent

2

pytrends

https://lazarinastoy.com/the-ultimate-guide-to

3.983786914

120

1727

Informational

3

lazarina stoy

https://lazarinastoy.com/the-ultimate-guide-to

10.12707499

22

1747

Unclassified

4

google trends api

https://lazarinastoy.com/the-ultimate-guide-to

2.470588235

16

51

Informational

5

pytrends example

https://lazarinastoy.com/the-ultimate-guide-to

4.164021164

189

32

Unclassified

6

pytrend

https://lazarinastoy.com/the-ultimate-guide-to

5.099337748

12

302

Informational

7

google trends api python

https://lazarinastoy.com/the-ultimate-guide-to

3.983786914

8

32

Unclassified

8

looker studio custom dimension

https://lazarinastoy.com/the-ultimate-guide-to

2.941176471

7

34

Informational

9

seo google sheets templates

https://lazarinastoy.com/resources/google-sh

+

≡

Import your

Screaming Frog crawl here

Content Audit Main Sheet

Query

Customize the classification
based on your own website, and
what makes sense

even better ✨
train a classifier
on your niche's content and queries

Prepare text training data for classification

[Send feedback](#)

This page describes how to prepare text data for use in a Vertex AI dataset to train single-label and multi-label classification models.

[Single-label classificati...](#)[Multi-label classification](#)

For single-label classification, training data consists of documents and the classification category that apply to those documents. Single-label classification allows a document to be assigned only one label.

Data requirements

- You must supply at least 20, and no more than 1,000,000, training documents.
- You must supply at least 2, and no more than 5000, unique category labels.
- You must apply each label to at least 10 documents.
- For multi-label classification, you can apply one or multiple labels to a document.
- You can include documents inline or reference TXT files that are in Cloud Storage buckets.

Best practices for text data used to train AutoML models

The following recommendations apply to datasets used to train AutoML models.

- Use training data that is as varied as the data on which predictions will be made. Include different lengths of documents, documents authored by different people, documents that use different wording or style, and so on.

Hello text data: Create a text classification dataset and import documents

[Send feedback](#)

To learn more, run the "Create, train, and deploy an AutoML text classification model" Jupyter notebook in one of the following environments:

[Open in Colab](#) | [Open in Vertex AI Workbench user-managed notebooks](#) | [View on GitHub](#)

Use the Vertex AI console to create a text classification dataset. After your dataset is created, use the CSV that you copied into your Cloud Storage bucket to import those documents into the dataset.

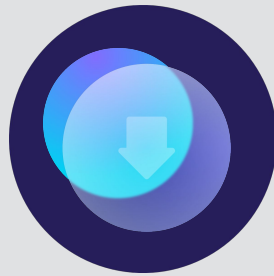
This tutorial has several pages:

1. [Setting up your project and environment.](#)
2. [Creating a text classification dataset .](#)
3. [Training an AutoML text classification model.](#)
4. [Deploy model to an endpoint and send a prediction.](#)
5. [Cleaning up your project.](#)

Each page assumes that you have already performed the instructions from the previous pages of the tutorial.

1. Go to the [Vertex AI console](#) .
2. From the **Get started with Vertex AI** page, click **Create dataset**.
3. Specify details about your dataset.
 - a. Specify a name for this dataset, such as `text_classification_tutorial`.

experiment and improve



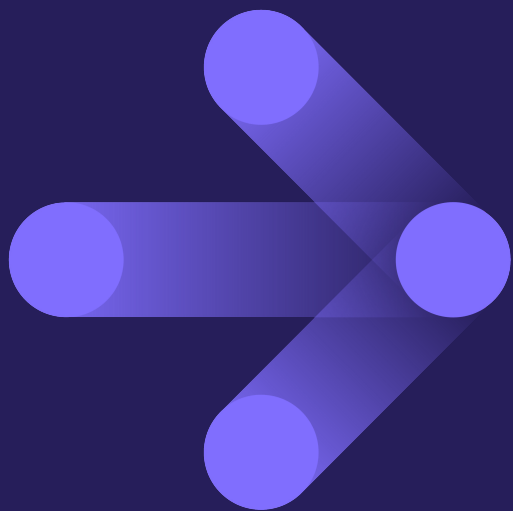
incorporating signals
from SERP features
ranked on Google



go deeper than
informational,
navigational, etc. and
pinpoint actions



decipher the user
knowledge level from
queries ranked, and
whether your content
matches that

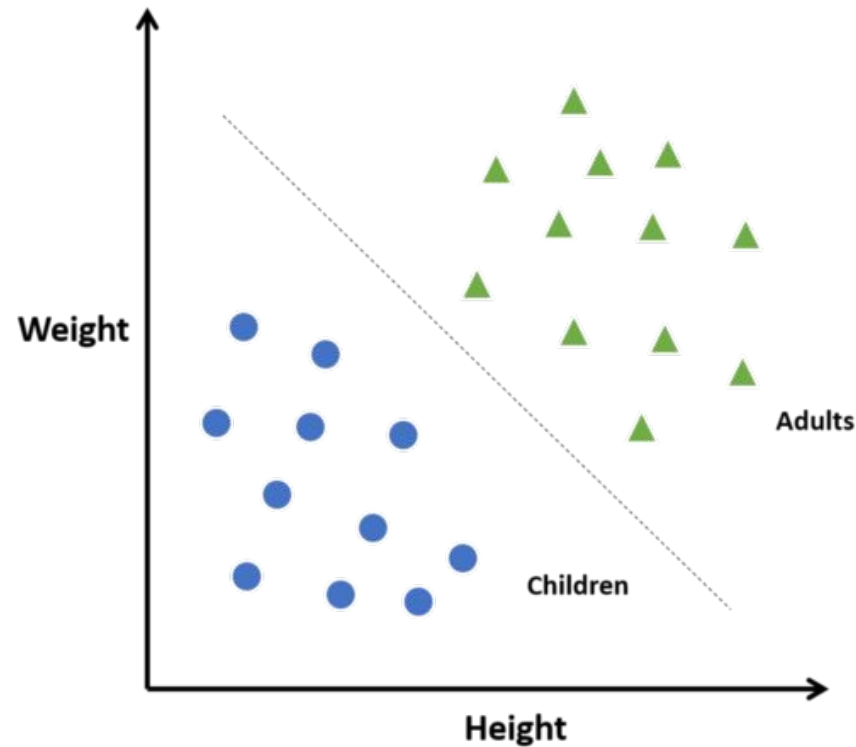


**understand content
topics and subtopics**

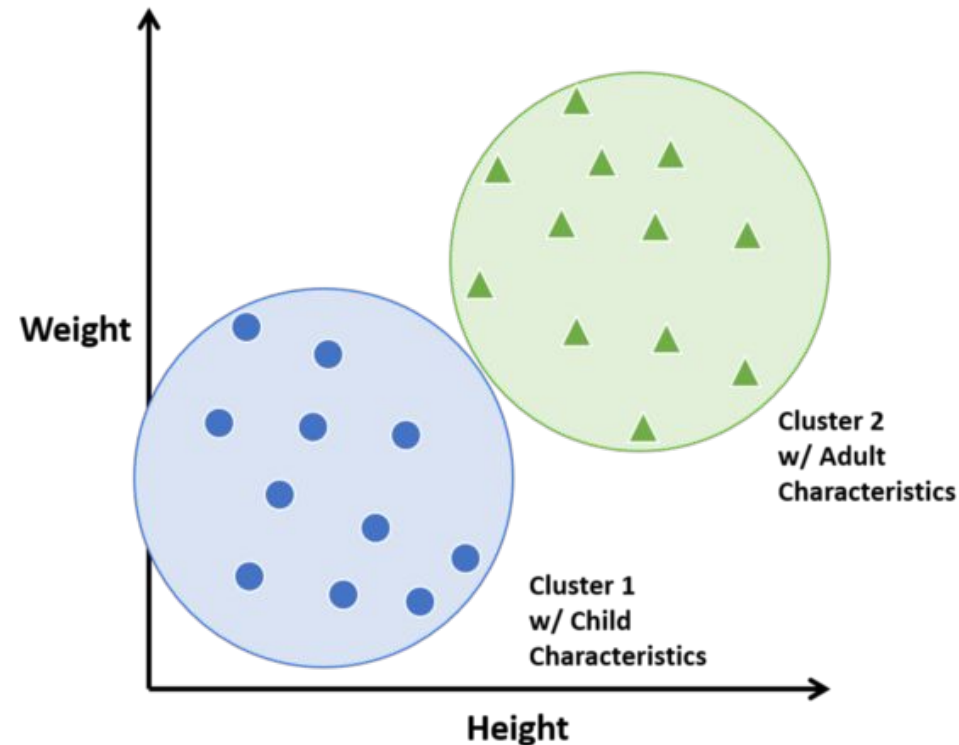
at a minimum, we should understand the topics
the content covers

roughly, there are two ways of doing this

Classification sorts data into specific categories **using a labeled dataset.**

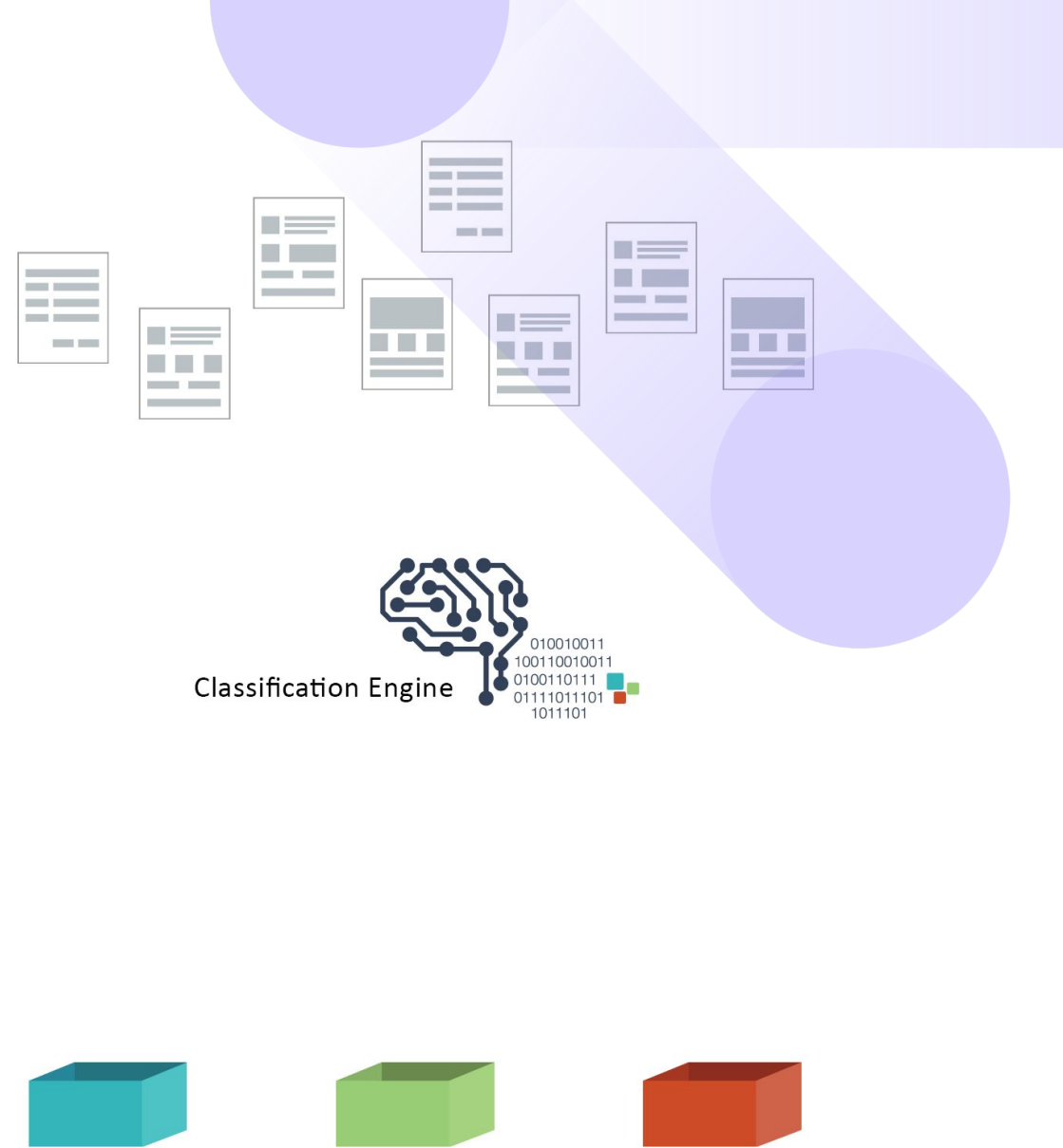


Clustering is partitioning an **unlabeled dataset into groups** of similar objects.





With Google's Natural Language API, you can classify documents in **1.3K+** predefined categories





Text Classification

How to do Text Classification with Google's Natural Language API in Google Sheets (Apps Script)

Lazarina Stoy. · Mar 27, 2024

Process will take no more than 20 minutes





100%

£

%

.0

.00

123

Defaul...

-

10

+

B

I



A



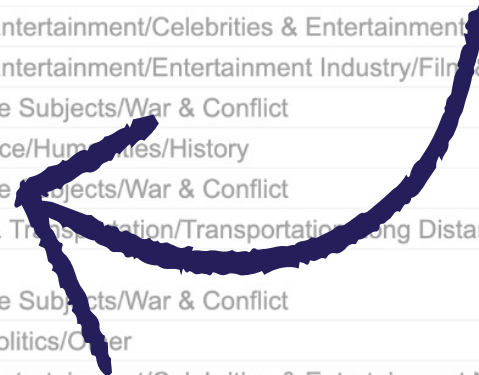
C2

/Sensitive Subjects/War & Conflict



| A | | B | | C | | D | |
|----|---|--|--|---|------------|---|--|
| 1 | URL | Content | | Classification Label | Confidence | | |
| 2 | https://www.bbc.com/blog/world-europe-68255490 | Exhausted Ukraine struggles to find new men for front | | /Sensitive Subjects/War & Conflict | 0.9567586 | | |
| 3 | https://www.bbc.com/news/world-us-canada-68395414 | South Carolina primary: Donald Trump easily defeats | | /News/Politics/Campaigns & Elections | 0.97098255 | | |
| 4 | https://www.bbc.com/news/entertainment-arts-68362811 | Kim Petras on sexual liberation and fighting TikTokPub | | /Arts & Entertainment/Celebrities & Entertainment News | 0.73509616 | | |
| 5 | https://www.bbc.com/news/entertainment-arts-68395354 | SAG Awards red carpet 2024: From Margot Robbie to | | /Arts & Entertainment/Celebrities & Entertainment News | 0.96840936 | | |
| 6 | https://www.bbc.com/news/entertainment-arts-68395355 | SAG Awards 2024: Oppenheimer dominates ahead of | | /Arts & Entertainment/Entertainment Industry/Film & TV Industry | 1 | | |
| 7 | https://www.bbc.com/news/world-middle-east-68395173 | US and UK carry out fresh strikes on Houthi targets in | | /Sensitive Subjects/War & Conflict | 1 | | |
| 8 | https://www.bbc.com/news/uk-scotland-glasgow-west-67 | Inside the long-abandoned tunnel beneath the ClydeP | | /Reference/Humanities/History | 0.46487474 | | |
| 9 | https://www.bbc.com/news/world-europe-68322527 | Ukraine war: Is Avdiivka's fall a sign Russia is turning | | /Sensitive Subjects/War & Conflict | 0.97672516 | | |
| 10 | https://www.bbc.com/news/uk-wales-68210255 | Travel: How a £525 bet gave birth to your morning con | | /Travel & Transportation/Transportation/Long Distance Bus & R | 0.824742 | | |
| 11 | https://www.bbc.com/news/world-europe-68384341 | Two years into Russia's invasion, exhausted Ukrainians | | /Sensitive Subjects/War & Conflict | 0.96695495 | | |
| 12 | https://www.bbc.com/news/world-europe-68393412 | Authorities return body of Alexei Navalny to mother 8 c | | /News/Politics/Other | 0.8510177 | | |
| 13 | https://www.bbc.com/news/entertainment-arts-68391330 | Wendy Williams thanks fans for support after dementia | | /Arts & Entertainment/Celebrities & Entertainment News | 0.9236957 | | |
| 14 | https://www.bbc.com/news/world-asia-68378651 | Japan naked festival: Women join Hadaka Matsuri for | | /People & Society/Religion & Belief | 0.9582382 | | |
| 15 | https://www.bbc.com/news/world-europe-68395030 | Alexei Navalny: Dissent is dangerous in Russia, but ac | | /News/Politics/Other | 1 | | |
| 16 | https://www.bbc.com/news/world-europe-68359252 | Rosenberg: How two years of war in Ukraine changed | | /Sensitive Subjects/War & Conflict | 0.95304227 | | |
| 17 | https://www.bbc.com/news/entertainment-arts-68395352 | SAG Award winners 2024: The full list of nominees and | | /Arts & Entertainment/Entertainment Industry/Film & TV Industry | 1 | | |
| 18 | https://www.bbc.com/news/entertainment-arts-68362811 | Stray Kids: How K-Pop took over the global charts in 2 | | /Arts & Entertainment/Music & Audio/World Music | 0.9936768 | | |
| 19 | https://www.bbc.com/news/entertainment-arts-68317730 | Gareth Edwards: The Creator director on shaking up H | | /Arts & Entertainment/Movies/Science Fiction & Fantasy Films | 0.9095848 | | |
| 20 | https://www.bbc.com/news/newsbeat-68382142 | Chuckie: 1Xtra presenter feels R&B has special year a | | /Arts & Entertainment/Music & Audio/Urban & Hip-Hop | 0.8976116 | | |

Enter your URLs and content





100%

£

%

.0

.00

123

Defaul...

-

10

+

B

I



A



C2

/Sensitive Subjects/War & Conflict

Run the script via a formula to get
classification label & confidence

| | A | B |
|----|---|--|
| 1 | URL | Content |
| 2 | https://www.bbc.com/news/world-europe-68254466 | Exhausted Ukraine struggles to find new men for front |
| 3 | https://www.bbc.com/news/world-us-canada-68395414 | South Carolina primary: Donald Trump easily defeats N |
| 4 | https://www.bbc.com/news/entertainment-arts-68362811 | Kim Petras on sexual liberation and fighting TikTokPul |
| 5 | https://www.bbc.com/news/entertainment-arts-68395354 | SAG Awards red carpet 2024: From Margot Robbie to B |
| 6 | https://www.bbc.com/news/entertainment-arts-68395355 | SAG Awards 2024: Oppenheimer dominates ahead of o |
| 7 | https://www.bbc.com/news/world-middle-east-68395173 | US and UK carry out fresh strikes on Houthis targets in |
| 8 | https://www.bbc.com/news/uk-scotland-glasgow-west-67 | Inside the long-abandoned tunnel beneath the WhitePa |
| 9 | https://www.bbc.com/news/world-europe-68322527 | Ukraine war: Is Avduka's fall a sign Russia is losing th |
| 10 | https://www.bbc.com/news/uk-wales-68210255 | Travel: How a £525 bet gave me a year's worth of mor |
| 11 | https://www.bbc.com/news/world-europe-68384341 | Two years into Russia's invasion, exhausted Ukrainians |
| 12 | https://www.bbc.com/news/world-europe-68393412 | Anthill resident "I'm no politician," confesses. Galeri |
| 13 | https://www.bbc.com/news/entertainment-arts-68391330 | Authorities return body of Alexei Navalny to mother 8 |
| 14 | https://www.bbc.com/news/world-asia-68378651 | Wendy Williams thanks fans for support after dementia |
| 15 | https://www.bbc.com/news/world-europe-68395030 | Japan naked festival: Women join Hadaka Matsuri for |
| 16 | https://www.bbc.com/news/world-europe-68359252 | Alexei Navalny: Dissent is dangerous in Russia, but ac |
| 17 | https://www.bbc.com/news/world-europe-68359252 | Rosenberg: How two years of war in Ukraine changed |
| 18 | https://www.bbc.com/news/entertainment-arts-68395352 | SAG Award winners 2024: The full list of nominees and |
| 19 | https://www.bbc.com/news/entertainment-arts-68362811 | Stray Kids: How K-Pop took over the global charts in 2 |
| 20 | https://www.bbc.com/news/entertainment-arts-68317736 | Gareth Edwards: The Creator director on shaking up M |
| | https://www.bbc.com/news/newsbeat-68382142 | Chuckie: 1Xtra presenter feels R&B has special year |

| C | D |
|---|------------|
| Classification Label | Confidence |
| /Sensitive Subjects/War & Conflict | 0.9567586 |
| /News/Politics/Campaigns & Elections | 0.97098255 |
| /Arts & Entertainment/Celebrities & Entertainment News | 0.73509616 |
| /Arts & Entertainment/Celebrities & Entertainment News | 0.96840936 |
| /Arts & Entertainment/Entertainment Industry/Film & TV Industry | 1 |
| /Sensitive Subjects/War & Conflict | 1 |
| /Reference/Humanities/History | 0.46487474 |
| /Sensitive Subjects/War & Conflict | 0.97672516 |
| /Travel & Transportation/Transportation/Long Distance Bus & R | 0.824742 |
| /Sensitive Subjects/War & Conflict | 0.96695495 |
| /News/Politics/Other | 0.8510177 |
| /Arts & Entertainment/Celebrities & Entertainment News | 0.9236957 |
| /People & Society/Religion & Belief | 0.9582382 |
| /News/Politics/Other | 1 |
| /Sensitive Subjects/War & Conflict | 0.95304227 |
| /Arts & Entertainment/Entertainment Industry/Film & TV Industry | 1 |
| /Arts & Entertainment/Music & Audio/World Music | 0.9936768 |
| /Arts & Entertainment/Movies/Science Fiction & Fantasy Films | 0.9095848 |
| /Arts & Entertainment/Music & Audio/Urban & Hip-Hop | 0.8976116 |

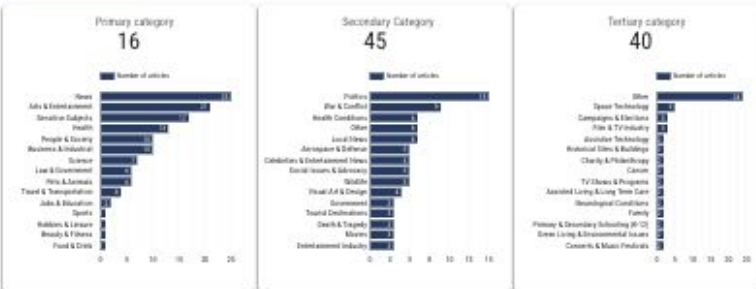


Simple Natural Language API Text Classification Visualization by [Lorenzini et al.](#) for [MFORSEO](#)
Read the [associated tutorial](#) and [grab your Google Sheets template](#)

at a glance

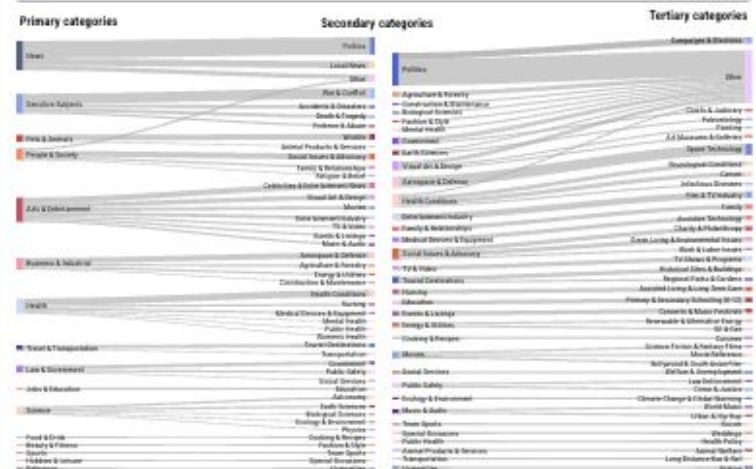
Executive Summary

A total of 12 items are contained in the selection. There are 16 unique primary categories, 45 unique secondary categories, 40 unique tertiary categories, and 2 unique quaternary categories.



playground

Click on any of the primary, secondary, or tertiary categories to filter the charts on the page, and understand the structure of content categories, and related categories.



Use the filters to identify content groups per URL, or pages that contain a certain keyword. You can also filter the page per classification label, using REGEX to view multiple content groups, or filter out low confidence categories.

| URL | Classification Label | Confidence | Primary category | Secondary category | Tertiary category | Quaternary category |
|---|--|------------|----------------------|---------------------|-----------------------|---------------------|
| https://www.bbc.com/news/world-us-canada-60795623 | News/Politics/Campaigns & Elections | 91% | News | Politics | Campaigns & Elections | null |
| https://www.bbc.com/news/world-us-canada-60795623 | News/Politics/Other | 89% | News | Politics | Other | null |
| https://www.bbc.com/news/world-us-canada-60795623 | News/Politics/Campaigns & Elections | 86% | News | Politics | Campaigns & Elections | null |
| https://www.bbc.com/news/world-middle-east-60795623 | Education Subjects/War & Conflict | 100% | Education Subjects | War & Conflict | null | null |
| https://www.bbc.com/news/world-middle-east-60795623 | Arts & Entertainment/Visual Art & Design | 92% | Arts & Entertainment | Visual Art & Design | Painting | null |
| https://www.bbc.com/news/world-middle-east-60795623 | News/Politics/Other | 100% | News | Politics | Other | null |
| https://www.bbc.com/news/world-middle-east-60795623 | News/Politics/Other | 89% | News | Politics | Other | null |
| https://www.bbc.com/news/world-middle-east-60795623 | Education Subjects/War & Conflict | 91% | Education Subjects | War & Conflict | null | null |
| https://www.bbc.com/news/world-middle-east-60795623 | Education Subjects/War & Conflict | 89% | Education Subjects | War & Conflict | null | null |
| https://www.bbc.com/news/world-middle-east-60795623 | Education Subjects/War & Conflict | 86% | Education Subjects | War & Conflict | null | null |

Plug-and-play template in Looker Studio



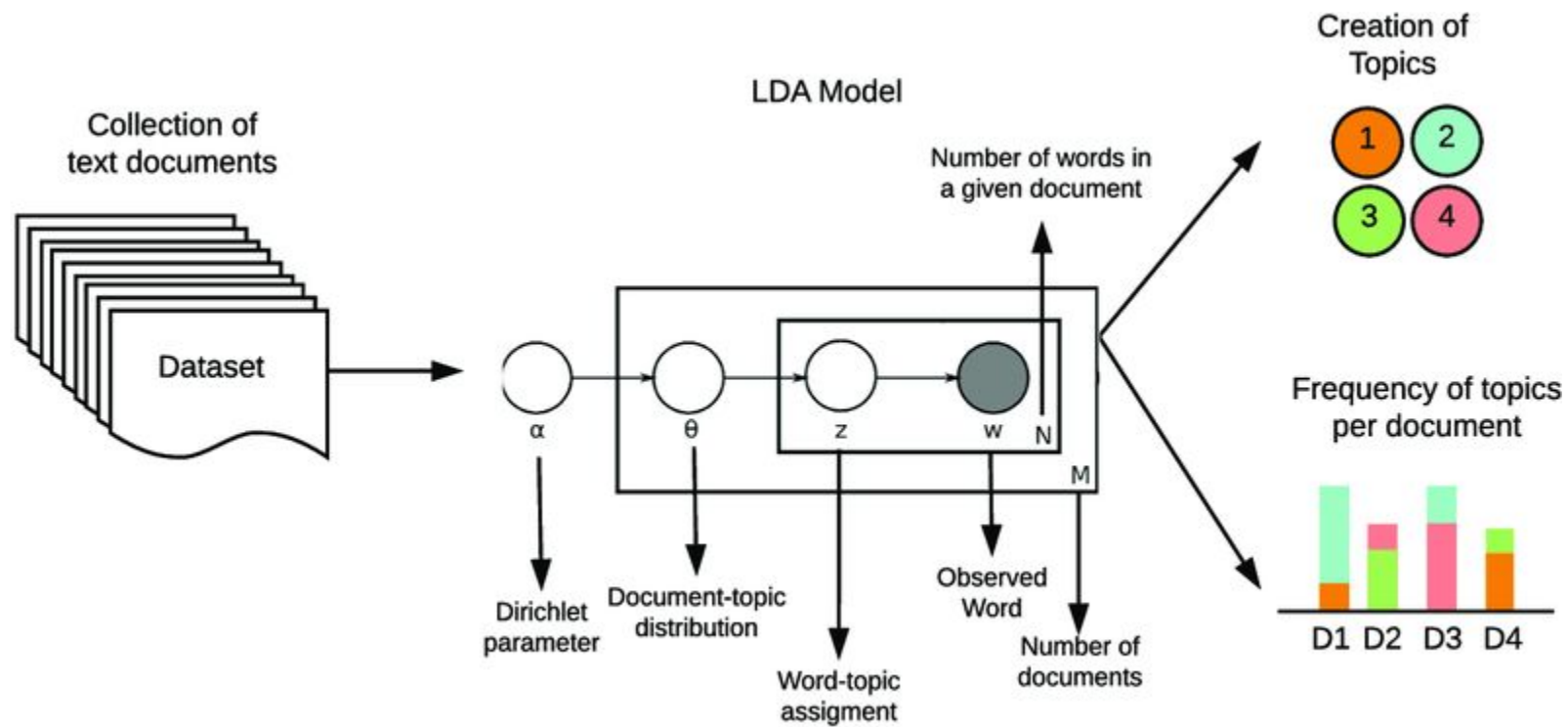




- ✓ Predictable categories
- ✓ Controlled training of model
- ✓ Accuracy indicated
- ✓ Great for scale and benchmarking
- ✓ Integration with LLMs (***new**)
- ✗ **Can't be used for uses outside of the main task (i.e. no clustering)**
- ✗ Can't be given custom lists
- ✗ Requires time and data for custom training models with AutoML

now, onto subtopic identification and tagging

enter LDA



Topics

| | |
|---------|------|
| gene | 0.04 |
| dna | 0.02 |
| genetic | 0.01 |
| ... | |

| | |
|----------|------|
| life | 0.02 |
| evolve | 0.01 |
| organism | 0.01 |
| ... | |

| | |
|--------|------|
| brain | 0.04 |
| neuron | 0.02 |
| nerve | 0.01 |
| ... | |

| | |
|----------|------|
| data | 0.02 |
| number | 0.02 |
| computer | 0.01 |
| ... | |

Documents

Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK—How many **genes** does an **organism** need to **survive**? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for **life**. One research team, using **computer** analyses to compare known **genomes**, concluded that today's **organisms** can be sustained with just 250 genes, and that the earliest life forms required a mere 128 **genes**. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, those **predictions**

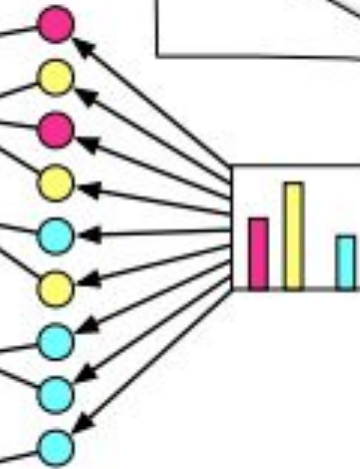
"are not all that far apart," especially in comparison to the 75,000 **genes** in the human genome, notes Siv Anderson, a geneticist at Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a **genetic** numbers game, particularly as more and more **genomes** are completely mapped and sequenced. "It may be a way of organizing any newly **sequenced genome**," explains Arcady Mushegian, a **computational** molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an

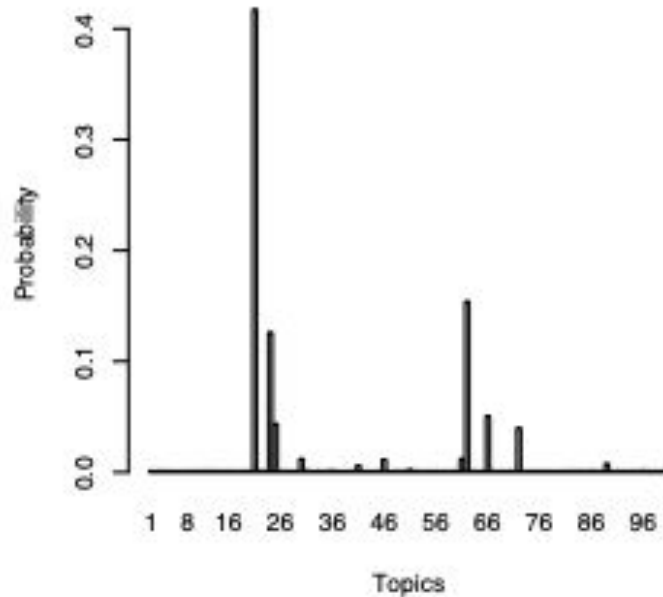
* Genome Mapping and Sequencing. Cold Spring Harbor, New York, May 8 to 12.

Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

SCIENCE • VOL. 272 • 24 MAY 1996

Topic proportions and assignments





“Genetics”

human
genome
dna
genetic
genes
sequence
gene
molecular
sequencing
map
information
genetics
mapping
project
sequences

“Evolution”

evolution
evolutionary
species
organisms
life
origin
biology
groups
phylogenetic
living
diversity
group
new
two
common

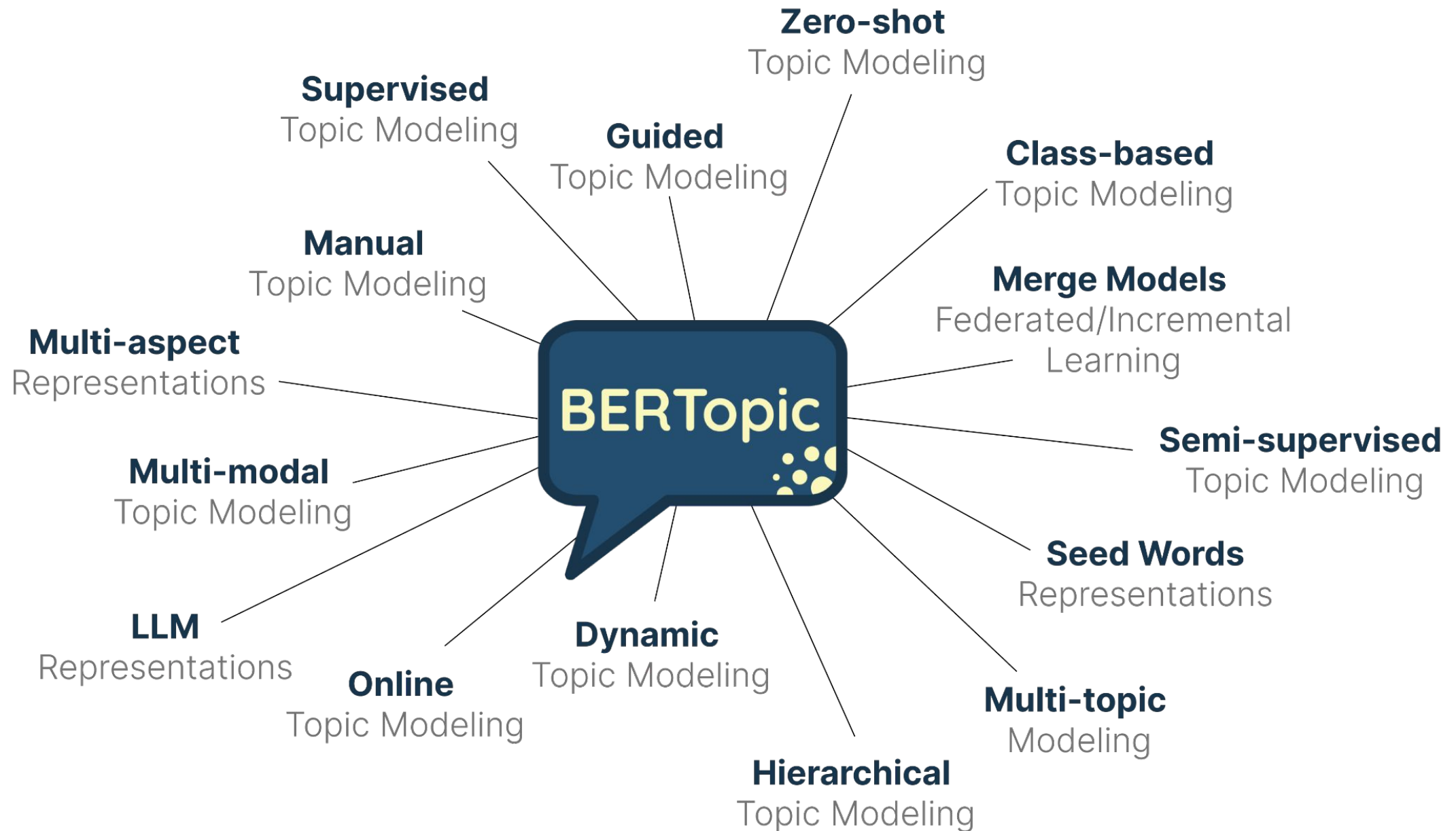
“Disease”

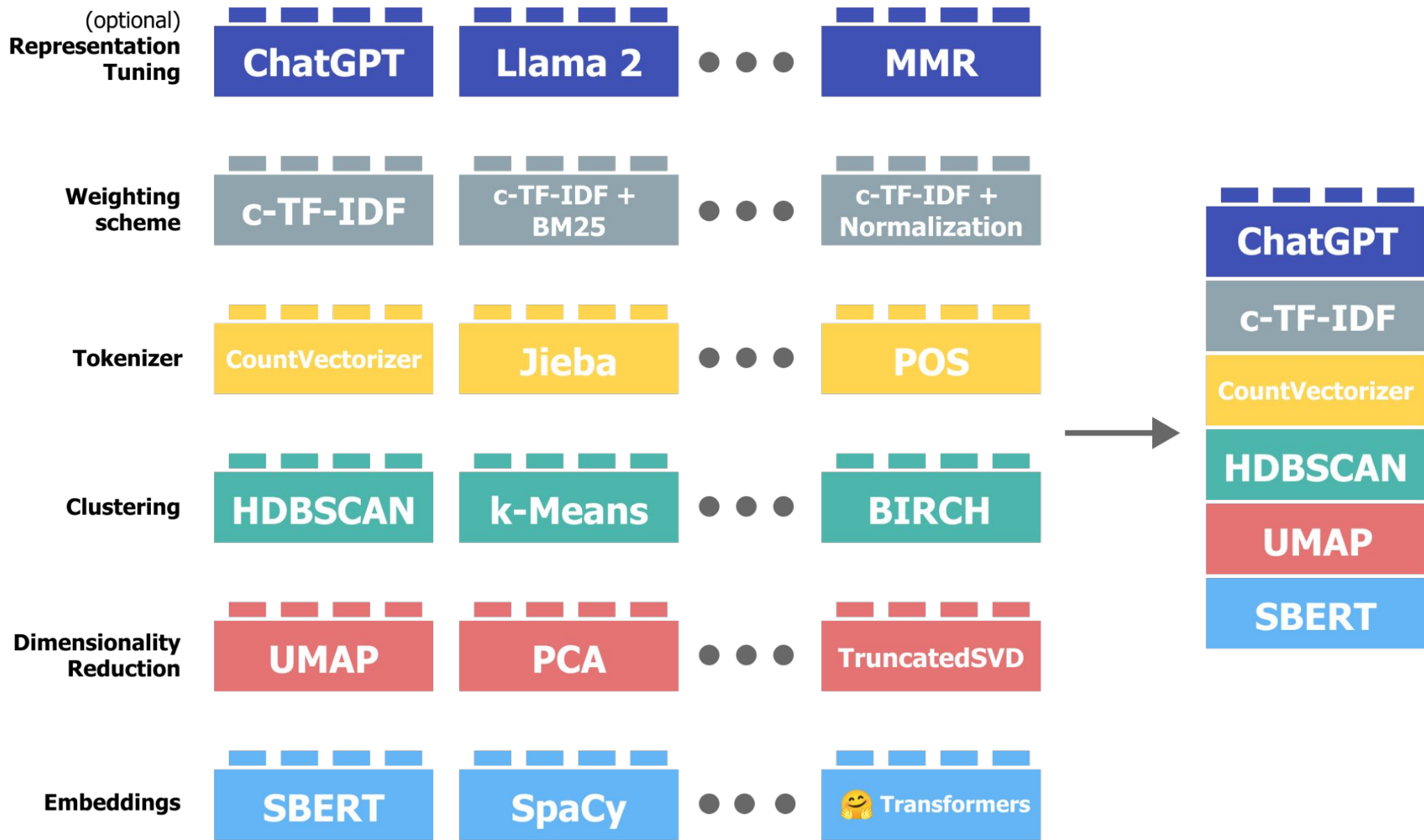
disease
host
bacteria
diseases
resistance
bacterial
new
strains
control
infectious
malaria
parasite
parasites
united
tuberculosis

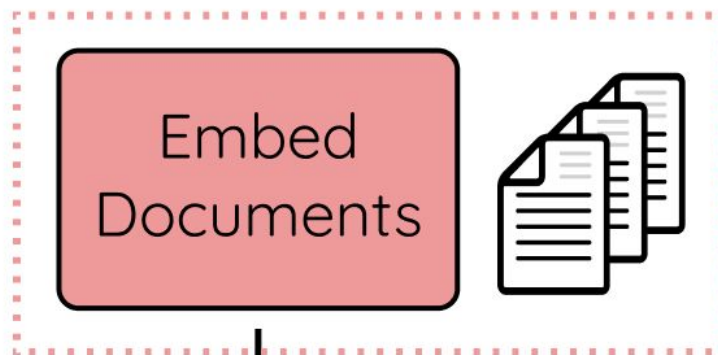
“Computers”

computer
models
information
data
computers
system
network
systems
model
parallel
methods
networks
software
new
simulations

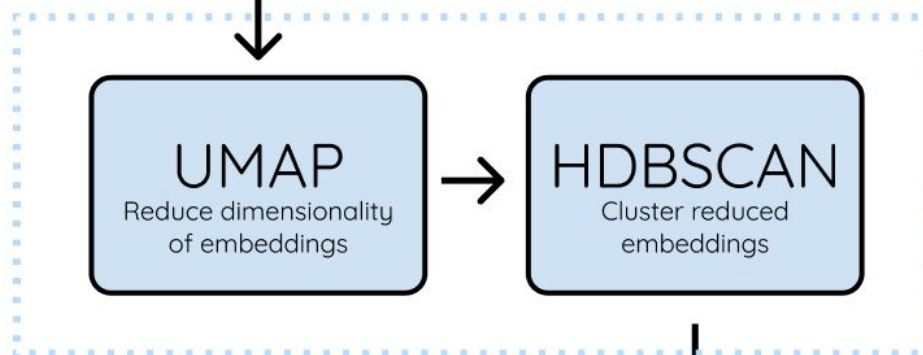
BERTopic is another, even better alternative





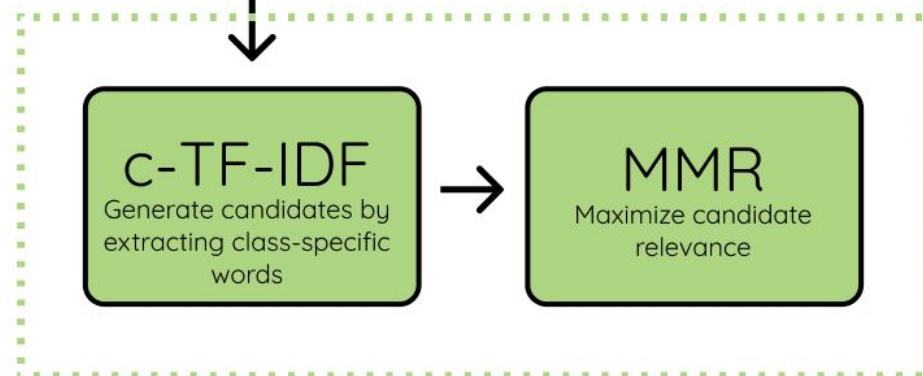


Although BERT is typically used for embedding documents, any embedding technique can be used.



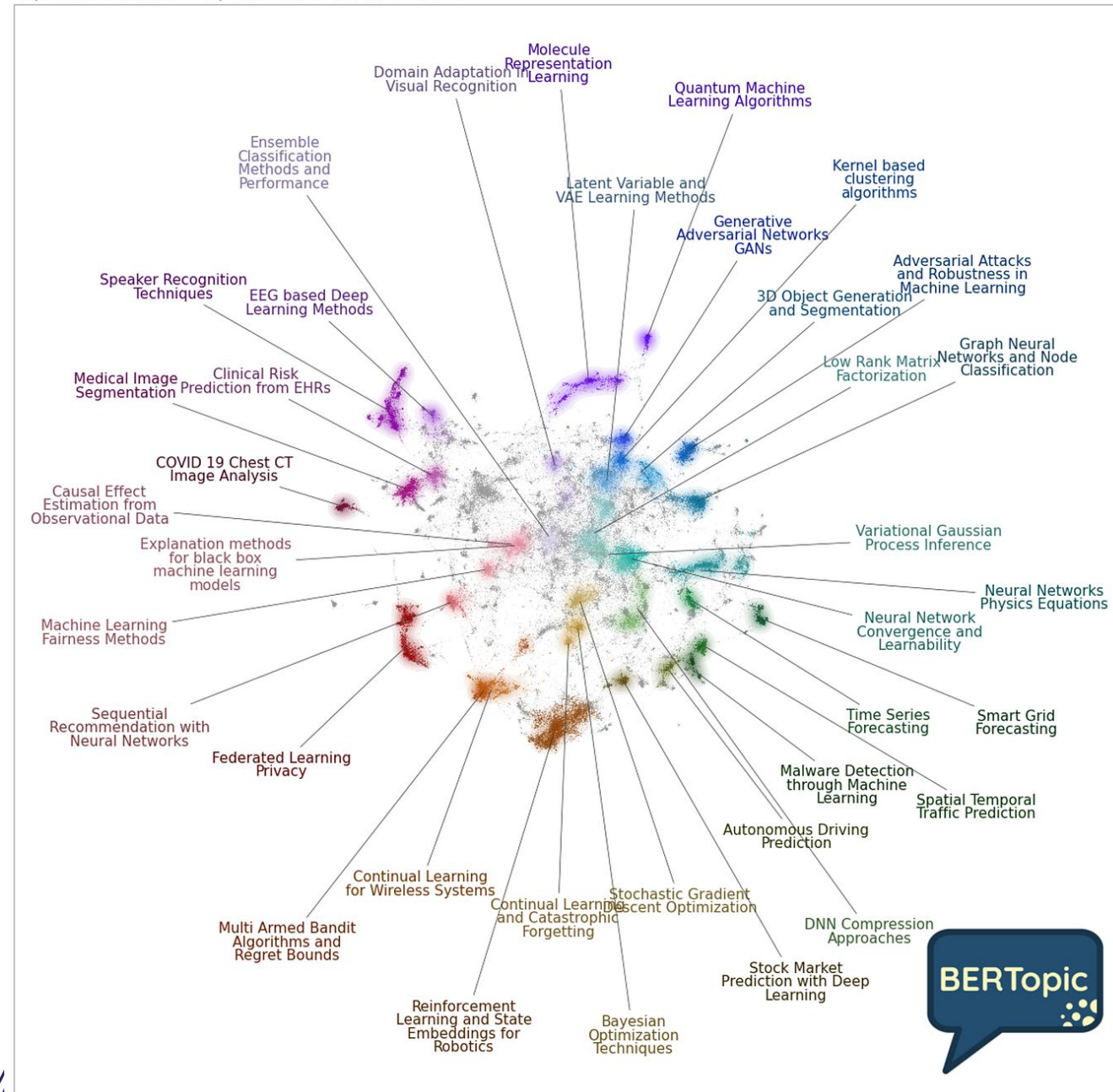
Cluster documents into semantically similar clusters

Create topic representations from clusters

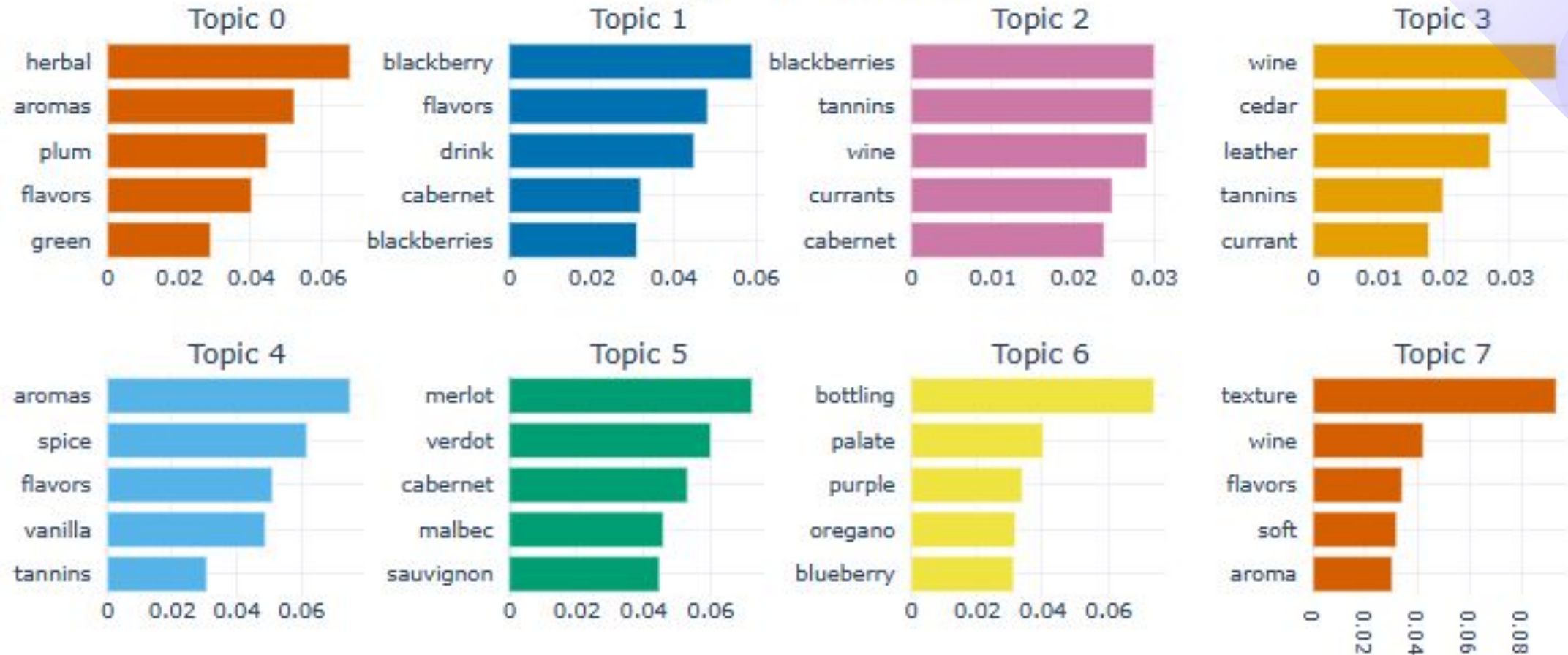


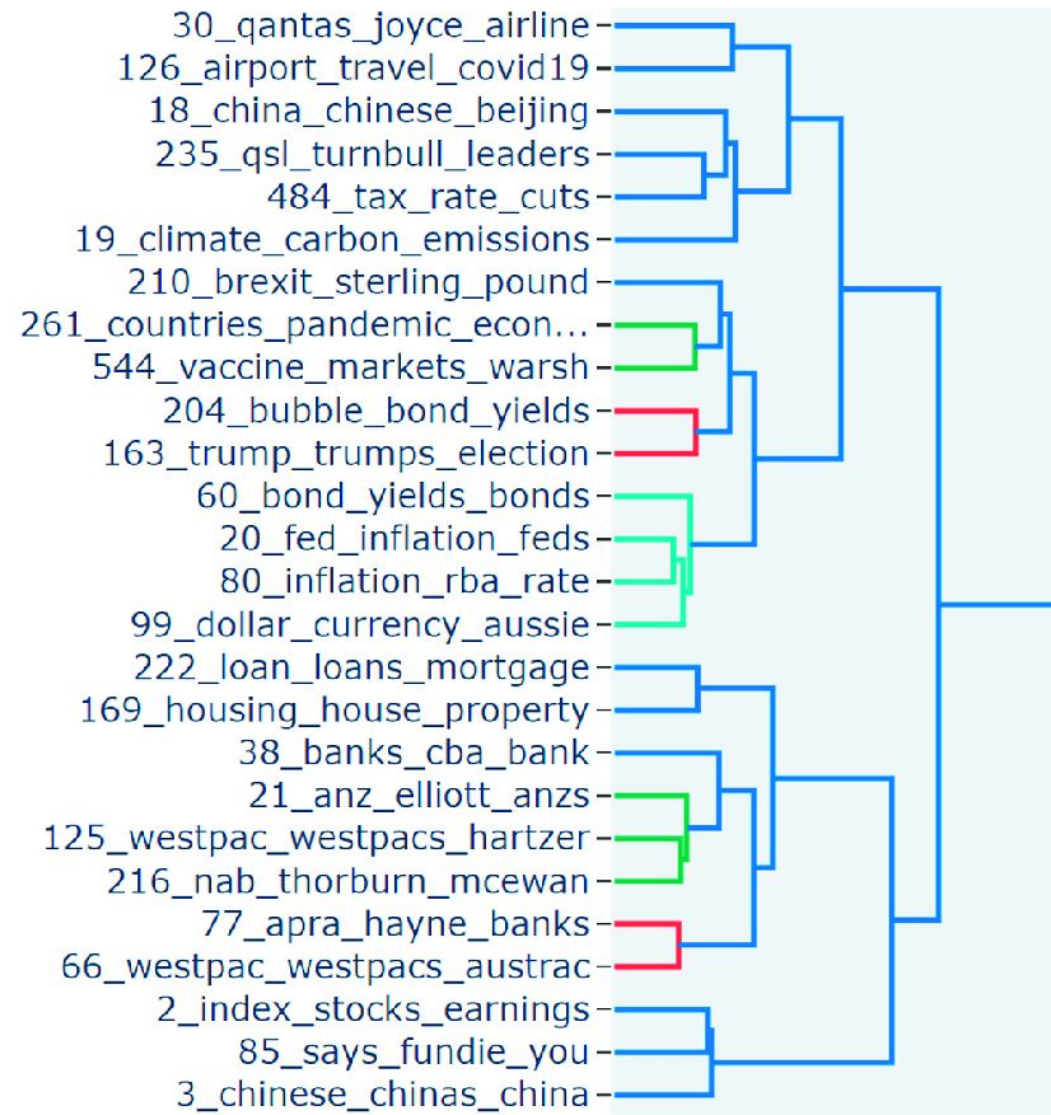
ArXiv - BERTopic

Topics labeled with `openhermes-2.5-mistral-7b`

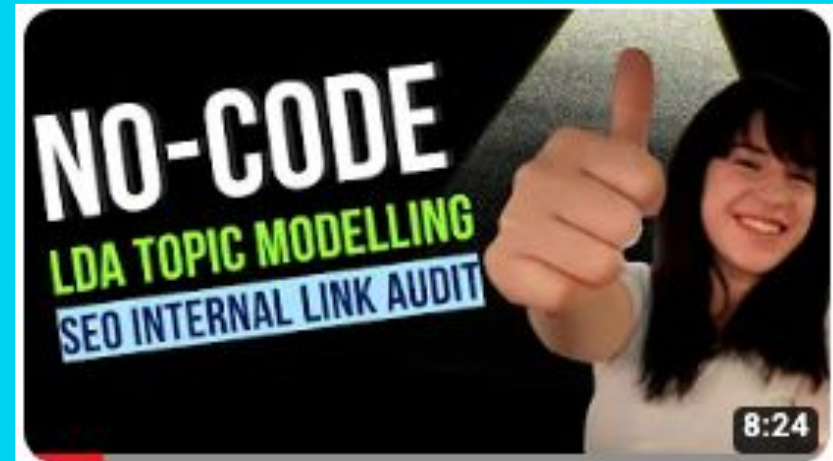


Topic Word Scores



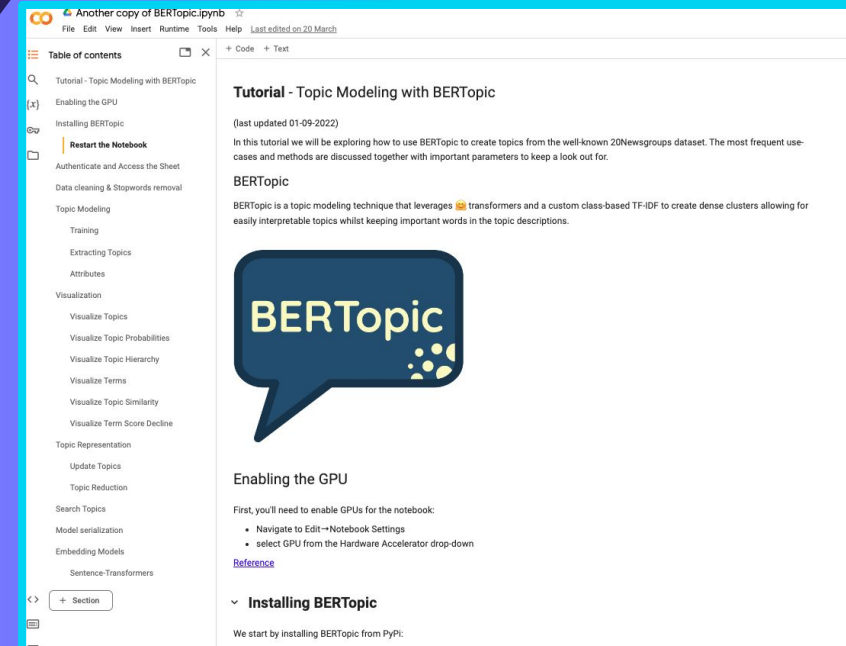


You can do
no-code LDA
topic modelling

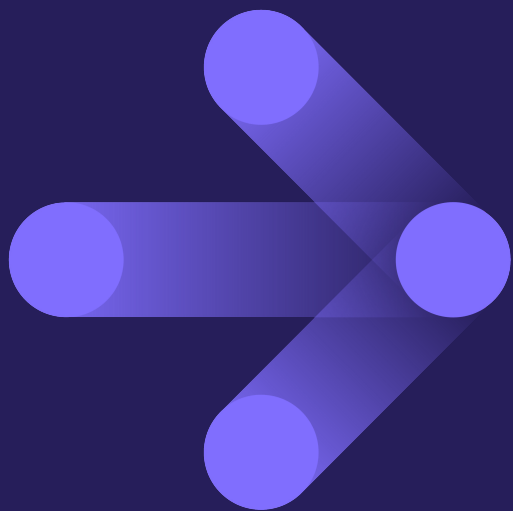


Topic Modeling with LDA web-based app for
finding internal linking opportunities | SEO...

To use BERTopic, you can use Python/Colab notebook

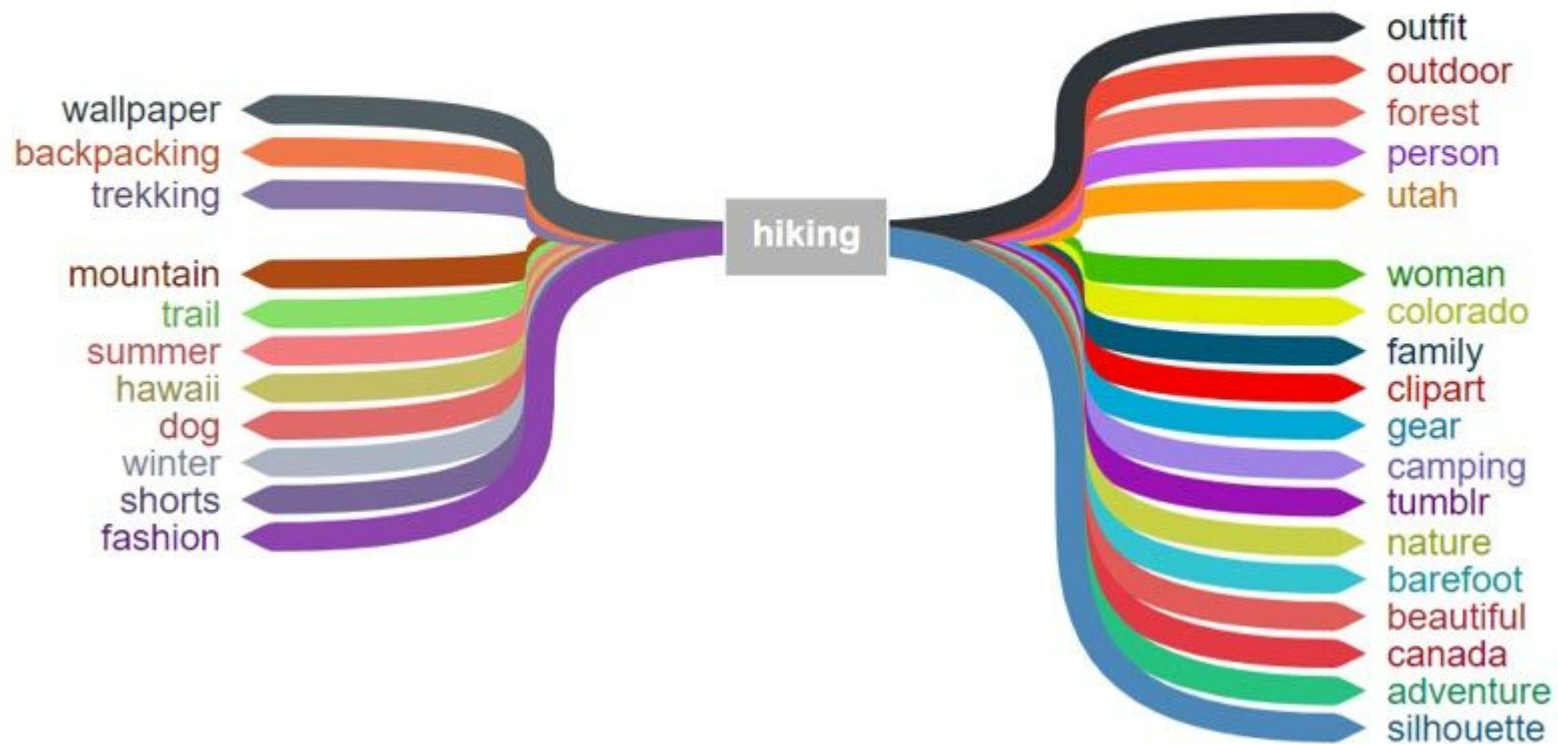


in both cases, you can feed the data
back into your audit spreadsheet



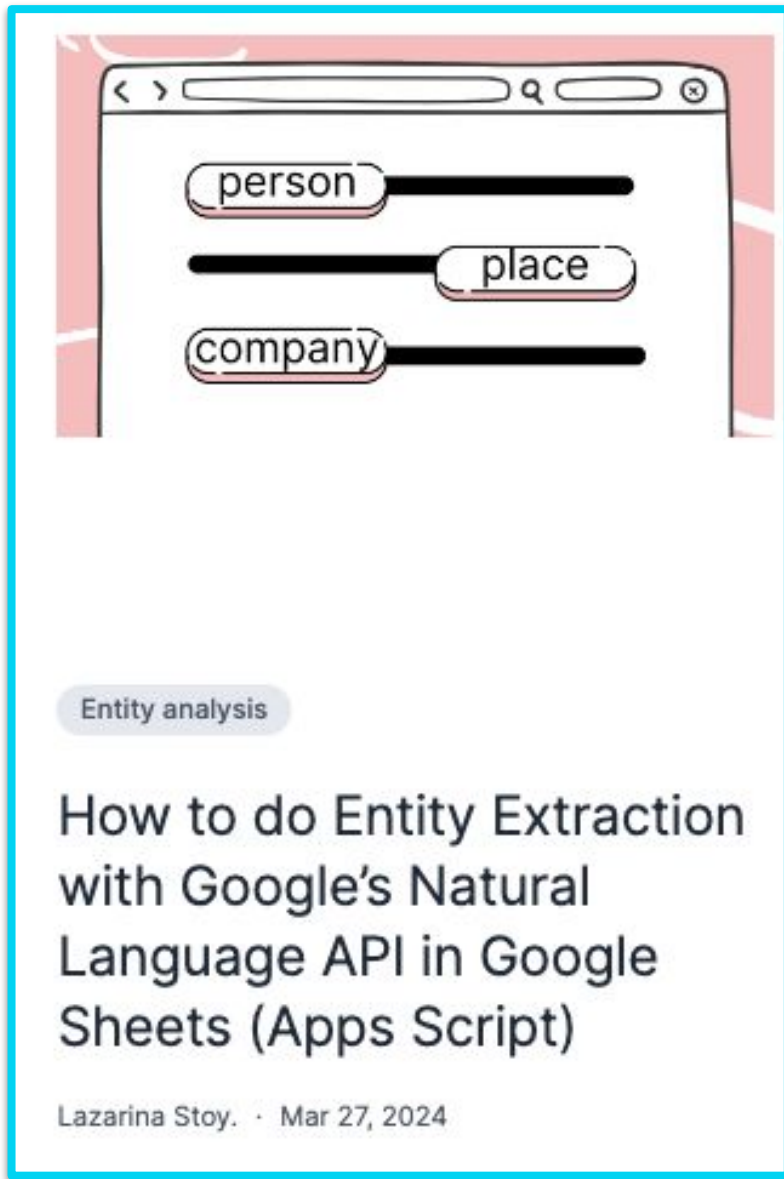
**identify and map
entities to topics**

hiking



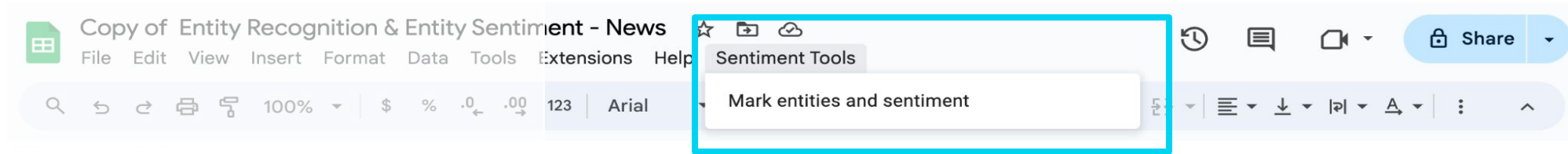


With Google's Natural Language API, you can **analyze entities**, and also get a bunch of other metrics



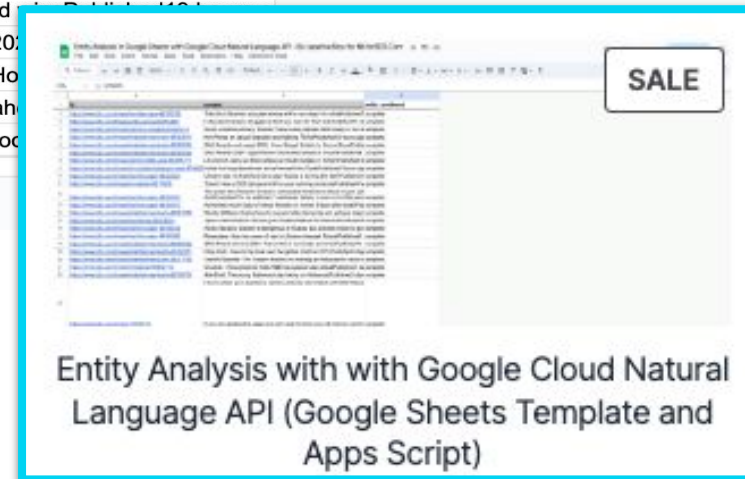
Process will take no
more than 20
minutes





Run the script via the
menu to extract
entities from content

| | B | C |
|----------|---|------------------|
| | comments | entity_sentiment |
| 6700 | The blind Ukrainian amputee whose wife's voice ke | complete |
| 68255490 | Exhausted Ukraine struggles to find new men for fr | complete |
| 68395354 | South Carolina primary: Donald Trump easily defe | complete |
| 68395355 | Kim Petras on sexual liberation and fighting TikTok | complete |
| 68395173 | SAG Awards red carpet 2024: From Margot Robbie | complete |
| 68395173 | SAG Awards 2024: Op | |
| 68395173 | US and UK carry out fresh strikes | |
| 68395173 | Inside the long-abandoned tunnel beneath the Clyde | |
| 68395173 | Ukraine war: Is Avdiivka's fall a sign | |
| 68395173 | Travel: How a £525 bet gave birth to | |
| 68395173 | Two years into Russian invasion, ex | |
| 68395173 | Authorities return body of Alexei Navalny to mother 8 days after death | |
| 68395173 | Japan naked festival: Women join Hadaka Matsuri for first time | |
| 68395173 | Alexei Navalny: Dissent is dangerous in Russia, but activists refuse to give up | |
| 68395173 | Rosenberg: How two years of war in Ukraine changed Russia | |
| 68395173 | SAG Award winners 2024: The full list of nominees and | |
| 68395173 | Stray Kids: How K-Pop took over the global charts in 20 | |
| 68395173 | Gareth Edwards: The Creator director on shaking up Ho | |
| 68395173 | Chuckie: 1Xtra presenter feels R&B has special year ah | |
| 68395173 | Alia Bhatt: The young Bollywood star taking on Hollywo | |



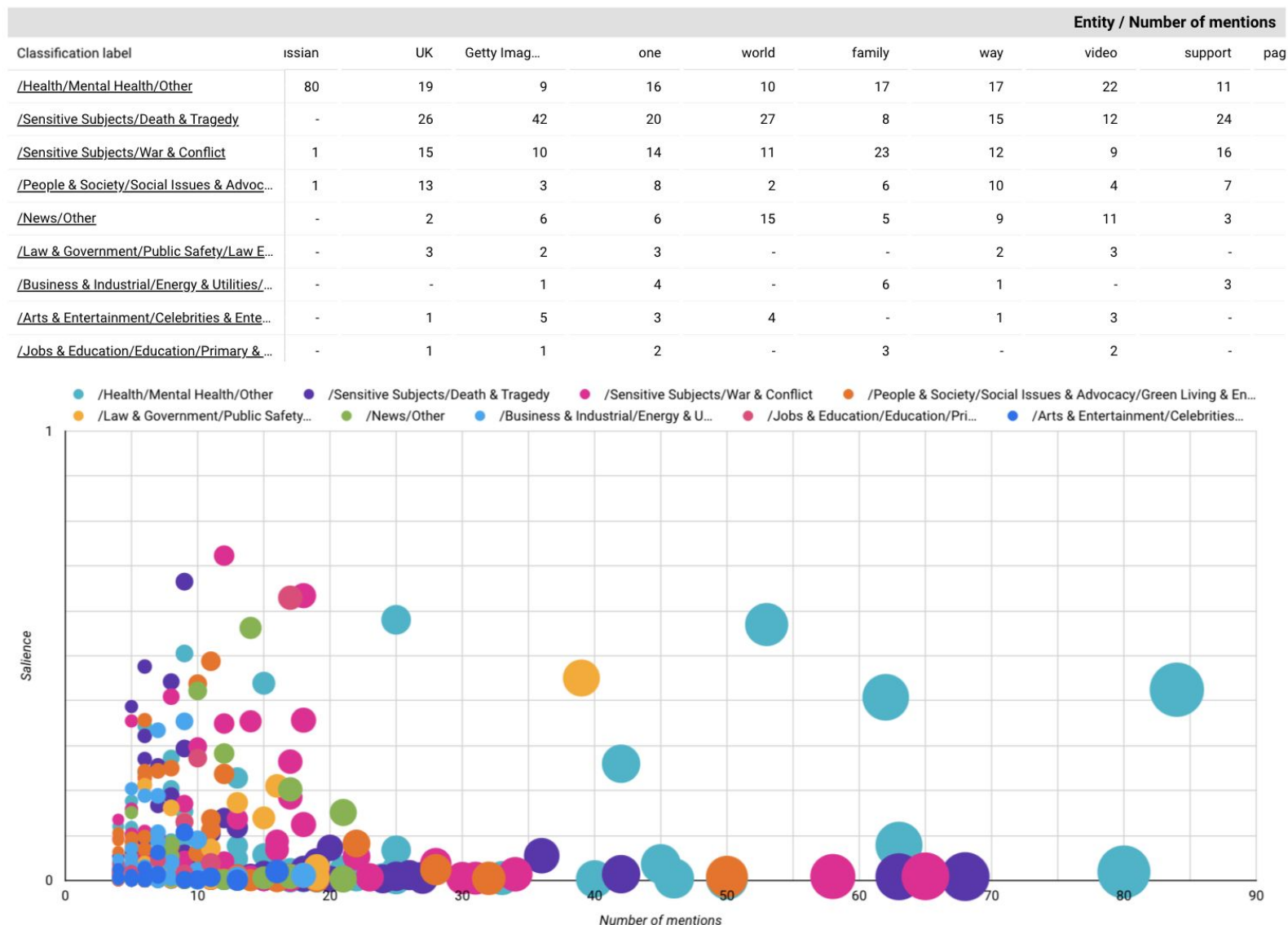
Entity Analysis with with Google Cloud Natural
Language API (Google Sheets Template and
Apps Script)

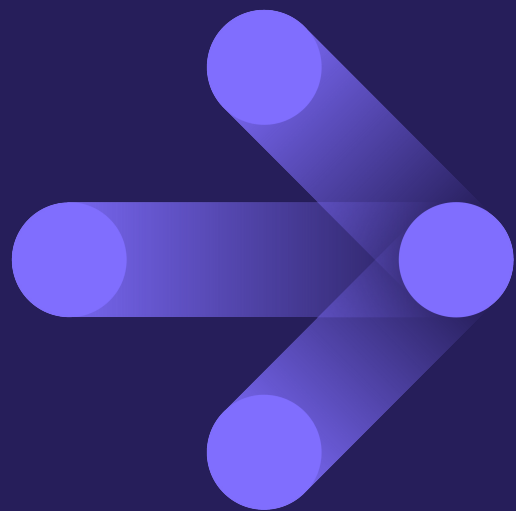
Menus 100% 123 arial 10 B I A

| ID | Entity | Type | Salience | Sentiment Score | Sentiment Magnitude | Number of mentions | Metadata | Mentions |
|---|------------------------|--------------|--------------|-----------------|---------------------|--------------------|---|--|
| https://www.bbc.com/news/world-europe-68376700 | Serhiy | PERSON | 0.34382942 | 0 | 2.6 | 6 | {} | Serhiy, Serhiy, Serhiy, Serhiy, Serhiy, Serhiy |
| https://www.bbc.com/news/world-europe-68376700 | Valeria | PERSON | 0.20473818 | 0 | 1.9 | 8 | {} | wife, Valeria, Valeria, Valeria, Valeria, Valeria |
| https://www.bbc.com/news/world-europe-68376700 | again.It | OTHER | 0.07993547 | 0 | 0.3 | 3 | {} | again.It, consciousness, pattern |
| https://www.bbc.com/news/world-europe-68376700 | amputee | PERSON | 0.03255884 | 0 | 1.3 | 1 | {} | amputee |
| https://www.bbc.com/news/world-europe-68376700 | men | PERSON | 0.023478702 | 0 | 0 | 2 | {} | men, men |
| https://www.bbc.com/news/world-europe-68376700 | consciousness | OTHER | 0.01324304 | 0 | 0 | 1 | {} | consciousness |
| https://www.bbc.com/news/world-europe-68376700 | legs | OTHER | 0.012497222 | 0 | 0 | 1 | {} | legs |
| https://www.bbc.com/news/world-europe-68376700 | voice | OTHER | 0.011640955 | 0 | 0 | 1 | {} | voice |
| https://www.bbc.com/news/world-europe-68376700 | voice | OTHER | 0.011640955 | 0 | 0 | 1 | {} | voice |
| https://www.bbc.com/news/world-europe-68376700 | wife | PERSON | 0.010622509 | 0 | 0 | 1 | {} | wife |
| https://www.bbc.com/news/world-europe-68376700 | thought | OTHER | 0.0101827 | 0 | 0 | 2 | {} | thought, relief |
| https://www.bbc.com/news/world-europe-68376700 | hospital bed | OTHER | 0.009335752 | 0 | 0 | 1 | {} | hospital bed |
| https://www.bbc.com/news/world-europe-68376700 | BakerBBC NewsAs Serhiy | OTHER | 0.00860053 | 0 | 0 | 1 | {} | BakerBBC NewsAs Serhiy |
| https://www.bbc.com/news/world-europe-68376700 | Ukrainian | LOCATION | 0.007102234 | 0 | 0 | 10 | { "mid": "/m/07121", "wikipedia_url": "https://en.wikipedia.org/wiki/Ukrainian", "wikipedia_title": "Ukrainian", "wikipedia_description": "Ukrainian, Ukrainian, Ukraine, Ukraine, Ukrainian" } | Ukrainian, Ukrainian, Ukrainian, Ukraine, Ukrainian, Ukrainian |
| https://www.bbc.com/news/world-europe-68376700 | soldier | PERSON | 0.0063414737 | 0 | 0 | 1 | {} | soldier |
| https://www.bbc.com/news/world-europe-68376700 | sharingRelated TopicsW | OTHER | 0.006290309 | 0 | 0 | 0 | {} | |
| https://www.bbc.com/news/world-europe-68376700 | Kyiv | LOCATION | 0.006290309 | 0 | 0 | 0 | {} | |
| https://www.bbc.com/news/world-europe-68376700 | pageCopy linkAbout | OTHER | 0.005526677 | 0 | 0 | 0 | {} | |
| https://www.bbc.com/news/world-europe-68376700 | agoShareclose panelSh | OTHER | 0.005526677 | 0 | 0 | 0 | {} | |
| https://www.bbc.com/news/world-europe-68376700 | UkraineBy Kelligh | OTHER | 0.005526677 | 0 | 0 | 0 | {} | |
| https://www.bbc.com/news/world-europe-68376700 | Russia | LOCATION | 0.005216035 | 0 | 0.2 | 6 | { "wiki": "Russia", "wiki_title": "Russia", "wiki_description": "Russia, Russia, Russia, Russia, Russia, Russia" } | Russia, Russia, Russia, Russia, Russia, Russia |
| https://www.bbc.com/news/world-europe-68376700 | tube | OTHER | 0.005082654 | 0 | 0 | 1 | {} | |
| https://www.bbc.com/news/world-europe-68376700 | consciousness | OTHER | 0.0050217225 | 0 | 0 | 2 | {} | |
| https://www.bbc.com/news/world-europe-68376700 | dreams | OTHER | 0.004847212 | -0.4 | 0.8 | 1 | {} | |
| https://www.bbc.com/news/world-europe-68376700 | wounds | OTHER | 0.004582253 | 0 | 0 | 2 | {} | |
| https://www.bbc.com/news/world-europe-68376700 | throat | OTHER | 0.004465256 | -0.1 | 0.1 | 1 | {} | throat |
| https://www.bbc.com/news/world-europe-68376700 | panic | OTHER | 0.004465256 | 0 | 0 | 1 | {} | panic |
| https://www.bbc.com/news/world-europe-68376700 | darkness | OTHER | 0.004465256 | 0 | 0 | 1 | {} | darkness |
| https://www.bbc.com/news/world-europe-68376700 | organisations | ORGANIZATION | 0.004451084 | 0 | 0 | 1 | {} | organisations |
| https://www.bbc.com/news/world-europe-68376700 | approach | OTHER | 0.0041478397 | 0 | 0 | 1 | {} | approach |
| https://www.bbc.com/news/world-europe-68376700 | veterans | PERSON | 0.0039000588 | 0 | 0.1 | 1 | {} | veterans |
| https://www.bbc.com/news/world-europe-68376700 | Ukrainians | PERSON | 0.0031437073 | 0 | 0 | 2 | { "wikipedia_url": "https://en.wikipedia.org/wiki/Ukrainians", "wikipedia_title": "Ukrainians", "wikipedia_description": "Ukrainians, Ukrainians" } | Ukrainians, Ukrainians |
| https://www.bbc.com/news/world-europe-68376700 | veteransHe | PERSON | 0.0030598007 | 0 | 0.2 | 1 | {} | veteransHe |
| https://www.bbc.com/news/world-europe-68376700 | family | PERSON | 0.002669544 | 0 | 0 | 1 | {} | family |
| https://www.bbc.com/news/world-europe-68376700 | injuries | OTHER | 0.0024970311 | 0 | 0 | 1 | {} | injuries |
| https://www.bbc.com/news/world-europe-68376700 | invasion | EVENT | 0.0024199213 | 0 | 0 | 1 | {} | invasion |
| https://www.bbc.com/news/world-europe-68376700 | consciousness | OTHER | 0.0023867677 | -0.3 | 0.3 | 1 | {} | consciousness |
| https://www.bbc.com/news/world-europe-68376700 | soldier | PERSON | 0.0023730078 | 0 | 0 | 1 | {} | soldier |

Get entity data in seconds

Map to topics,
labels or
intent
classifications





**evaluate content
helpfulness and EEAT**

Aleyda has this checklist with questions from Google's guidelines



| Questions to assess your content Quality, Helpfulness and Experience | Area | Expected Answer |
|---|-------------------------------------|-----------------|
| Does the content seem to be serving the genuine interests of visitors to the site or does it seem to exist solely by someone attempting to guess what might rank well in search engines? | Helpfulness and People Satisfaction | No |
| Are you producing lots of content on many different topics in hopes that some of it might perform well in search results? | Helpfulness and People Satisfaction | No |
| Are you using extensive automation to produce content on many topics? | Helpfulness and People Satisfaction | No |
| Are you mainly summarizing what others have to say without adding much value? | Helpfulness and People Satisfaction | No |
| Are you writing about things simply because they seem trending and not because you'd write about them otherwise for your existing audience? | Helpfulness and People Satisfaction | No |
| Does your content leave readers feeling like they need to search again to get better information from other sources? | Helpfulness and People Satisfaction | No |
| Are you writing to a particular word count because you've heard or read that Google has a preferred word count? | Helpfulness and People Satisfaction | No |
| Does your content actually answer the question that it promised to answer? | Helpfulness and People Satisfaction | Yes |
| Do you have an existing or intended audience for your business or site that would find the content useful if they came directly to you? | Helpfulness and People Satisfaction | Yes |
| Does your site have a primary purpose or focus? | Helpfulness and People Satisfaction | Yes |
| Will someone reading your content leave feeling like they've had a satisfying experience and learned enough about a topic to help achieve their goal? | Helpfulness and People Satisfaction | Yes |
| Does the content provide substantial value when compared to other pages in search results? | Helpfulness and People Satisfaction | Yes |
| Does the content provide original information, reporting, research, or analysis? | Quality | Yes |
| Does the content provide a substantial, complete, or comprehensive description of the topic? | Quality | Yes |
| Does the content provide insightful analysis or interesting information that is beyond the obvious? | Quality | Yes |
| If the content draws on other sources, does it avoid simply copying or rewriting those sources, and instead provide substantial additional value and originality? | Quality | Yes |
| Does the main heading or page title provide a descriptive, helpful summary of the content? | Quality | Yes |
| Does the main heading or page title avoid exaggerating or being shocking in nature? | Quality | Yes |
| Is this the sort of page you'd want to bookmark, share with a friend, or recommend? | Quality | Yes |
| Would you expect to see this content in or referenced by a printed magazine, encyclopedia, or book? | Quality | Yes |
| Is it self-evident to your visitors who authored your content, do pages carry a byline, where one might be expected? | Expertise and Experience | Yes |
| Do bylines lead to further information about the author or authors involved, giving background about them and the areas they write about? | Expertise and Experience | Yes |
| If automation is used to substantially generate content, is the use of automation, including AI-generation, self-evident to visitors through disclosures or in other ways? | Expertise and Experience | Yes |
| Are you providing background about how automation or AI-generation was used to create content (in case it was)? | Expertise and Experience | Yes |
| Are you explaining why automation or AI was seen as useful to produce content (in case it was)? | Expertise and Experience | Yes |
| Does the content present information in a way that makes you want to trust it, such as clear sourcing, evidence of the expertise involved, background about the author or the site that publishes it, such as through links to an author page or a site's About page? | Expertise and Experience | Yes |
| If someone researched the site producing the content, would they come away with an impression that it is well-trusted or widely-recognized as an authority on its topic? | Expertise and Experience | Yes |
| Is this content written by an expert or enthusiast who demonstrably knows the topic well? | Expertise and Experience | Yes |
| Is the content free from easily-verified factual errors? | Expertise and Experience | Yes |
| Does your content clearly demonstrate first-hand expertise and a depth of knowledge (for example, expertise that comes from having actually used a product or service, or visiting a place)? | Expertise and Experience | Yes |
| Is the content free from spelling or stylistic issues? | Presentation and production | Yes |
| Was the content produced well, or does it appear sloppy or hastily produced? | Presentation and production | Yes |
| Is the content mass-produced by or outsourced to a large number of creators, or spread across a large network of sites, so that individual pages or sites don't get as much attention or care? | Presentation and production | Yes |
| Do pages have good Core Web Vitals? | Page Experience | Yes |
| Are pages served in a secure fashion? | Page Experience | Yes |
| Does content display well for mobile devices when viewed on them? | Page Experience | Yes |
| Does the content lack an excessive amount of ads that distract from or interfere with the main content? | Page Experience | Yes |
| Do pages lack intrusive interstitials or pop-ups? | Page Experience | Yes |
| Can visitors navigate to or locate the main content of your pages easily? | Page Experience | Yes |
| Is the page designed so visitors can easily distinguish the main content from other content on your page? | Page Experience | Yes |

with a bit of ✨ creativity,
you can automate some checks

Are you producing lots of content on **many different topics** in hopes that some of it might perform well in search results?



of content pieces
topic models
of topics

Does your content **leave readers feeling like they need to search again** to get better information from other sources?



Bounce rate
Time on page

Is it self-evident to your visitors
who authored your content, do
pages carry a byline, where one
might be expected?



Author byline

Do **bylines lead to further information** about the author or authors involved, giving background about them and the areas they write about?



Author page

Green – can be checked easily
 Yellow – can be checked with the help of a custom model/calculation
 Red – difficult to check at scale



| Questions to assess your content Quality, Helpfulness and Experience | Area | Expected Answer |
|---|-------------------------------------|-----------------|
| Does the content seem to be serving the genuine interests of visitors to the site or does it seem to exist solely by someone attempting to guess what might rank well in search engines? | Helpfulness and People Satisfaction | No |
| Are you producing lots of content on many different topics in hopes that some of it might perform well in search results? | Helpfulness and People Satisfaction | No |
| Are you using extensive automation to produce content on many topics? | Helpfulness and People Satisfaction | No |
| Are you mainly summarizing what others have to say without adding much value? | Helpfulness and People Satisfaction | No |
| Are you writing about things simply because they seem trending and not because you'd write about them otherwise for your existing audience? | Helpfulness and People Satisfaction | No |
| Does your content leave readers feeling like they need to search again to get better information from other sources? | Helpfulness and People Satisfaction | No |
| Are you writing to a particular word count because you've heard or read that Google has a preferred word count? | Helpfulness and People Satisfaction | No |
| Does your content actually answer the question that it promised to answer? | Helpfulness and People Satisfaction | Yes |
| Do you have an existing or intended audience for your business or site that would find the content useful if they came directly to you? | Helpfulness and People Satisfaction | Yes |
| Does your site have a primary purpose or focus? | Helpfulness and People Satisfaction | Yes |
| Will someone reading your content leave feeling like they've had a satisfying experience and learned enough about a topic to help achieve their goal? | Helpfulness and People Satisfaction | Yes |
| Does the content provide substantial value when compared to other pages in search results? | Helpfulness and People Satisfaction | Yes |
| Does the content provide original information, reporting, research, or analysis? | Quality | Yes |
| Does the content provide a substantial, complete, or comprehensive description of the topic? | Quality | Yes |
| Does the content provide insightful analysis or interesting information that is beyond the obvious? | Quality | Yes |
| If the content draws on other sources, does it avoid simply copying or rewriting those sources, and instead provide substantial additional value and originality? | Quality | Yes |
| Does the main heading or page title provide a descriptive, helpful summary of the content? | Quality | Yes |
| Does the main heading or page title avoid exaggerating or being shocking in nature? | Quality | Yes |
| Is this the sort of page you'd want to bookmark, share with a friend, or recommend? | Quality | Yes |
| Would you expect to see this content in or referenced by a printed magazine, encyclopedia, or book? | Quality | Yes |
| Is it self-evident to your visitors who authored your content, do pages carry a byline, where one might be expected? | Expertise and Experience | Yes |
| Do bylines lead to further information about the author or authors involved, giving background about them and the areas they write about? | Expertise and Experience | Yes |
| If automation is used to substantially generate content, Is the use of automation, including AI-generation, self-evident to visitors through disclosures or in other ways? | Expertise and Experience | Yes |
| Are you providing background about how automation or AI-generation was used to create content (in case it was)? | Expertise and Experience | Yes |
| Are you explaining why automation or AI was seen as useful to produce content (in case it was)? | Expertise and Experience | Yes |
| Does the content present information in a way that makes you want to trust it, such as clear sourcing, evidence of the expertise involved, background about the author or the site that publishes it, such as through links to an author page or a site's About page? | Expertise and Experience | Yes |
| If someone researched the site producing the content, would they come away with an impression that it is well-trusted or widely-recognized as an authority on its topic? | Expertise and Experience | Yes |
| Is this content written by an expert or enthusiast who demonstrably knows the topic well? | Expertise and Experience | Yes |
| Is the content free from easily-verified factual errors? | Expertise and Experience | Yes |
| Does your content clearly demonstrate first-hand expertise and a depth of knowledge (for example, expertise that comes from having actually used a product or service, or visiting a place)? | Expertise and Experience | Yes |
| Is the content free from spelling or stylistic issues? | Presentation and production | Yes |
| Was the content produced well, or does it appear sloppy or hastily produced? | Presentation and production | Yes |
| Is the content mass-produced by or outsourced to a large number of creators, or spread across a large network of sites, so that individual pages or sites don't get as much attention or care? | Presentation and production | Yes |
| Do pages have good Core Web Vitals? | Page Experience | Yes |
| Are pages served in a secure fashion? | Page Experience | Yes |
| Does content display well for mobile devices when viewed on them? | Page Experience | Yes |
| Does the content lack an excessive amount of ads that distract from or interfere with the main content? | Page Experience | Yes |
| Do pages lack intrusive interstitials or pop-ups? | Page Experience | Yes |
| Can visitors navigate to or locate the main content of your pages easily? | Page Experience | Yes |
| Is the page designed so visitors can easily distinguish the main content from other content on your page? | Page Experience | Yes |



audit semantics

Semantic Content Audit Roadmap



✗

Title starts with the primary keyword

✓

Use Prominent,
Popular and
Relevant Pairs of

Entity

Attribute

Noun

Verb/Predicate

Adjective

✗

Limited to specific keyword

✓

Covering all
angles of a query
network (query
semantics)

Covering all
relevant words
relations (lexical
semantics)

Sequential queries

Correlative queries

Query Context

Query paths

Query Augmentation

Represented Queries

Representative Queries

Canonical queries

Holonym

Meronym

Hypernym

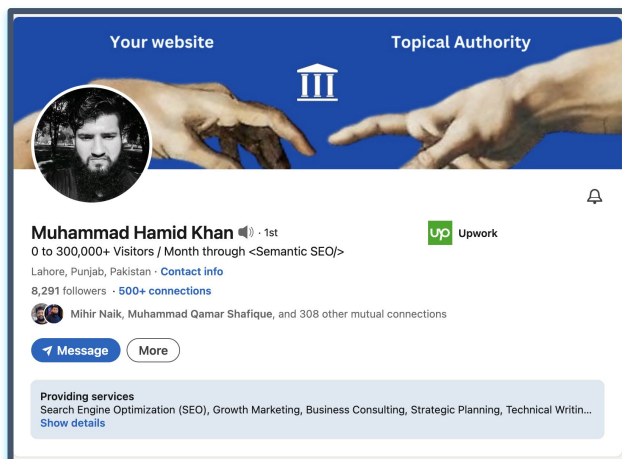
Hyponym

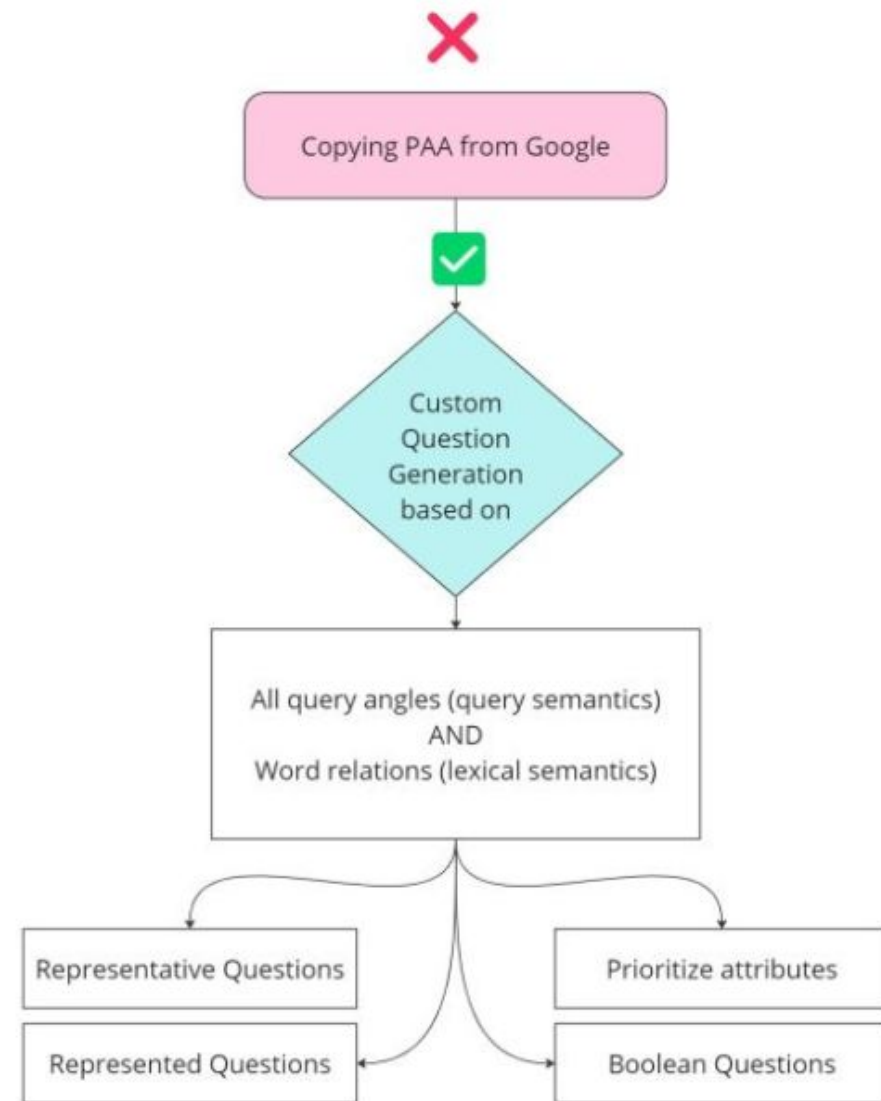
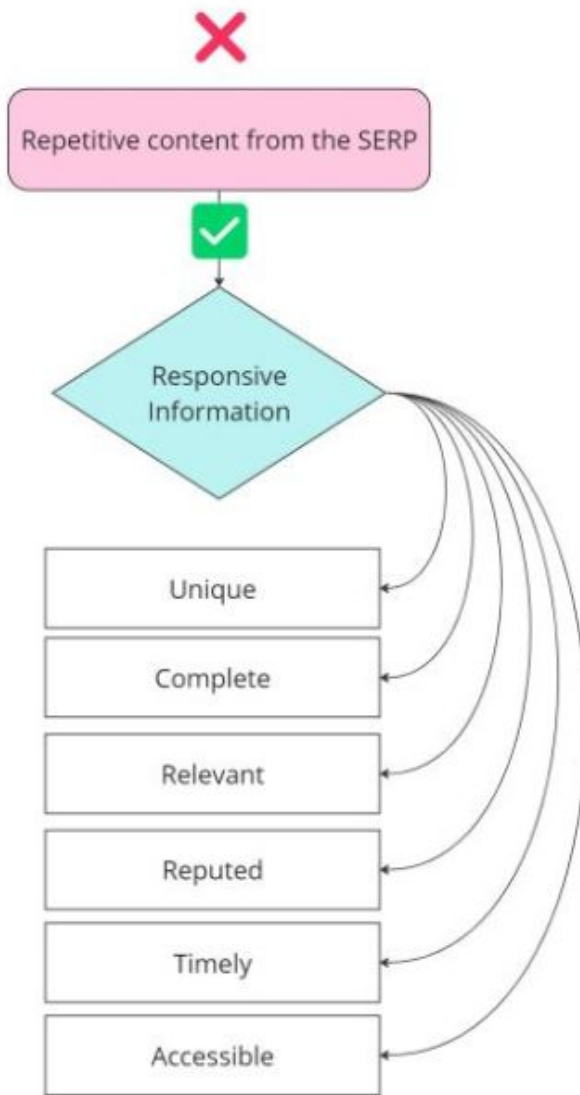
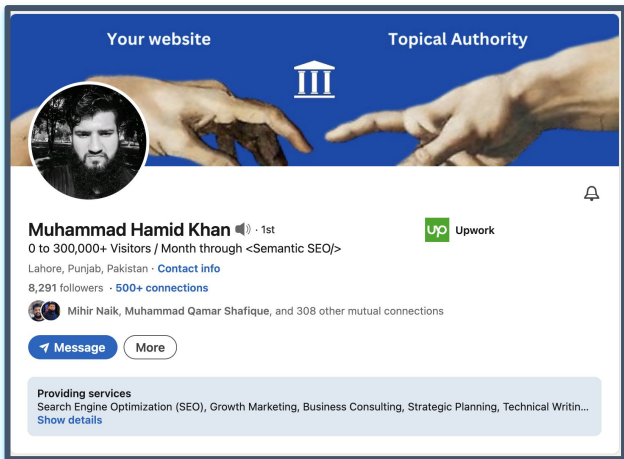
Synonym

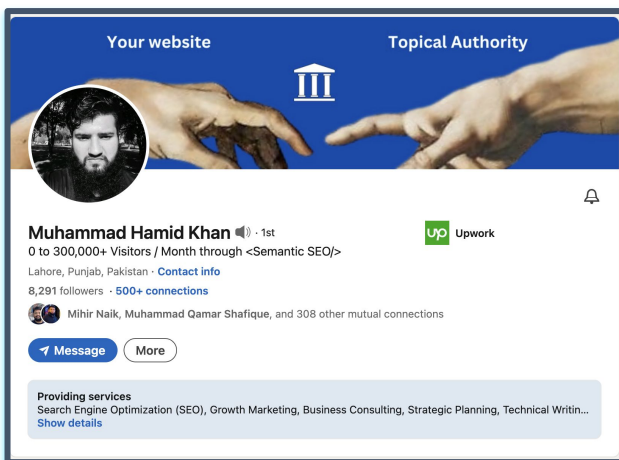
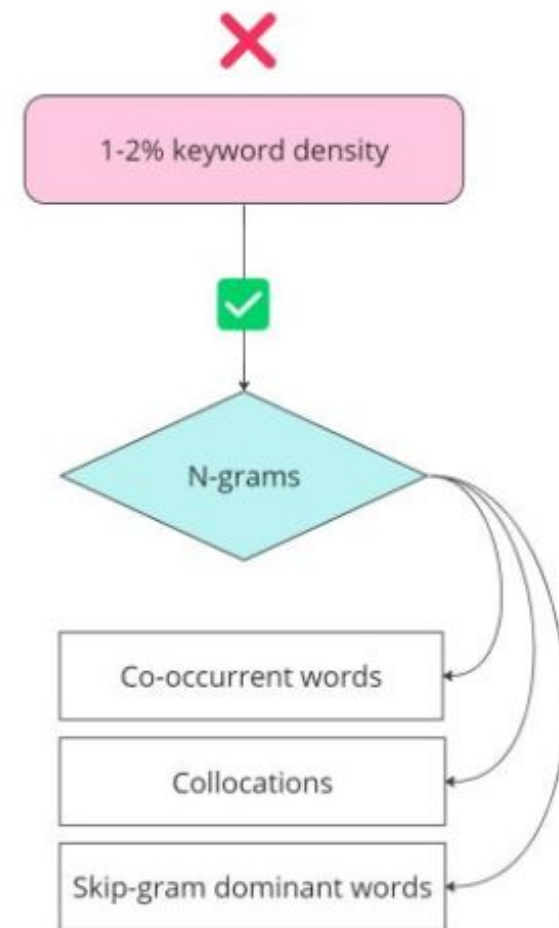
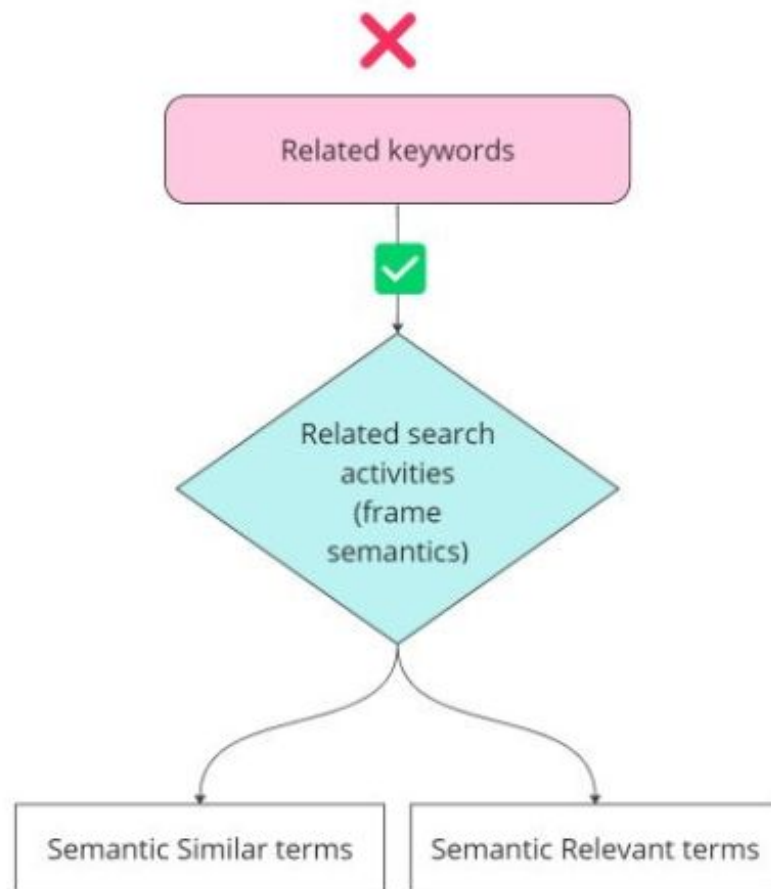
Antonym

Polyseme

Homonym







Build custom metrics



keyword density
entity density



topic relevance



query-content similarity (fuzzy
matching, n-gram analysis)

Organize your database – crawl, content, metrics

Understand the data – Categorize your inventory

Evaluate performance and assign action items to underperforming content

Implement & Review impact – Prioritize what and when to implement and repeat

goals at this stage are pretty
straightforward

01

consistency

you need consistent evaluation criteria

02

adaptable

you need a system and process that you can adapt for the website or industry

03

simple

don't over-complicate things unnecessarily

04

forward-looking

design your process and system here with the intention to hand the insights to departments/ team members, who will be implementing

If/else logic = your bff at this stage

build a
checklist with
best practices

| Area of SEO | | | |
|-------------|------------------------|---|--|
| A | B | C | D |
| Area of SEO | Check | Acceptable Criteria | Unacceptable Criteria (flagged when) |
| On-page | URL | URL is clean from special characters | Contains special characters |
| On-page | URL | URL is lowercase | Contains uppercase characters |
| On-page | URL | URL is unique | Is duplicate with other URL |
| On-page | URL | URL is short and descriptive | Is too long (over 100 characters) |
| On-page | URL | URL contains descriptive keywords | Contains stop-words (is, and, etc) |
| On-page | URL | Spaces in the URL are replaced by dashes, no spaces are contained in the final URL | Contains spaces |
| On-page | Title | Title is clean from special characters or emojis | Contains special characters or emojis |
| On-page | Title | Title is unique | Is duplicate with other URL |
| On-page | Title | Title is short and descriptive | Title is too long (over 70 characters) |
| On-page | Title | Title is not over 80% similar to another title of a page on the site | Title is over 80% similar to another title of page on the website |
| On-page | Title | Title is alligned with the standard format per section | Title is not alligned with the standard per section - autopopulated sections have been deleted, proposed structure is not kept |
| On-page | Title | A title is present | A title is not present (empty title field) |
| On-page | Title | Keywords are present in the title (main action, main value proposition) | Keywords are not present in the title (no context provided via the title of what the page is about) |
| Meta Data | Meta Description | Meta Description is clean from special characters | Meta Description Contains special characters |
| Meta Data | Meta Description | Meta Description is unique | Is duplicate with other Meta Description |
| Meta Data | Meta Description | Meta Description is descriptive and enticeful | Meta Description contains boiler plate language and is not insightful of the article |
| Meta Data | Meta Description | Meta Description is long enough to be visible in the SERP (150-170 characters) | Meta Description is too long (over 170 characters) |
| On-page | Conversion | All call-to-actions on the page are clear and strong, prompting the user to the desired conversion. | All call-to-actions on the page contain boiler plate language or are too generic (e.g. "Contact sales", or "click here") |
| On-page | Headings Copy & Format | Titles are written based on the primary purpose of the document. If a document is primarily a tutorial, but it has a conceptual introduction, write a task-based title. | Titles are mixed when considering the purpose of the page. |



| A | B | C |
|-------------|---|--|
| Area of SEO | Check | Combined if/else Statement |
| On-page URL | URL is clean from special characters | if re.search(r'^a-zA-Z0-9\-_\.\/:', url): 'Contains special characters' else: 'URL is clean' |
| On-page URL | URL is lowercase | if url != url.lower(): 'Contains uppercase characters' else: 'URL is lowercase' |
| On-page URL | URL is unique | if url in url_list: 'Is duplicate with other URL' else: 'URL is unique' |
| On-page URL | URL is short and descriptive | if len(url) > 100: 'Is too long (over 100 characters)' else: 'URL is short and descriptive' |
| On-page URL | URL contains descriptive keywords | if any(stop_word in url for stop_word in stop_words_list): 'Contains stop-words' else: 'URL contains descriptive keywords' |
| On-page URL | Spaces in the URL are replaced by dashes, no spaces | if ' ' in url: 'Contains spaces' else: 'Spaces are replaced by dashes' |

Translate each item on
your checklist into
conditional (if/else)
statements.



Translate the
if/else statements
to code for your
platform of choice



Each assessment should have
a corresponding action item + responsible person/department



Looker Studio dashboards (and custom dimensions) can also be magical for saving time for future audits



Zapier can also be amazing for automating integrations, or for setting up automated notifications for underperforming content

Organize your database – crawl, content, metrics

Understand the data – Categorize your inventory

Evaluate performance and assign action items to underperforming content

~~Implement & Review impact~~ **Build a prioritised list and organize deliverable**

what a good audit deliverable looks like to me:



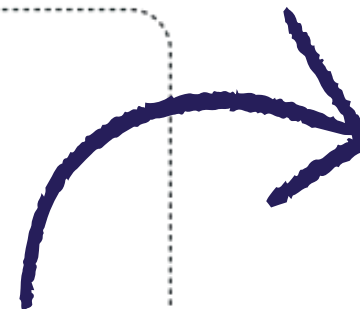
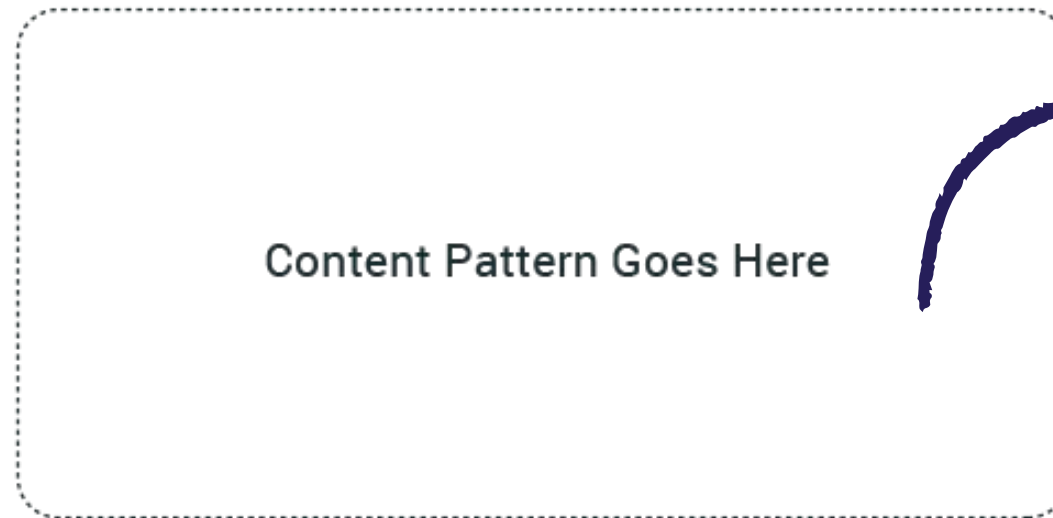
- immediately actionable
- extensive and thorough
- insightful
- contextualized and relevant for multiple stakeholder groups
- promotes action
- evokes immediate FOMO



Audit Section



Why it matters



Who should action



Auto Fill Google Doc Templates from Google Sheets Data



21:22

**Auto Fill Google Doc Template from Google Sheets
Data Using Google Apps Script**

Create integrations between Google Sheets and Google Docs to automate any workflow

Start with email for free



Start with Google for free



WHEN THIS HAPPENS

New Spreadsheet



AUTOMATICALLY DO THIS

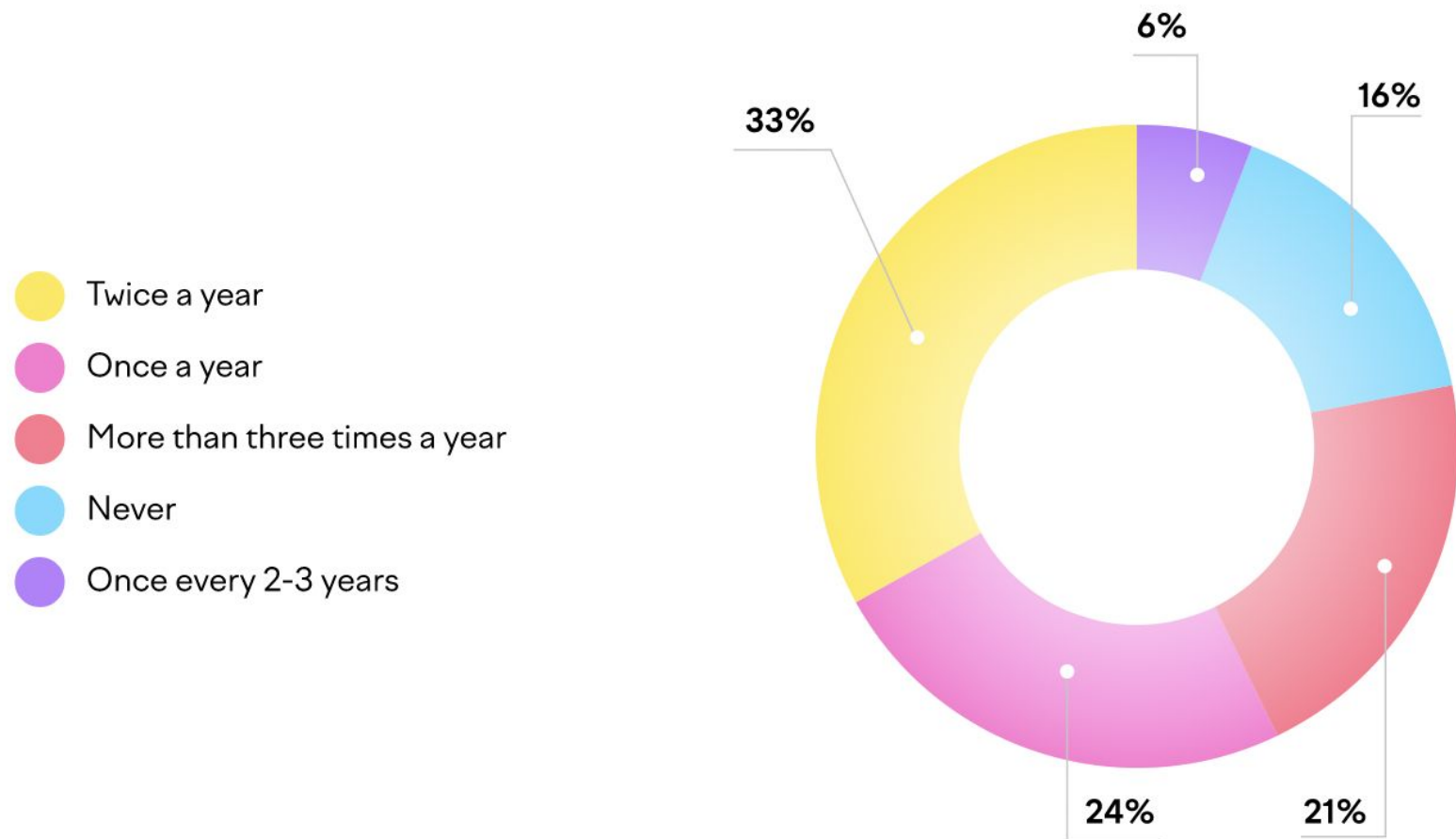
Custom Actions (Beta)



Endless possibilities

before I go...

How often do you conduct content audits?



your system



RESULTS

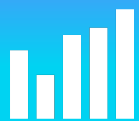
1% Better

make your process **1% better** every time

1% worse

TIME

ways to make it *even* better



Rank analysis



SERP analysis



Competitor
content audit
& content gap



Cross-platform
opportunity
analysis &
Automated
Content
Transformation



User feedback
(first party) for
content
revisions

what I do my website

seo@lazarinastoy.com

social handle

best way to get in touch



