



How to automate your content audit with ML APIs

Lazarina Stoy.

MLforSEO

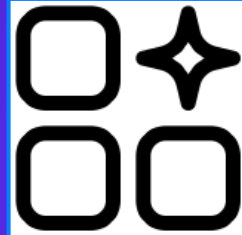
 @lazarinastoy

 @lazarinastoy

 lazarinastoy.com/resources/conference-decks-and-presentations/



Organize your database - crawl,
content, metrics



Understand the data - Categorize your
inventory

Evaluate performance and assign
action items to underperforming content



Organise your deliverable - Build a
prioritised list of recommendations

so, we'll go through the same steps

build better **systems**

implement **ML APIs**

A man with a shaved head, wearing a blue uniform, is shown from the chest up. He has a large black cylindrical object resting on his shoulders. He is making peace signs with both hands. The background is a dark, industrial-looking setting with some blue lighting.

automate

the auditing process

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

Organize your database - crawl, content, metrics

Understand the data - Categorize your inventory

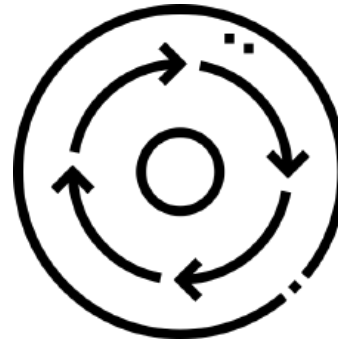
Evaluate performance and assign action items to underperforming content

Organise your deliverable - Build a prioritised list of recommendations

goals at this stage are pretty simple



accuracy



consistency

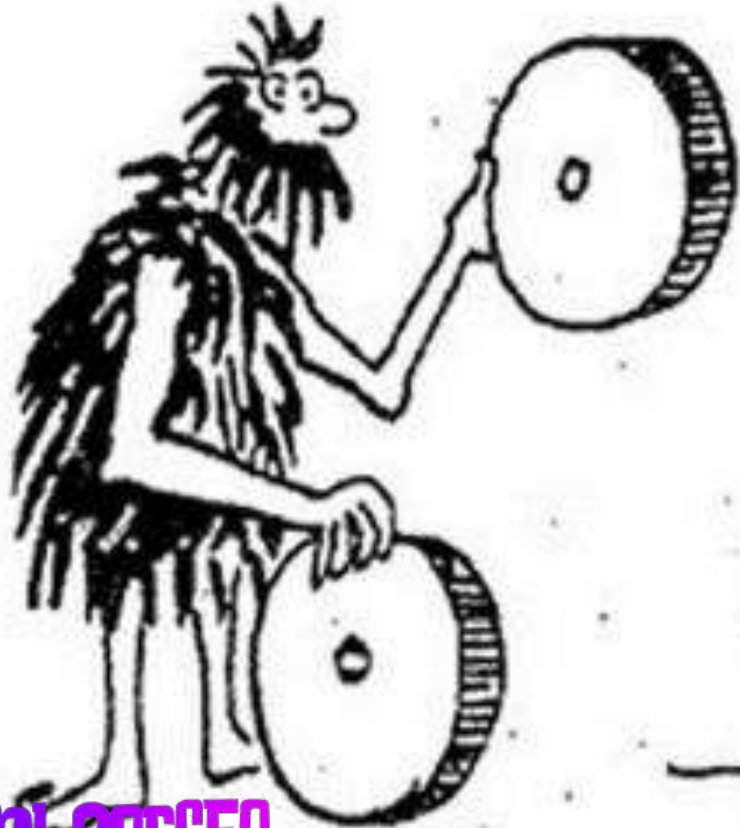


thoroughness

Crawling and scraping content

No thanks!

We are too busy



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/

Map important performance metrics



User Engagement Data

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 Data



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



Search Traffic Data (discover/news/web)

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

we can go a step further

	<p>Internal Links</p>	<p># internal links incoming # internal links outgoing # external links outgoing</p>
	<p>Content readability</p>	<p>Flesch Reading Ease Score Readability Score Word Count Sentence Count Average words per sentence</p>
	<p>Backlinks</p>	<p># backlinks # follow backlinks # nofollow backlinks # text links # referring domains</p>

And even further...

	Custom Metrics	<ul style="list-style-type: none">% branded vs non-branded keywordsIntent categoryContent decay scoreContent Type - Seasonal, Evergreen, Interview, etc.% of clicks brought by article0-click pages
--	-----------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

putting this data together is a matter of preference

Most likely



Google Sheets

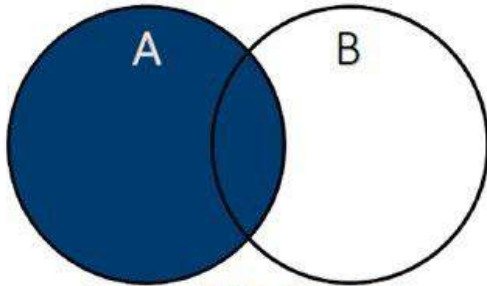


BigQuery

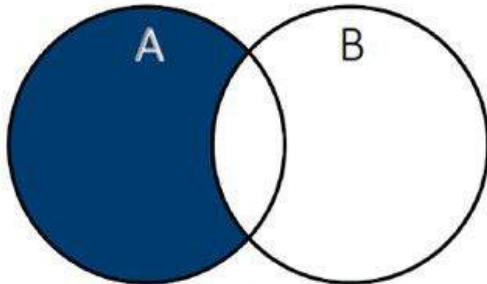
Less likely but will be useful down the line to visualise



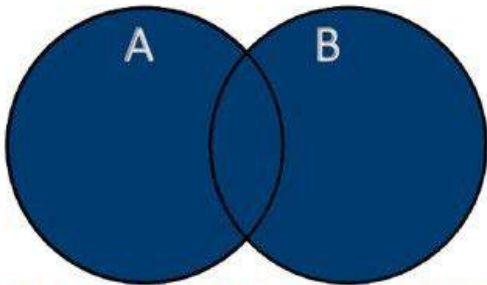
Looker Studio



LEFT INCLUSIVE

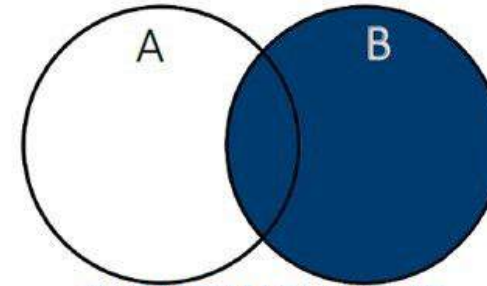


LEFT EXCLUSIVE

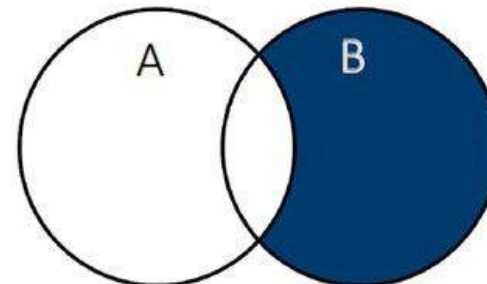


FULL OUTER INCLUSIVE

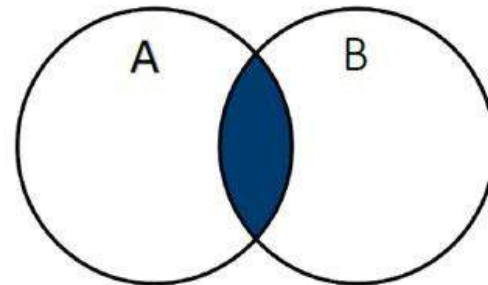
SQL JOINS	
<p>LEFT INCLUSIVE</p> <pre>SELECT [Select List] FROM TableA A LEFT OUTER JOIN TableB B ON A.Key= B.Key</pre>	<p>RIGHT INCLUSIVE</p> <pre>SELECT [Select List] FROM TableA A RIGHT OUTER JOIN TableB B ON A.Key= B.Key</pre>
<p>LEFT EXCLUSIVE</p> <pre>SELECT [Select List] FROM TableA A LEFT OUTER JOIN TableB B ON A.Key= B.Key WHERE B.Key IS NULL</pre>	<p>RIGHT EXCLUSIVE</p> <pre>SELECT [Select List] FROM TableA A LEFT OUTER JOIN TableB B ON A.Key= B.Key WHERE A.Key IS NULL</pre>
<p>FULL OUTER INCLUSIVE</p> <pre>SELECT [Select List] FROM TableA A FULL OUTER JOIN TableB B ON A.Key = B.Key</pre>	<p>FULL OUTER EXCLUSIVE</p> <pre>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</pre>
<p>INNER JOIN</p> <pre>SELECT [Select List] FROM TableA A INNER JOIN TableB B ON A.Key = B.Key</pre>	



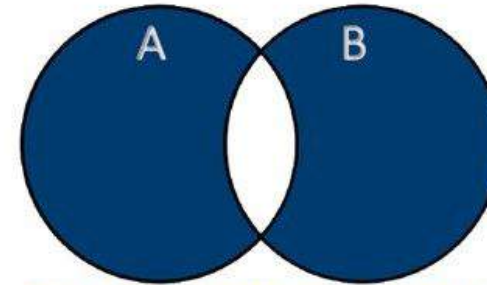
RIGHT INCLUSIVE



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

Organise your deliverable - Build a prioritised list of recommendations

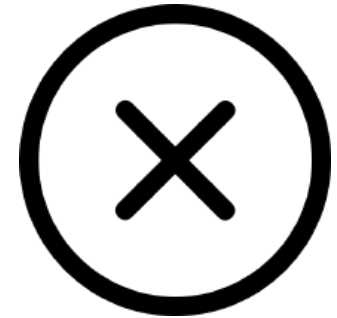
goals at this stage are pretty simple



label as much as possible



see what works



see what fails

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 au	https://www.distilled.net/technical-audit-seo-audit/	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

 @lazarinastoy | #brightonSEO

www.mlforseo.com

	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/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	-

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**





Mode Configuration Bulk Export Reports Sitemaps Visualisations Crawl Analysis Licence Help

ming

Spider

Content

robots.txt

URL Rewriting

CDNs

Include

Exclude

Speed

User-Agent

HTTP Header

Custom

API Access

Authentication

Segments

Crawl Analysis

Profiles

Enter URL to spider

URL Page Titles Meta Description Meta Keywords H1 H2 Content Images Canonicals Page

Export

Address	Content Type	Status Code	Sta
---------	--------------	-------------	-----

AutoML

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

The screenshot shows an Amazon product page for a Kohler toilet. Several elements are highlighted with blue bounding boxes and labeled with AutoML predictions:

- main product image**: A large image of a white toilet, labeled with a confidence of 90.14%.
- product name**: The product title, "11-0 Santa Rosa Comfort Elongated 1.6 GPF Toilet with AquaPiston Flush Technology and Left-Hand Trip Lever, White", labeled with 94.47%.
- product description**: The detailed description text, labeled with 98.80%.
- additional product images**: A row of smaller thumbnail images below the main product image, labeled with 99.40%.

Other visible elements on the page include the Amazon logo, search bar, delivery location (Netherlands), and various navigation links.

Target intent and update needs

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

```

=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|risk|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

```

URL	Content Type	Content length	Target intent
https://lazarinastoy.com/	LAZARINA STOY - SEO, Analytics & Storytelling		Unclassified
https://lazarinastoy.com/how-to-translate-content-collections-with-wix/	How to Translate Content Collections with Wix - LAZARINA		Informational
https://lazarinastoy.com/consulting/	Consulting - LAZARINA STOY.		Commercial
https://lazarinastoy.com/resources/courses/	Courses - LAZARINA STOY.		Unclassified
https://lazarinastoy.com/analytics/	Analytics - LAZARINA STOY.		Unclassified
https://lazarinastoy.com/data-science/	Data Science - LAZARINA STOY.		Unclassified
https://lazarinastoy.com/contact/	Get in touch. - LAZARINA STOY.		Unclassified
https://lazarinastoy.com/resources/google-looker-studio-dashboard-templates/	Google Looker Studio Dashboard Templates - LAZARINA S		Informational
https://lazarinastoy.com/resources/	Resources - LAZARINA STOY.		Informational
https://lazarinastoy.com/resources/google-sheets-templates/	Google Sheets Templates - LAZARINA STOY.		Informational
https://lazarinastoy.com/10-common-data-storytelling-mistakes-and-how-to-avoid-them/	10 Common Data Storytelling Mistakes and How to Avoid Th		Informational
https://lazarinastoy.com/about/	Who is Lazarina Stoy? - LAZARINA STOY.		Informational
https://lazarinastoy.com/resources/conference-decks-and-presentations/	Conference Decks and Presentations - LAZARINA STOY.		Unclassified
https://lazarinastoy.com/seo/	SEO - LAZARINA STOY.		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(IF(S(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|risk|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
fx 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|payment|finance|lease|rent|rental|subscribe|subscription|free|trial|book|reserve|availability|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-cu	3.983786914	120	1727	Informational
3	lazarina stoy	https://lazarinastoy.com/the-ultimate-guide-to-cu	10.666666667	3	127	Unclassified
4	google trends api	https://lazarinastoy.com/the-ultimate-guide-to-cu	10.12707499	22	1747	Informational
5	pytrends example	https://lazarinastoy.com/the-ultimate-guide-to-cu	2.470588235	16	51	Informational
6	pytrend	https://lazarinastoy.com/the-ultimate-guide-to-cu	10.12707499	12	189	Unclassified
7	google trends api python	https://lazarinastoy.com/the-ultimate-guide-to-cu	5.099337748	12	302	Informational
8	looker studio custom dimension	https://lazarinastoy.com/ultimate-guide-to-cu	3.8125	8	32	Unclassified
9	seo google sheets templates	https://lazarinastoy.com/the-ultimate-guide-to-cu	10.666666667	17	34	Informational

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

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



Mark Williams-Cook  @thetafferboy · Sep 16



Question: I have 100M rows of query data, each query has been assigned a category. If I wanted to build a classifier trained on this data, that means you could enter a query and it would classify it into one of those categories, based on that data - how would you go about it?



Lazarina Stoy
@lazarinastoy



Hi Mark 100 shared sentiment!

Look into custom training with AutoML Natural Language
youtube.com/watch?v=ieaqfU...



Look into costs.

Otherwise, look into training your own text classifier for short-form text (supervised machine learning models - e.g. naive Bayes, decision tree,..)



youtube.com

Using AutoML Natural Language for custom text classificati
In this episode of AI Adventures, Yufeng introduces how to use AutoML Natural Language to perform custom text ...

Vertex AI > Documentation Was this helpful?  

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.

Vertex AI > Documentation Was this helpful?  

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](#)

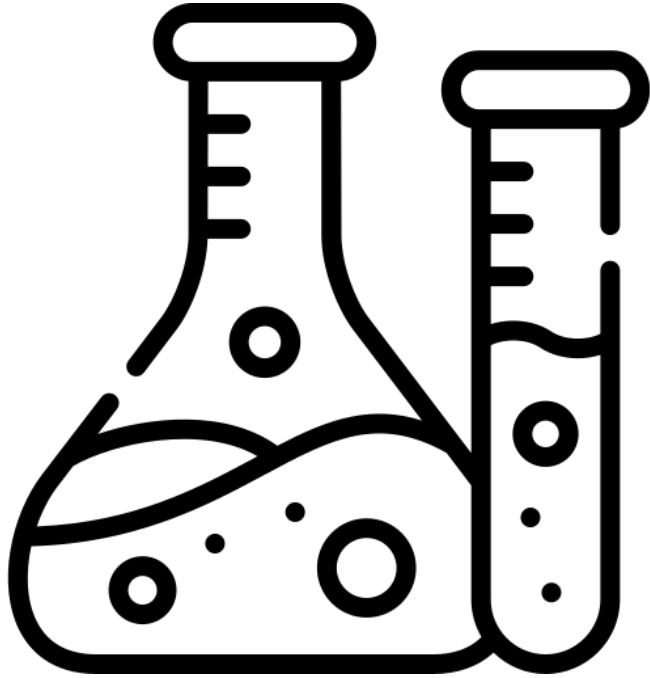
Use the Vertex AI console to create a text classification dataset. After your dataset is created, use the CSV that you copied into your Cloud Storage bucket to import those documents into the dataset.

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.

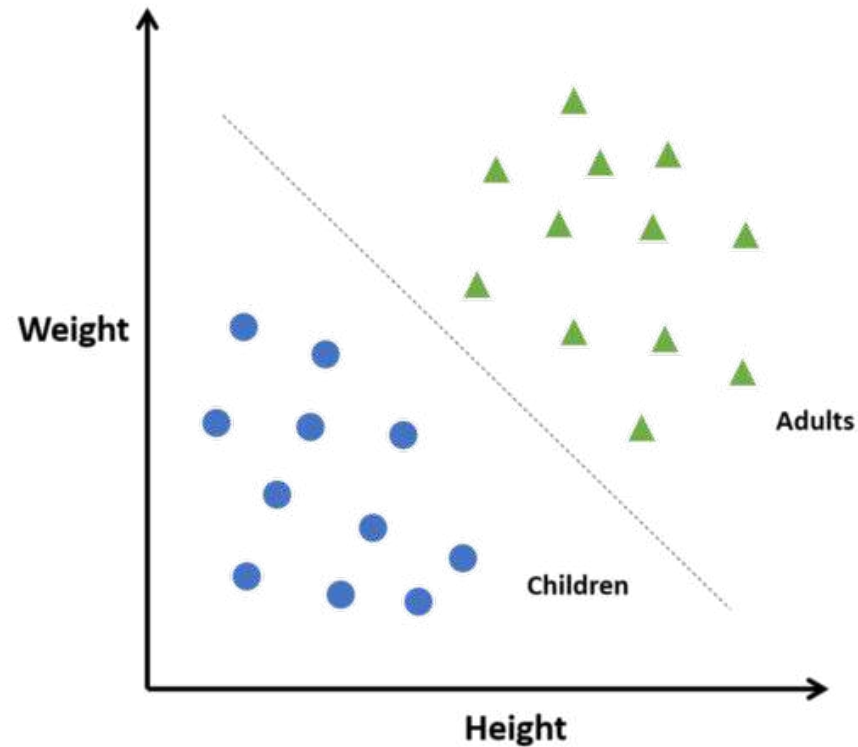
1. Go to the [Vertex AI console](#).
2. From the **Get started with Vertex AI** page, click **Create dataset**.
3. Specify details about your dataset.
 - a. Specify a name for this dataset, such as `text_classification_tutorial`.



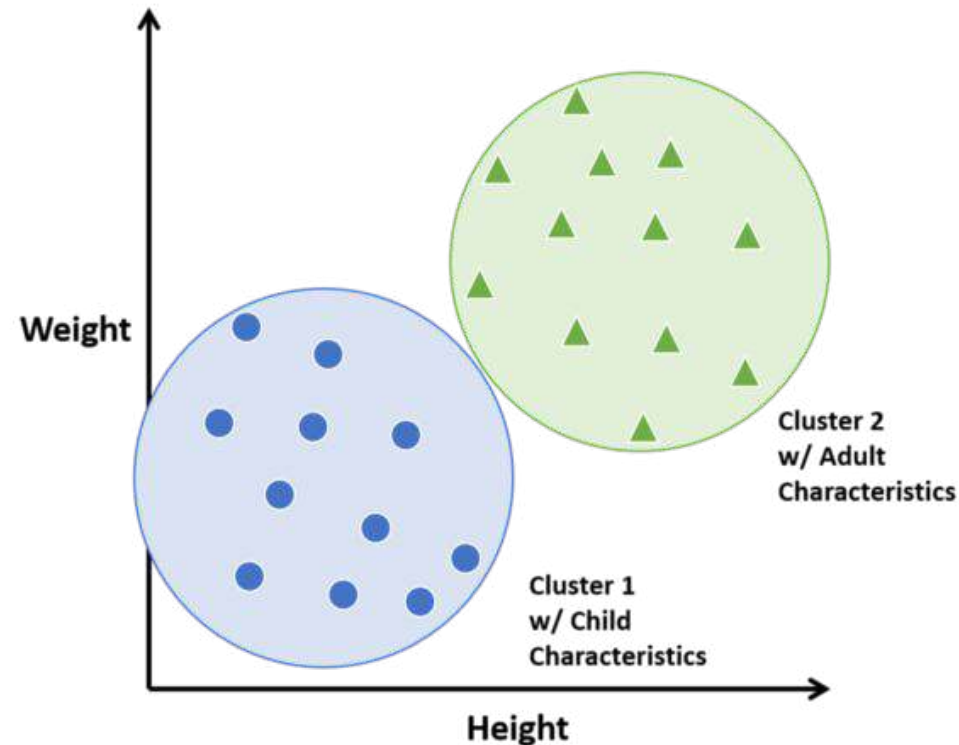
- Incorporate signals from **SERP features**
- Tag **user knowledge level** in queries ranked
- Pinpoint **user action wanted**, JTBD, implicit intent

Understand content topics and subtopics

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

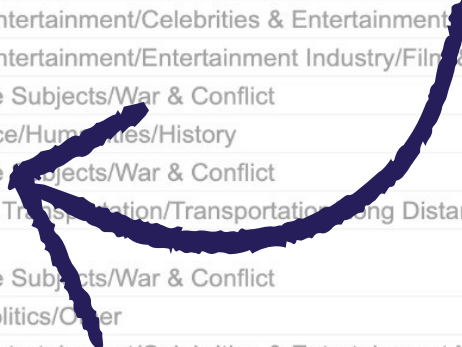
Process will take no more than 20 minutes



C2 /Sensitive Subjects/War & Conflict

	A	B	C	D
1	URL	Content	Classification	Score
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-68362810	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-68395354	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-68395173	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 Ukrainian	/Sensitive Subjects/War & Conflict	0.96695495
12	https://www.bbc.com/news/world-europe-68393412	Authorities return body of Alexei Navalny to mother 8 d	/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 an	/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-68317736	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



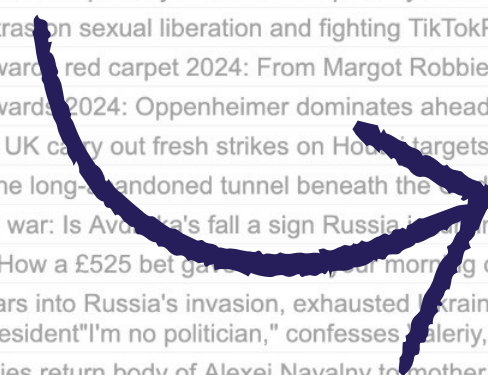


C2 /Sensitive Subjects/War & Conflict

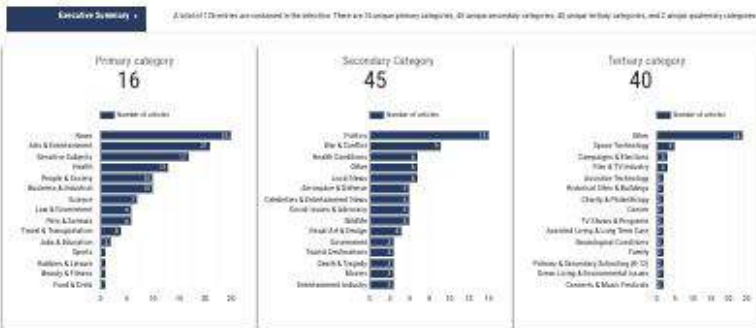
Run the script via a formula to get classification label & confidence

	A	B
1	URL	Content
2	https://www.bbc.com/news/world-us-canada-68395414	South Carolina primary: Donald Trump easily defeats...
3	https://www.bbc.com/news/world-us-canada-68395414	South Carolina primary: Donald Trump easily defeats...
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...
6	https://www.bbc.com/news/entertainment-arts-68395355	SAG Awards 2024: Oppenheimer dominates ahead of...
7	https://www.bbc.com/news/world-middle-east-68395173	US and UK carry out fresh strikes on Hamas targets in...
8	https://www.bbc.com/news/uk-scotland-glasgow-west-68395173	Inside the long-abandoned tunnel beneath the Glasgow F...
9	https://www.bbc.com/news/world-europe-68322527	Ukraine war: Is Avdiivka's fall a sign Russia is making...
10	https://www.bbc.com/news/uk-wales-68210255	Travel: How a £525 bet gave me a free trip to my morri...
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	Authorities return body of Alexei Navalny to mother 8...
13	https://www.bbc.com/news/entertainment-arts-68391330	Wendy Williams thanks fans for support after dementia...
14	https://www.bbc.com/news/world-asia-68378651	Japan naked festival: Women join Hadaka Matsuri for...
15	https://www.bbc.com/news/world-europe-68395030	Alexei Navalny: Dissent is dangerous in Russia, but ac...
16	https://www.bbc.com/news/world-europe-68359252	Rosenberg: How two years of war in Ukraine changed...
17	https://www.bbc.com/news/entertainment-arts-68395352	SAG Award winners 2024: The full list of nominees and...
18	https://www.bbc.com/news/entertainment-arts-68362811	Stray Kids: How K-Pop took over the global charts in 20...
19	https://www.bbc.com/news/entertainment-arts-68317736	Gareth Edwards: The Creator director on shaking up M...
20	https://www.bbc.com/news/newsbeat-68382142	Chuckie: 1Xtra presenter feels R&B has special year...

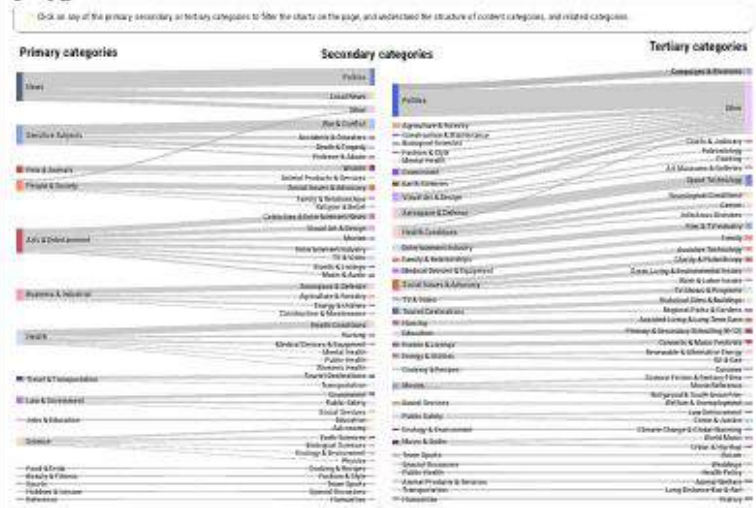
C	D
Classification Label	Confidence
/Sensitive Subjects/War & Conflict	0.9567586
/News/Politics/Campaigns & Elections	0.97098255
/Arts & Entertainment/Celebrities & Entertainment News	0.73509616
/Arts & Entertainment/Celebrities & Entertainment News	0.96840936
/Arts & Entertainment/Entertainment Industry/Film & 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 & Ra	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



at a glance



playground



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

URL: Equals

Content: Equals

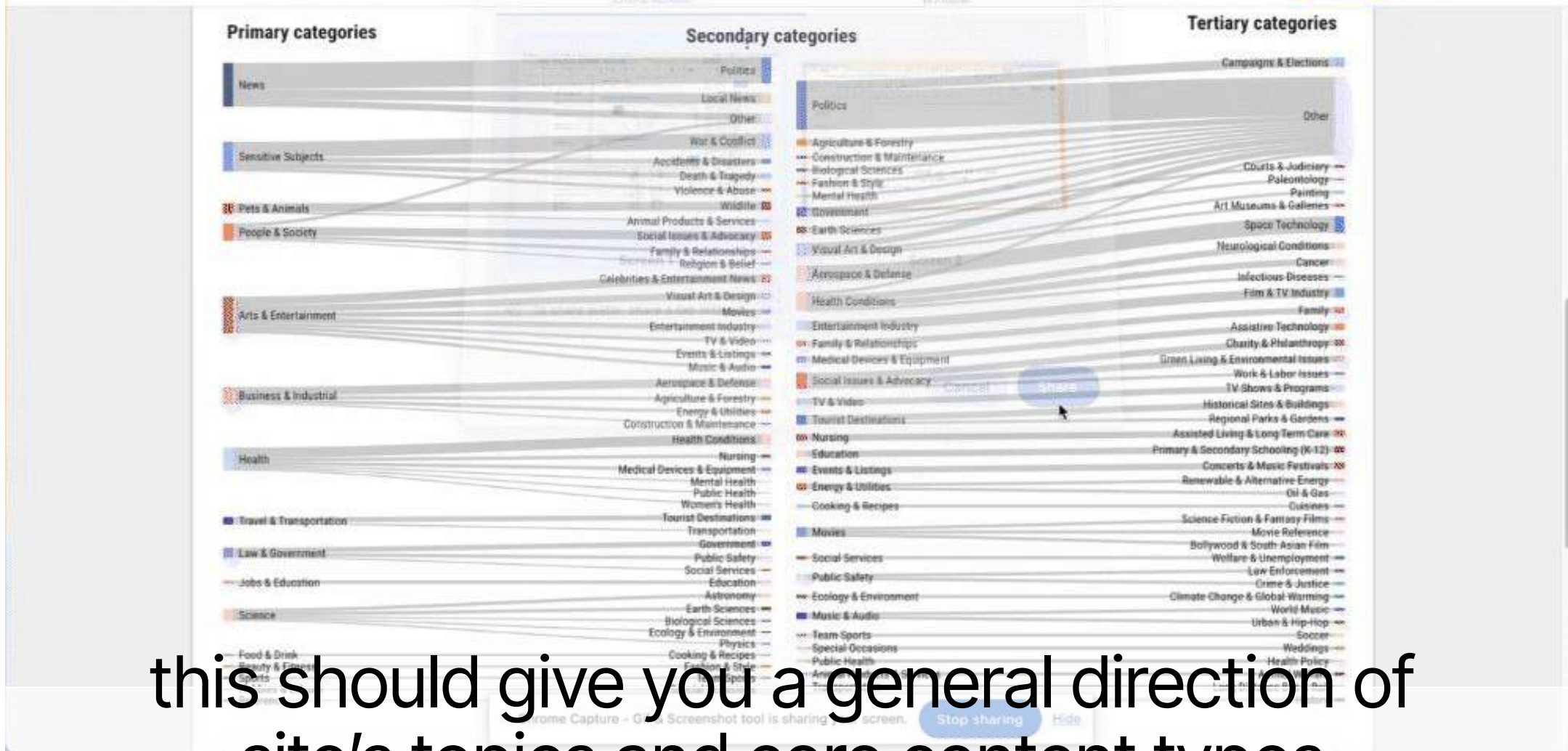
Classification Label: Equals

Confidence:

URL	Classification Label	Confidence	Primary category	Secondary category	Tertiary category	Quaternary category
https://www.fox.com/news/us/politics/2018/02/02/180202-1	News, Politics, Campaigns & Elections	91%	News	Politics	Campaigns & Elections	null
https://www.fox.com/news/us/politics/2018/02/02/180202-2	News, Politics & Elections	85%	News	Politics	Other	null
https://www.fox.com/news/us/politics/2018/02/02/180202-3	News, Politics, Campaigns & Elections	93%	News	Politics	Campaigns & Elections	null
https://www.fox.com/news/us/politics/2018/02/02/180202-4	Demotivator Subjects, War & Conflict	100%	Demotivator Subjects	War & Conflict	null	null
https://www.fox.com/news/us/politics/2018/02/02/180202-5	News & Entertainment, Travel, Art & Design	42%	News & Entertainment	Travel, Art & Design	Traveling	null
https://www.fox.com/news/us/politics/2018/02/02/180202-6	News, Politics & Elections	100%	News	Politics	Other	null
https://www.fox.com/news/us/politics/2018/02/02/180202-7	News, Politics & Elections	85%	News	Politics	Other	null
https://www.fox.com/news/us/politics/2018/02/02/180202-8	Demotivator Subjects, War & Conflict	91%	Demotivator Subjects	War & Conflict	null	null
https://www.fox.com/news/us/politics/2018/02/02/180202-9	Demotivator Subjects, War & Conflict	85%	Demotivator Subjects	War & Conflict	null	null
https://www.fox.com/news/us/politics/2018/02/02/180202-10	Demotivator Subjects, War & Conflict	96%	Demotivator Subjects	War & Conflict	null	null

Plug-and-play template in Looker Studio





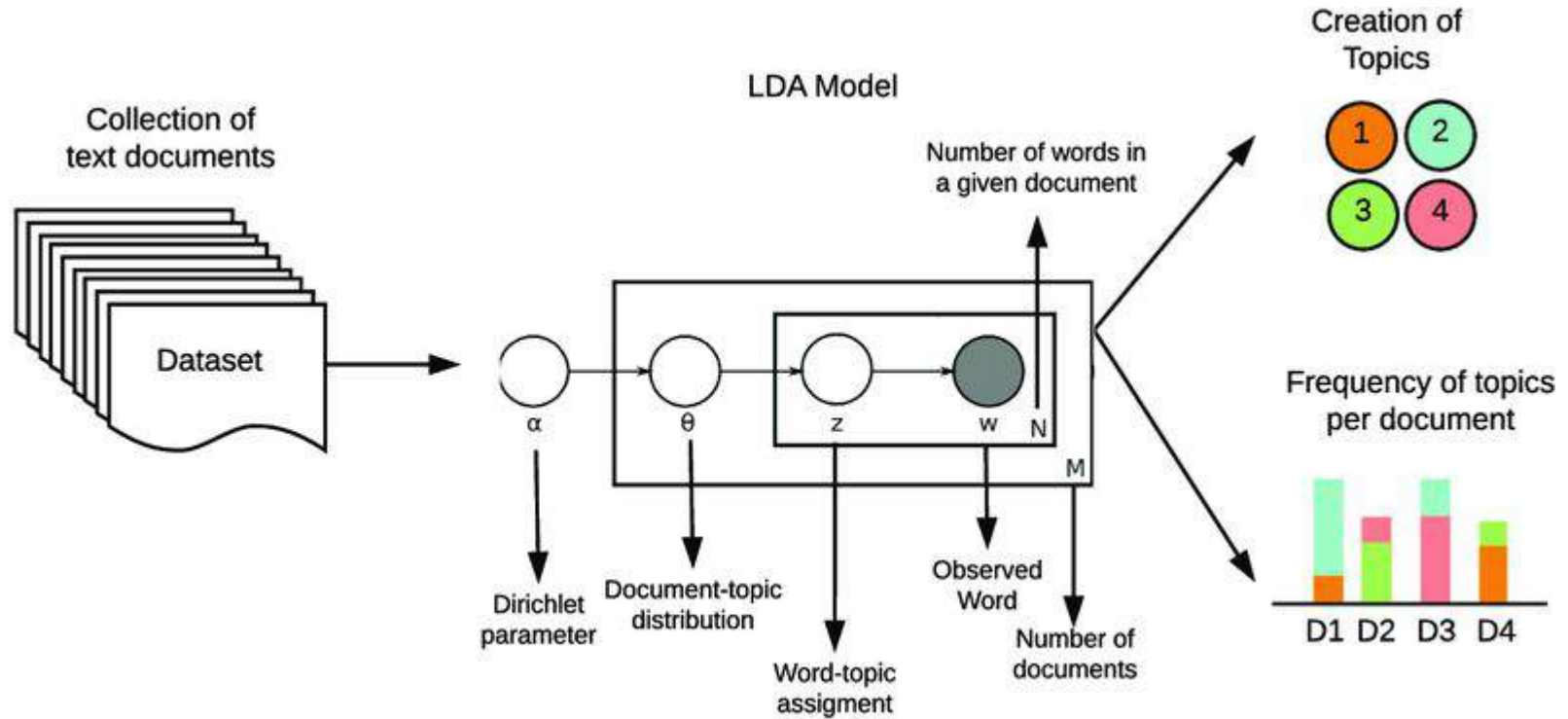
this should give you a general direction of site's topics and core content types



- ✓ 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

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.

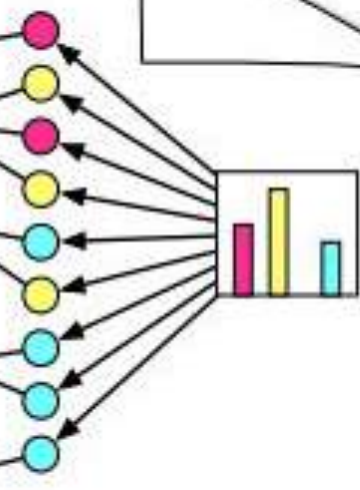
Although the numbers don't match precisely, those predictions

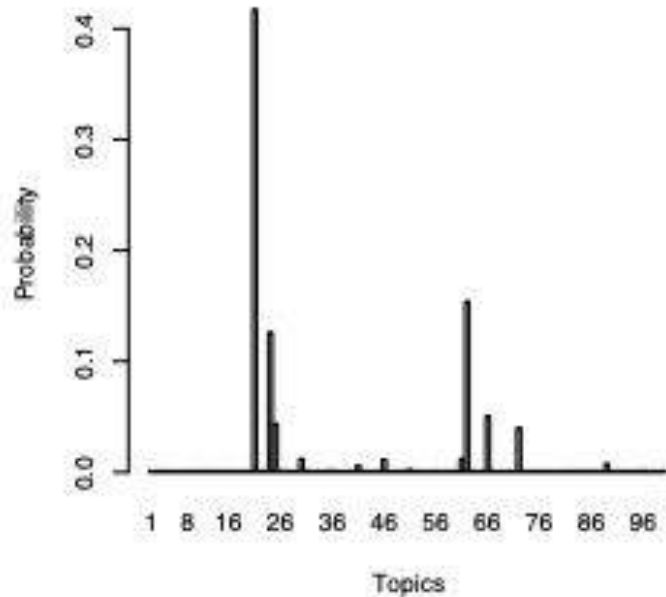
are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson, a biochemist at Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a numbers game, particularly as more and more genomes 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

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

SCIENCE • VOL. 272 • 24 MAY 1996

Topic proportions and assignments





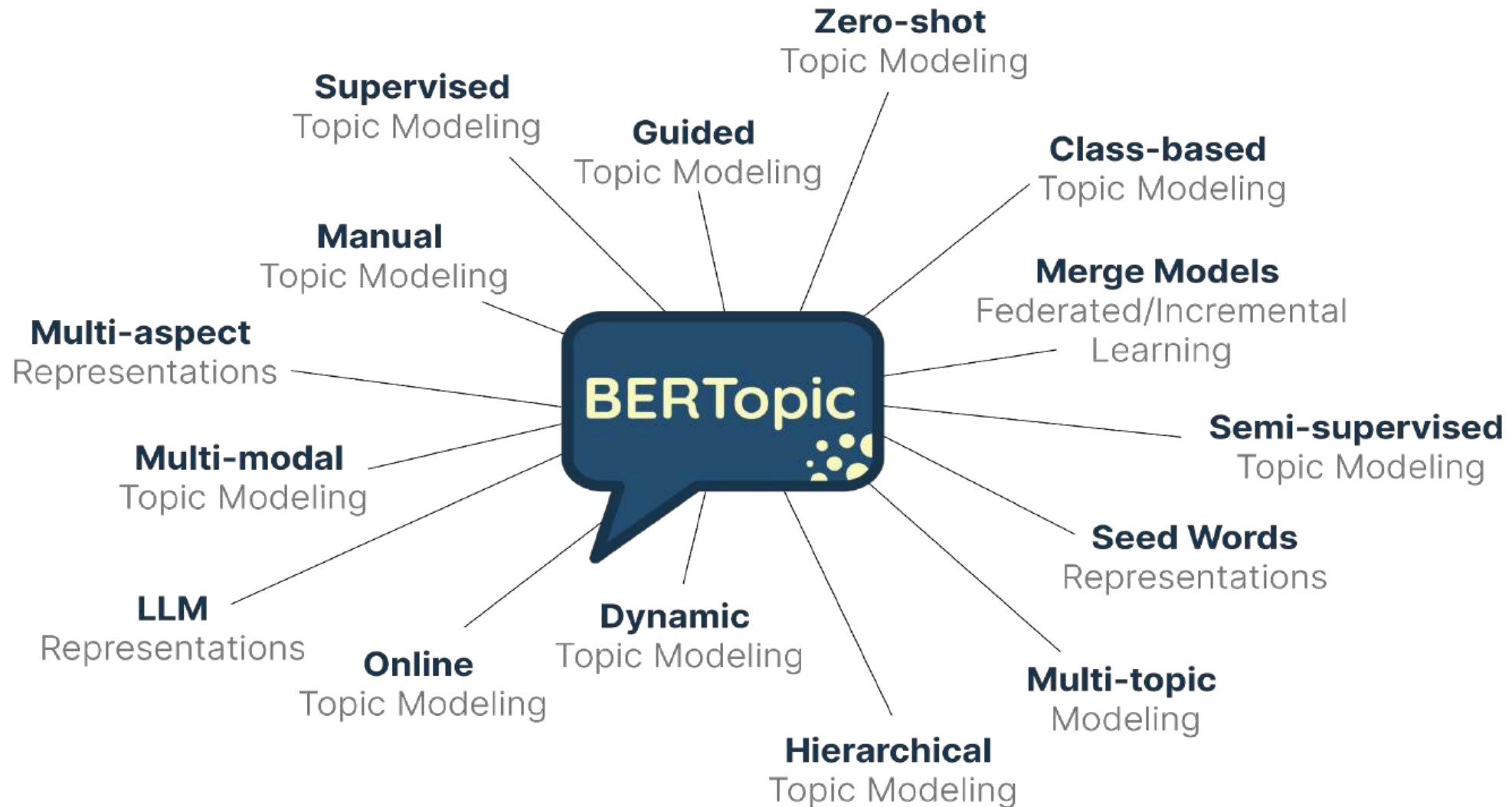
“Genetics”
 human
 genome
 dna
 genetic
 genes
 sequence
 gene
 molecular
 sequencing
 map
 information
 genetics
 mapping
 project
 sequences

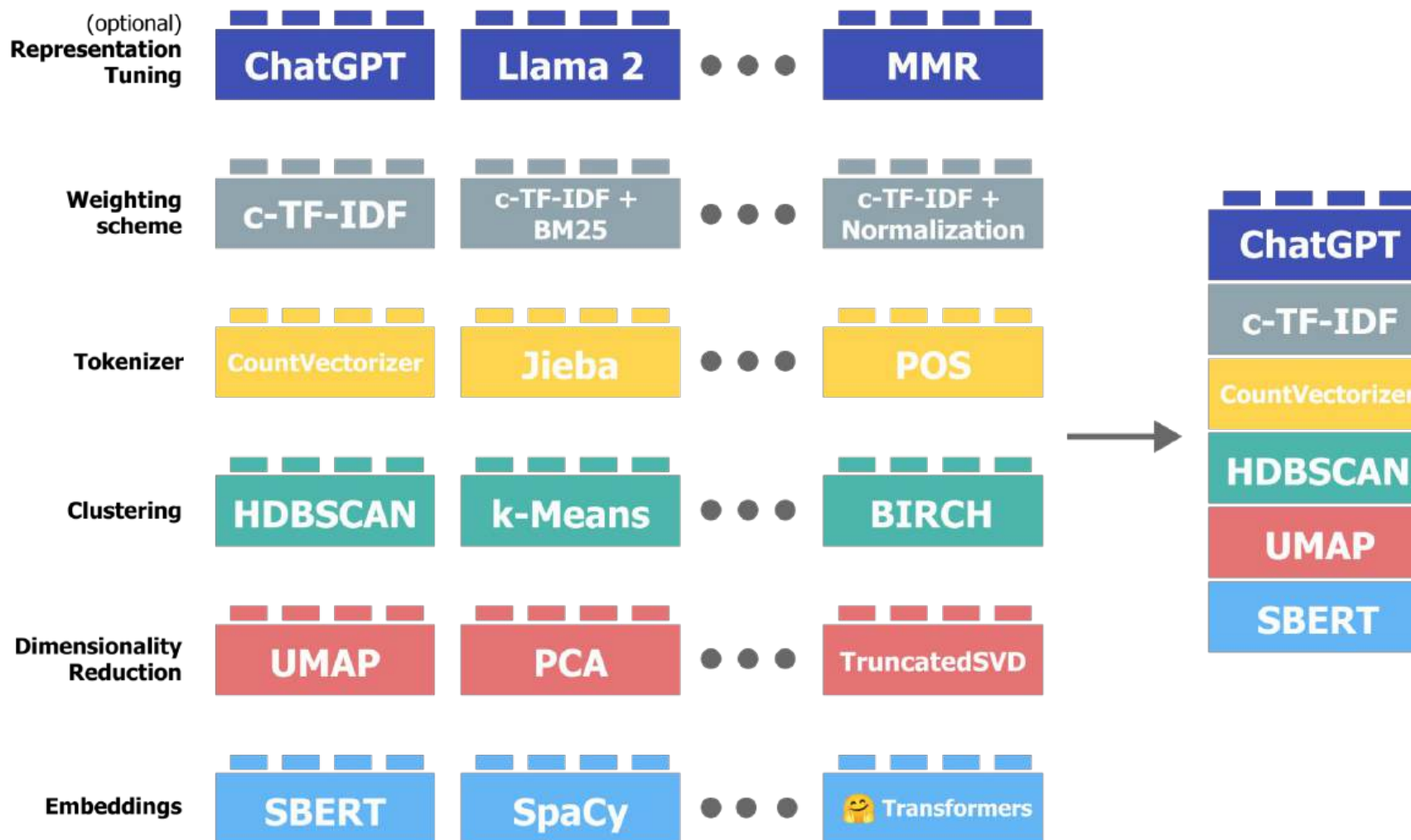
“Evolution”
 evolution
 evolutionary
 species
 organisms
 life
 origin
 biology
 groups
 phylogenetic
 living
 diversity
 group
 new
 two
 common

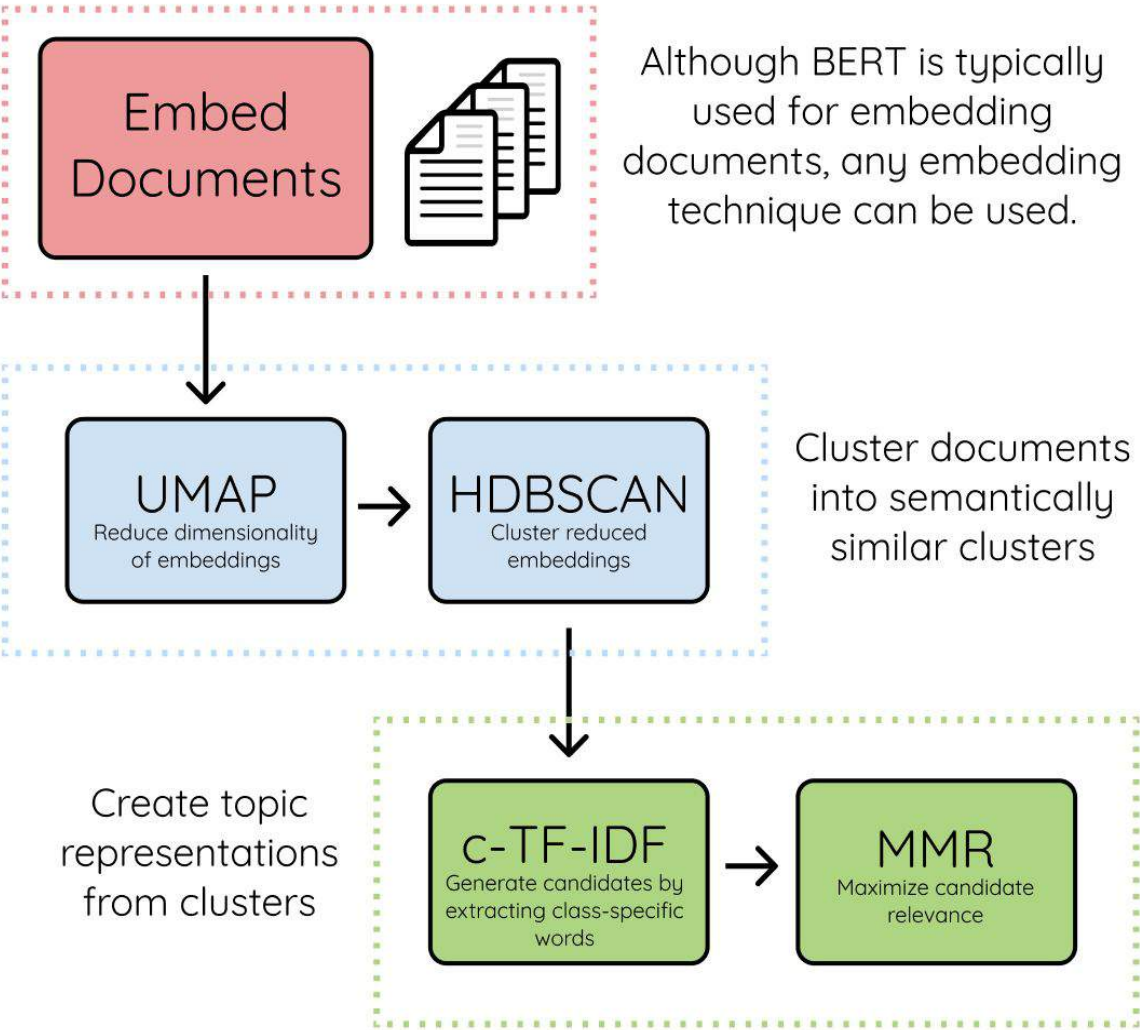
“Disease”
 disease
 host
 bacteria
 diseases
 resistance
 bacterial
 new
 strains
 control
 infectious
 malaria
 parasite
 parasites
 united
 tuberculosis

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

BERTopic is another, even better alternative

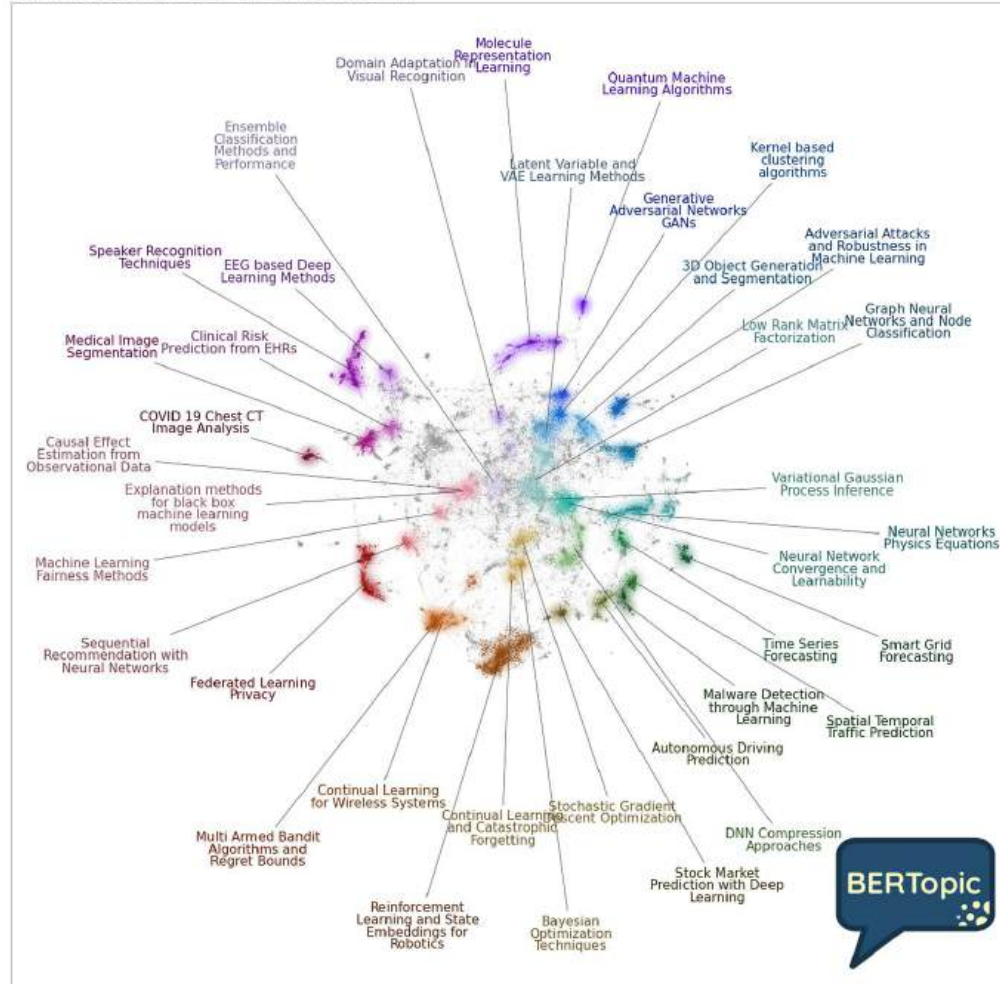




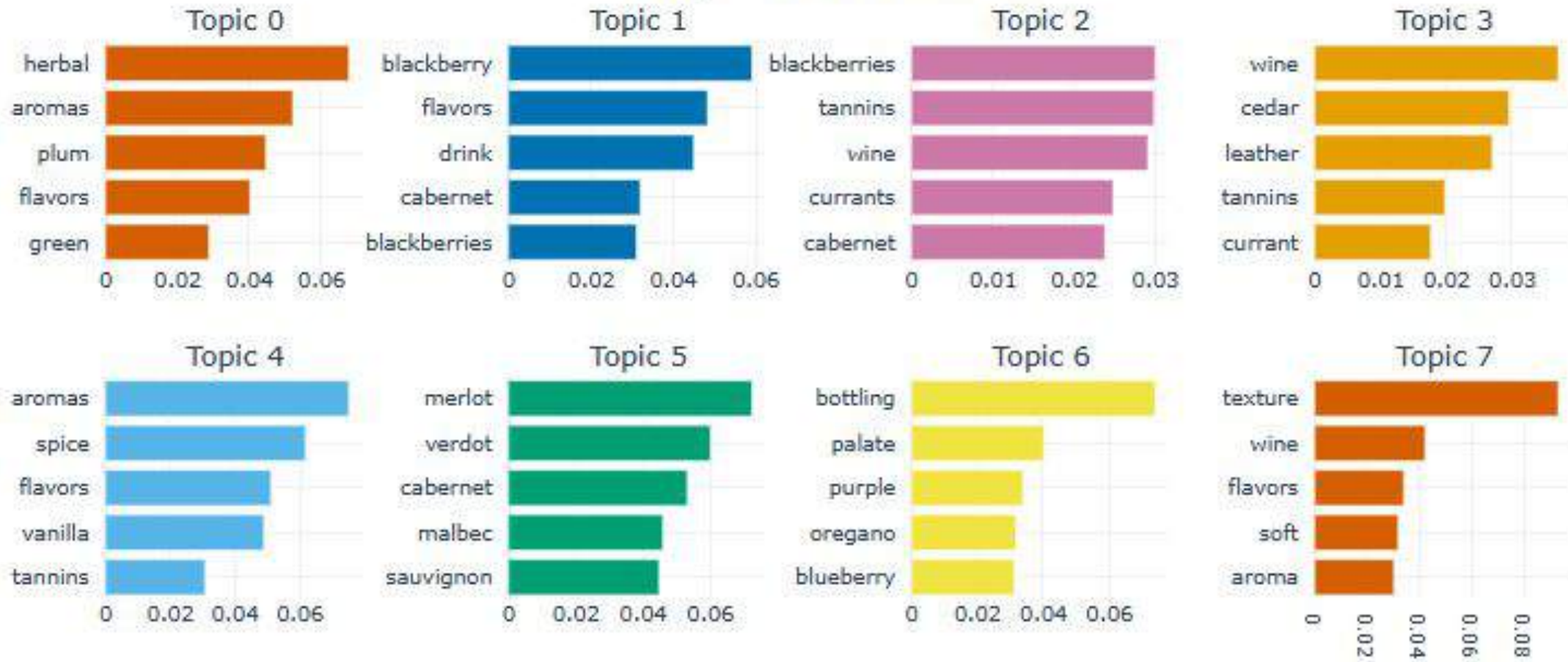


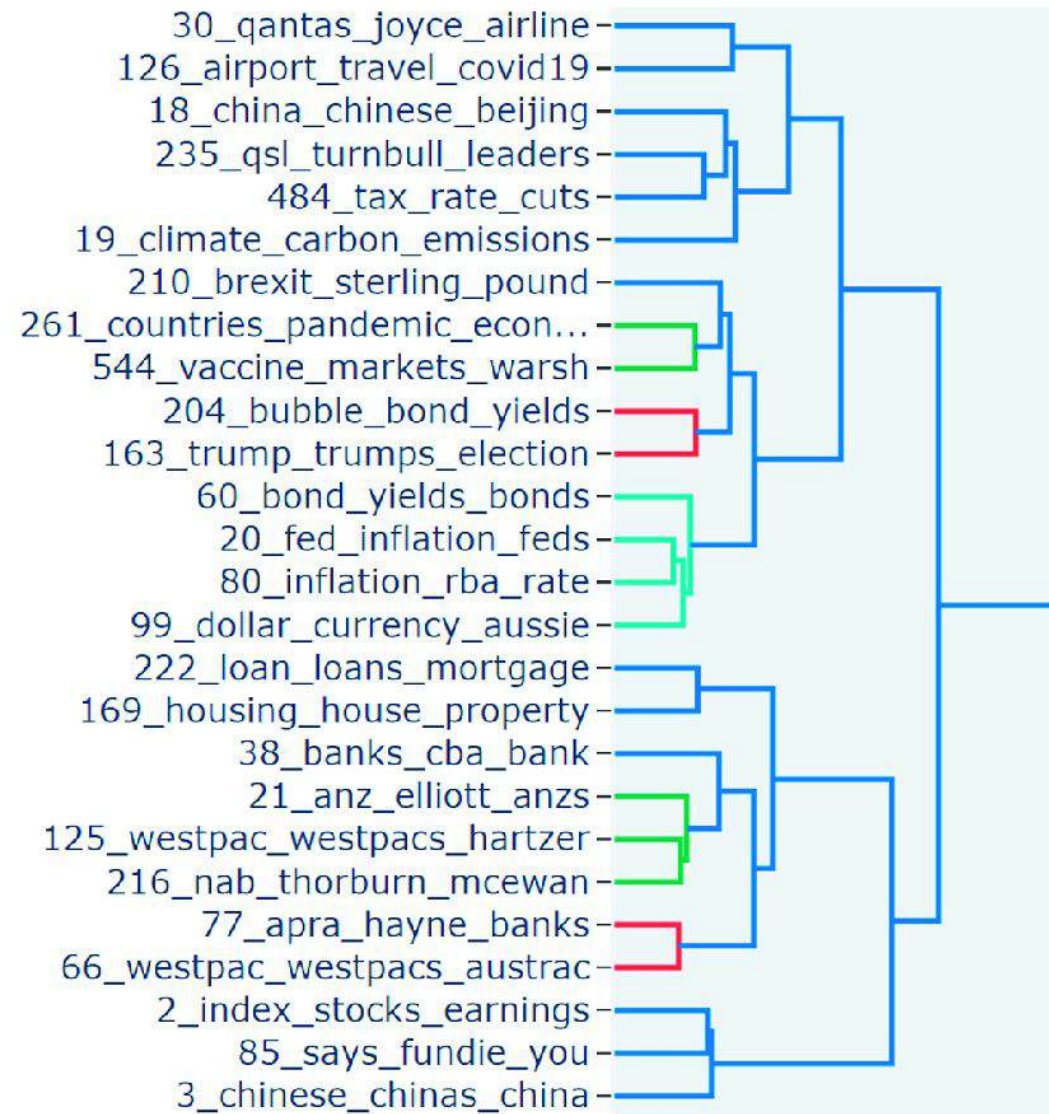
ArXiv - BERTopic

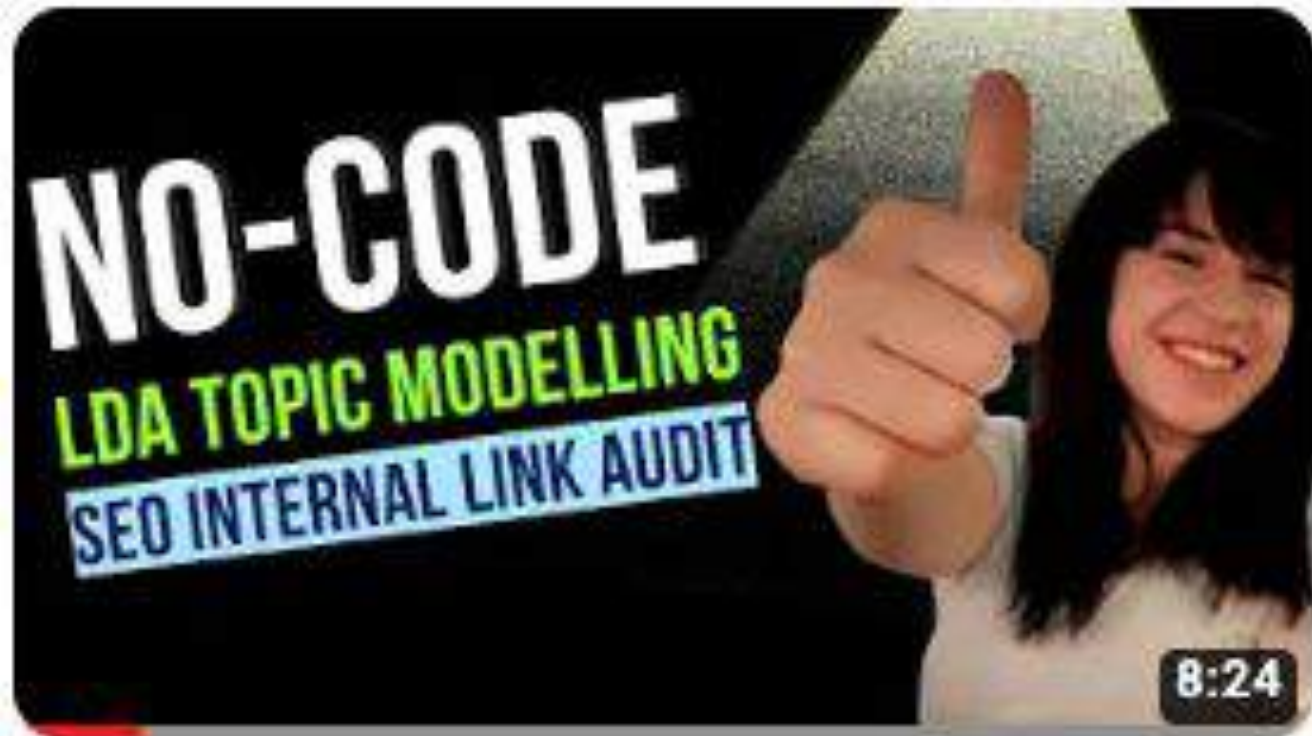
Topics labeled with 'openhermes-2.5-mistral-7b'



Topic Word Scores







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

Another copy of BERTopic.ipynb

File Edit View Insert Runtime Tools Help Last edited on 20 March

Table of contents

- Tutorial - Topic Modeling with BERTopic
- Enabling the GPU
- Installing BERTopic
- Restart the Notebook**
- Authenticate and Access the Sheet
- Data cleaning & Stopwords removal
- Topic Modeling
 - Training
 - Extracting Topics
 - Attributes
- Visualization
 - Visualize Topics
 - Visualize Topic Probabilities
 - Visualize Topic Hierarchy
 - Visualize Terms
 - Visualize Topic Similarity
 - Visualize Term Score Decline
- Topic Representation
 - Update Topics
 - Topic Reduction
- Search Topics
- Model serialization
- Embedding Models
 - Sentence-Transformers

+ Code + Text


Tutorial - Topic Modeling with BERTopic

(last updated 01-09-2022)

In this tutorial we will be exploring how to use BERTopic to create topics from the well-known 20Newsgroups dataset. The most frequent use-cases and methods are discussed together with important parameters to keep a look out for.

BERTopic

BERTopic is a topic modeling technique that leverages 🧠 transformers and a custom class-based TF-IDF to create dense clusters allowing for easily interpretable topics whilst keeping important words in the topic descriptions.



Enabling the GPU

First, you'll need to enable GPUs for the notebook:

- Navigate to Edit → Notebook Settings
- select GPU from the Hardware Accelerator drop-down

[Reference](#)

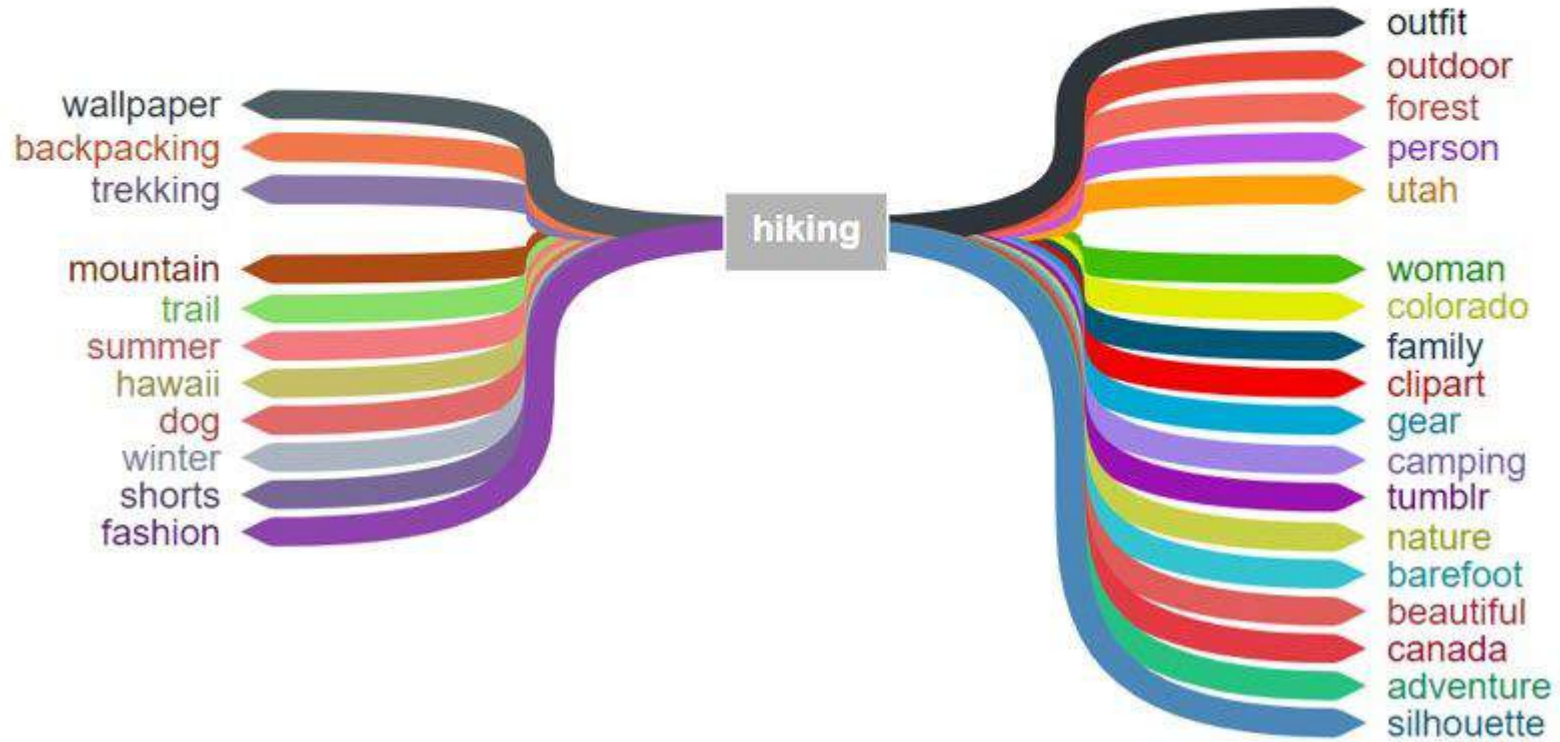
Installing BERTopic

We start by installing BERTopic from PyPi:

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
	The blind Ukrainian amputee whose wife's voice ke	complete
	Exhausted Ukraine struggles to find new men for fr	complete
	South Carolina primary: Donald Trump easily defea	complete
	Kim Petras on sexual liberation and fighting TikTok	complete
	SAG Awards red carpet 2024: From Margot Robbie	complete
	SAG Awards 2024: Op... mates ahead of Oscars	Published8 hours a
	US and UK carry out fresh strikes...outhi targets in Yemen	Published12 hours a
	Inside the long-abandoned tunnel beneath the Clyde	Published2 hours agoShare
	Ukraine war: Is Avdiivka's fall...ign...ussia is turning the tide?	Published17 Febru
	Travel: How a £525 bet...birth to...our morning commute	Published4 hours ago
	Two years into Russian invasion, exhausted Ukrainians refuse to give up	Publishe
	Authorities return body of Alexei Navalny to mother 8 days after death	Published1:
	...ains thanks fans for support after dementia and aphasia diagnosis	Put
	Japan naked festival: Women join Hadaka Matsuri for first time	Published10 hours
	Alexei Navalny: Dissent is dangerous in Russia, but activists refuse to give up	Put
	Rosenberg: How two years of war in Ukraine changed Russia	Published3 days ag
	SAG Award winners 2024: The full list of nominees and	
	Stray Kids: How K-Pop took over the global charts in 20	
	Gareth Edwards: The Creator director on shaking up Ho	
	Chuckie: 1Xtra presenter feels R&B has special year ah	
	Alia Bhatt: The young Bollywood star taking on Hollywoo	

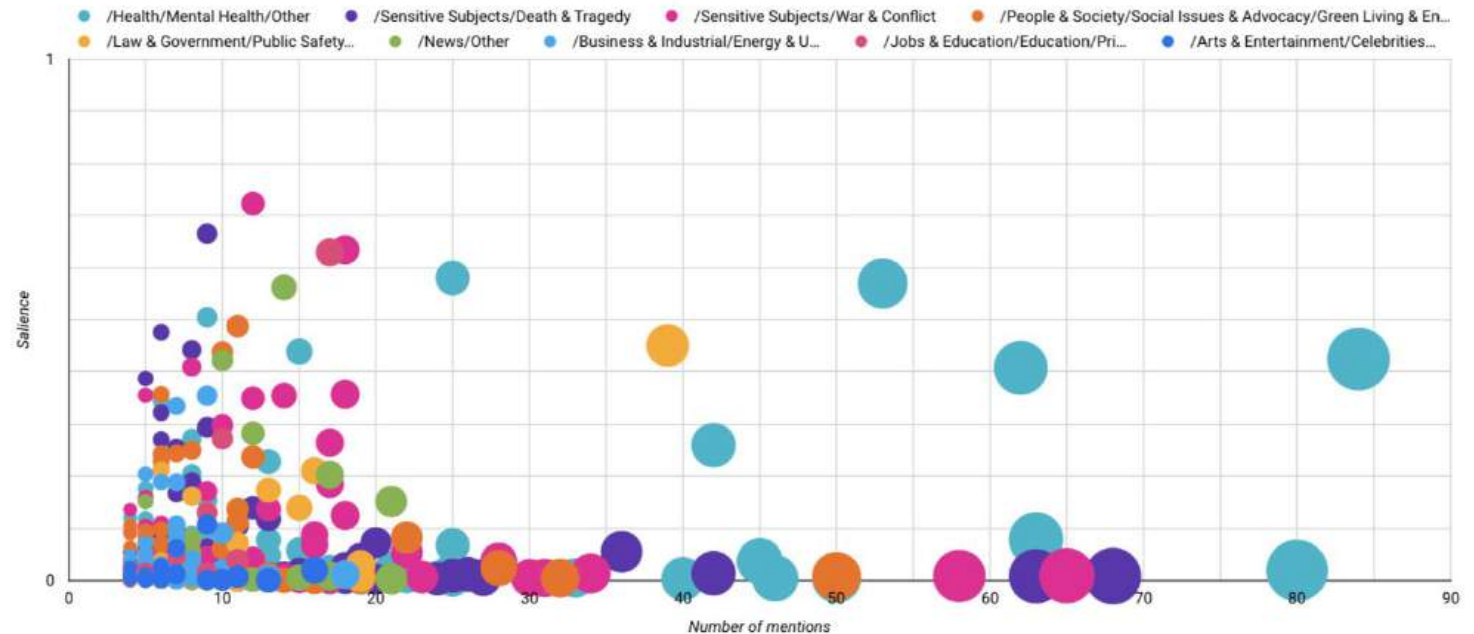
ID	Entity	Type	Salience	Sentiment Score	Sentiment Magnitud	Number of menti	Metadata	Mentions
https://www.bbc.com/news/world-europe-68376700	Serhiy	PERSON	0.34382942	0	2.6	6		Serhiy, Serhiy, Serhiy, Serhiy, Serhiy, Sei
https://www.bbc.com/news/world-europe-68376700	Valeria	PERSON	0.20473818	0	1.9	8		wife, Valeria, Valeria, Valeria, Valeria, Va
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	0.010182	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 Serf	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","wikip	Ukrainian, Ukrainian, Ukraine, U
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	...sia, 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://e	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



Classification label	Entity / Number of mentions									
	issian	UK	Getty Imag...	one	world	family	way	video	support	pag
/Health/Mental Health/Other	80	19	9	16	10	17	17	22	11	
/Sensitive Subjects/Death & Tragedy	-	26	42	20	27	8	15	12	24	
/Sensitive Subjects/War & Conflict	1	15	10	14	11	23	12	9	16	
/People & Society/Social Issues & Advoc...	1	13	3	8	2	6	10	4	7	
/News/Other	-	2	6	6	15	5	9	11	3	
/Law & Government/Public Safety/Law E...	-	3	2	3	-	-	2	3	-	
/Business & Industrial/Energy & Utilities/...	-	-	1	4	-	6	1	-	3	
/Arts & Entertainment/Celebrities & Ente...	-	1	5	3	4	-	1	3	-	
/Jobs & Education/Education/Primary &...	-	1	1	2	-	3	-	2	-	



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

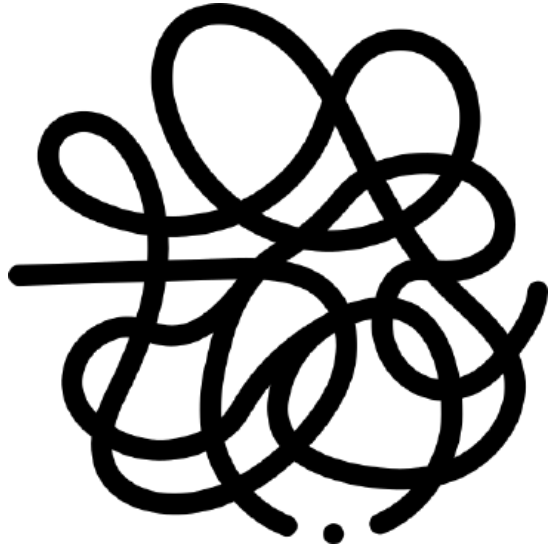
Red - difficult to check at scale



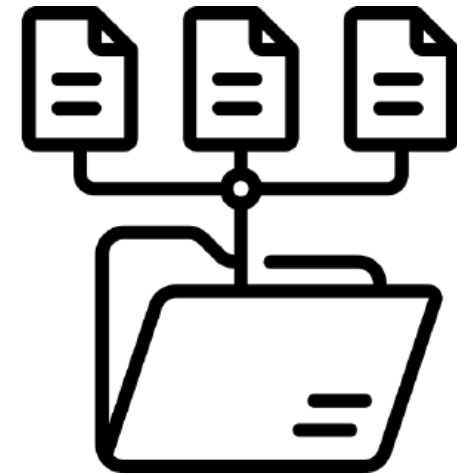
Questions to assess your content Quality, Helpfulness and Experience	Area	Expected Answer
Does the content seem to be serving the genuine interests of visitors to the site or does it seem to exist solely by someone attempting to guess what might rank well in search engines?	Helpfulness and People Satisfaction	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 ✨

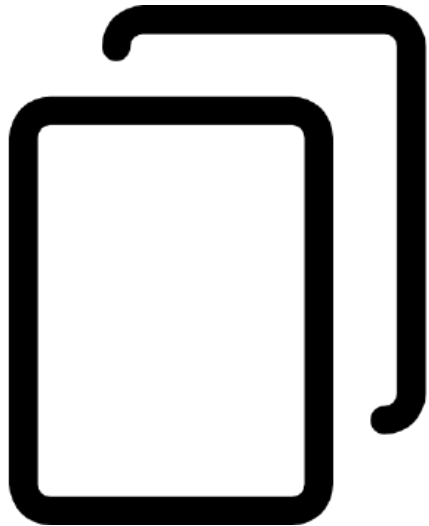
Instead of this...



Do this ...



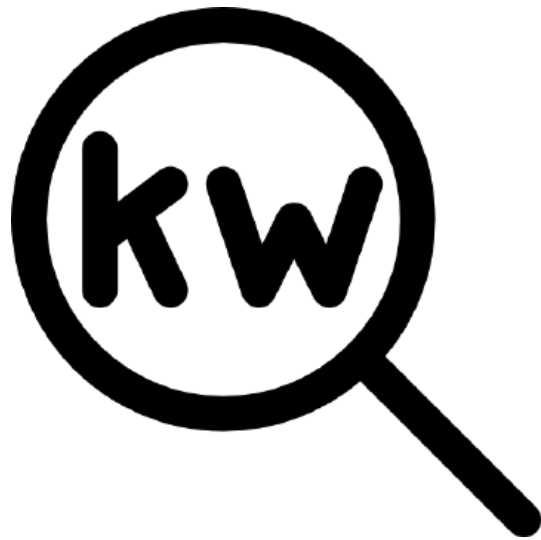
Instead of this...



Do this ...



Instead of this...



Do this ...



Your website Topical Authority

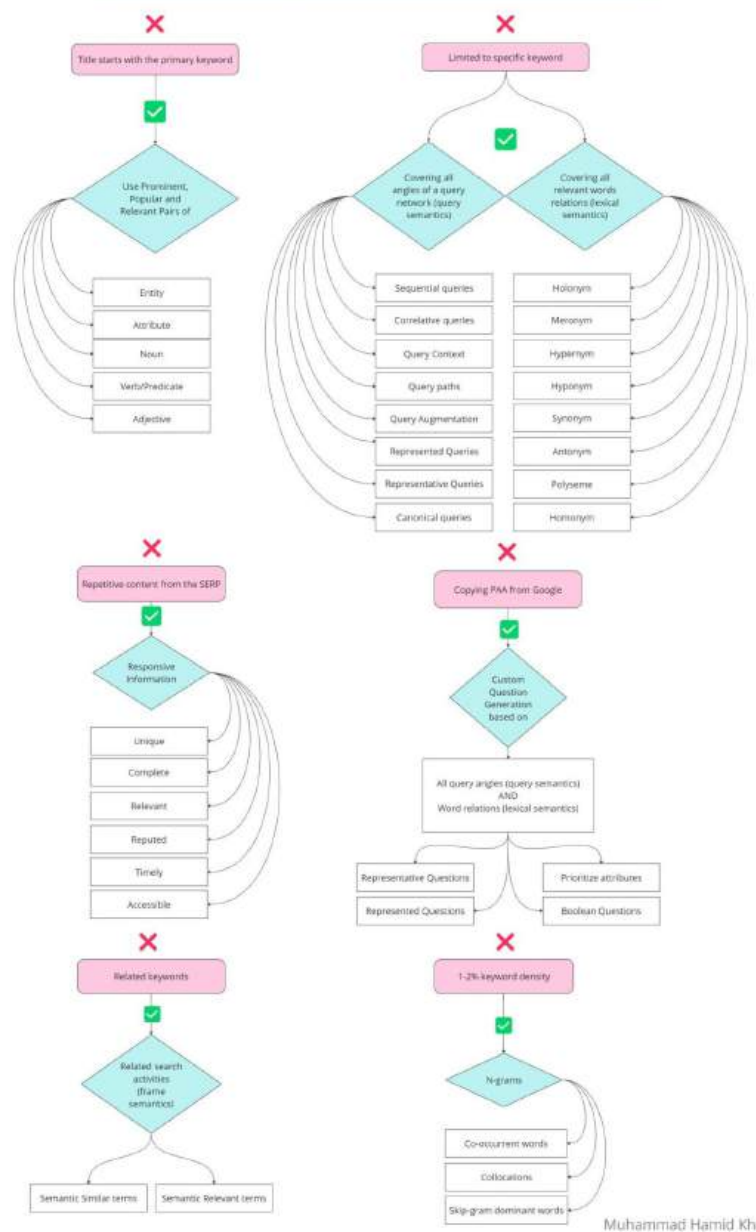


Muhammad Hamid Khan 1st
0 to 300,000+ Visitors / Month through <Semantic SEO>
Lahore, Punjab, Pakistan · Contact info
8,291 followers · 500+ connections
Mihir Naik, Muhammad Qamar Shafique, and 308 other mutual connections

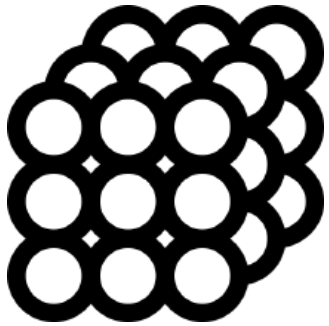
Providing services
Search Engine Optimization (SEO), Growth Marketing, Business Consulting, Strategic Planning, Technical Writin...
Show details



Semantic Content Audit Roadmap



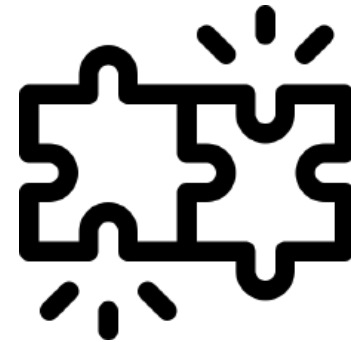
Build custom metrics



keyword density
entity density

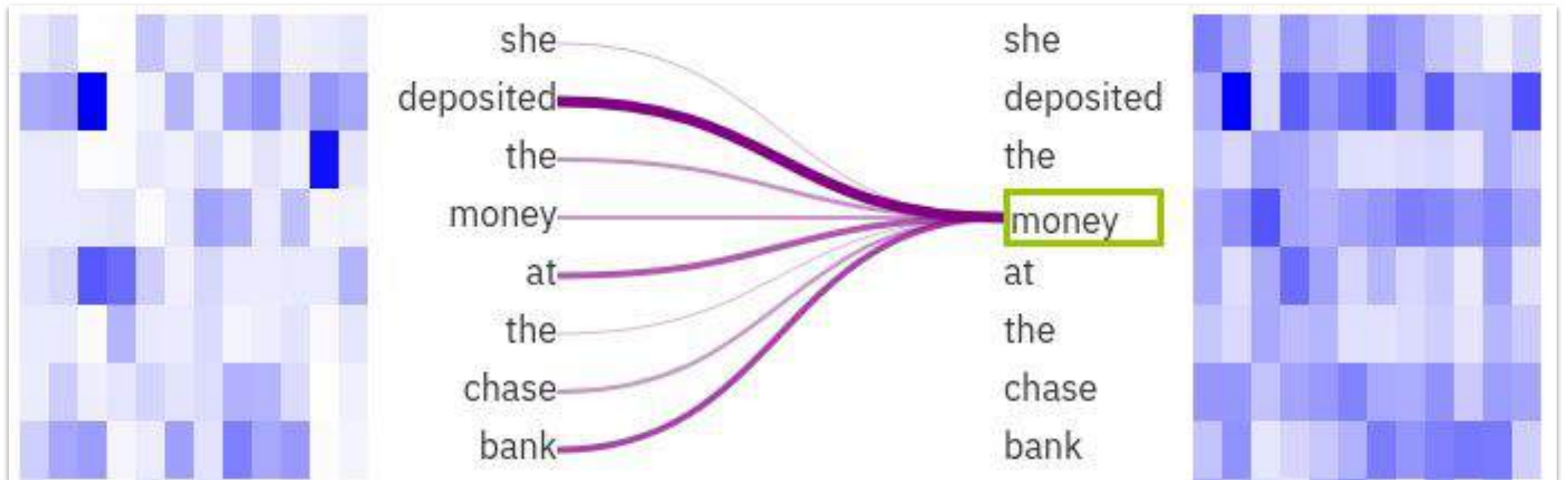


topic relevance



query-content match (see
KeyBert 🧐)

KeyBERT



```
#Basic usage - keyword extraction
```

```
from keybert import KeyBERT
```

```
doc = """
```

```
Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labeled training data consisting of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances. This requires the learning algorithm to generalize from the training data to unseen situations in a 'reasonable' way (see inductive bias).
```

```
"""
```

```
kw_model = KeyBERT()
```

```
keywords = kw_model.extract_keywords(doc)
```



```
#highlight keywords in the document
```

```
keywords = kw_model.extract_keywords(doc, highlight=True)
```

```
Supervised learning is the machine learning task of learning function that maps an input to an output based on example input output pairs It infers function from labeled training data consisting of set of training examples In supervised learning each example is pair consisting of an input object typically vector and desired output value also called the supervisory signal supervised learning algorithm analyzes the training data and produces an inferred function which can be used for mapping new examples An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances This requires the learning algorithm to generalize from the training data to unseen situations in reasonable way see inductive bias
```

Organize your database - crawl, content, metrics

Understand the data - Categorize your inventory

Evaluate performance and assign action items to underperforming content

Organise your deliverable - Build a prioritised list of recommendations

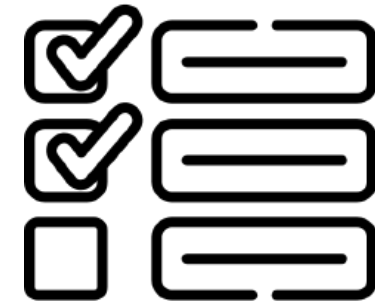
goals at this stage are pretty simple



Simple, adaptable
system



Consistent
evaluation criteria



Forward-looking
(who's next)

If/else logic = your bff at this stage

build a
checklist with
best practices



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 enticingful	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.

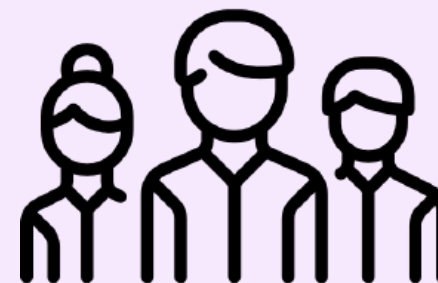


Translate the if/else statements to code for your platform of choice

Each **assessment** should have



a corresponding action item



responsible person/department



Looker Studio dashboards 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

Organise your deliverable - Build a prioritised list of recommendations

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

what a good audit deliverable looks like to me: *you're not auditing for me... so ask your client what it looks like for them*





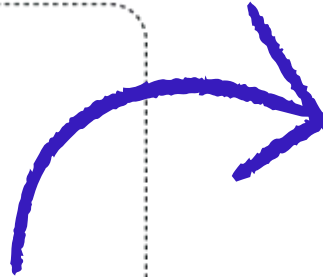
Audit Section



Why it matters



Content Pattern Goes Here



Who should action

Downloadable list with all relevant data, actions, roles

Auto Fill Google Doc Templates from Google Sheets Data



21:22

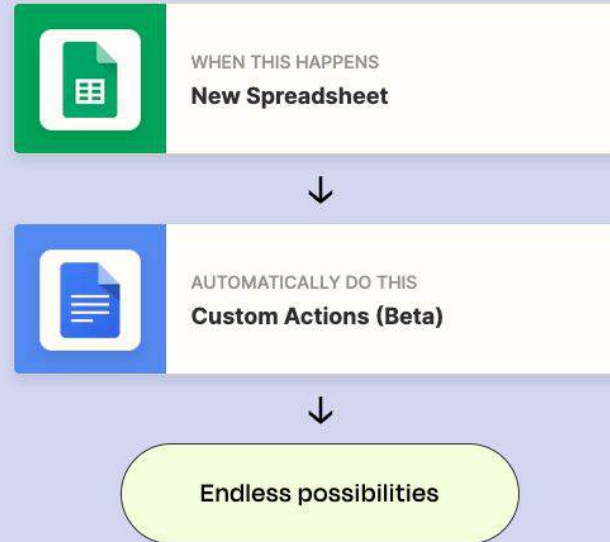
Auto Fill Google Doc Template from Google Sheets
Data Using Google Apps Script

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

Start with email for free

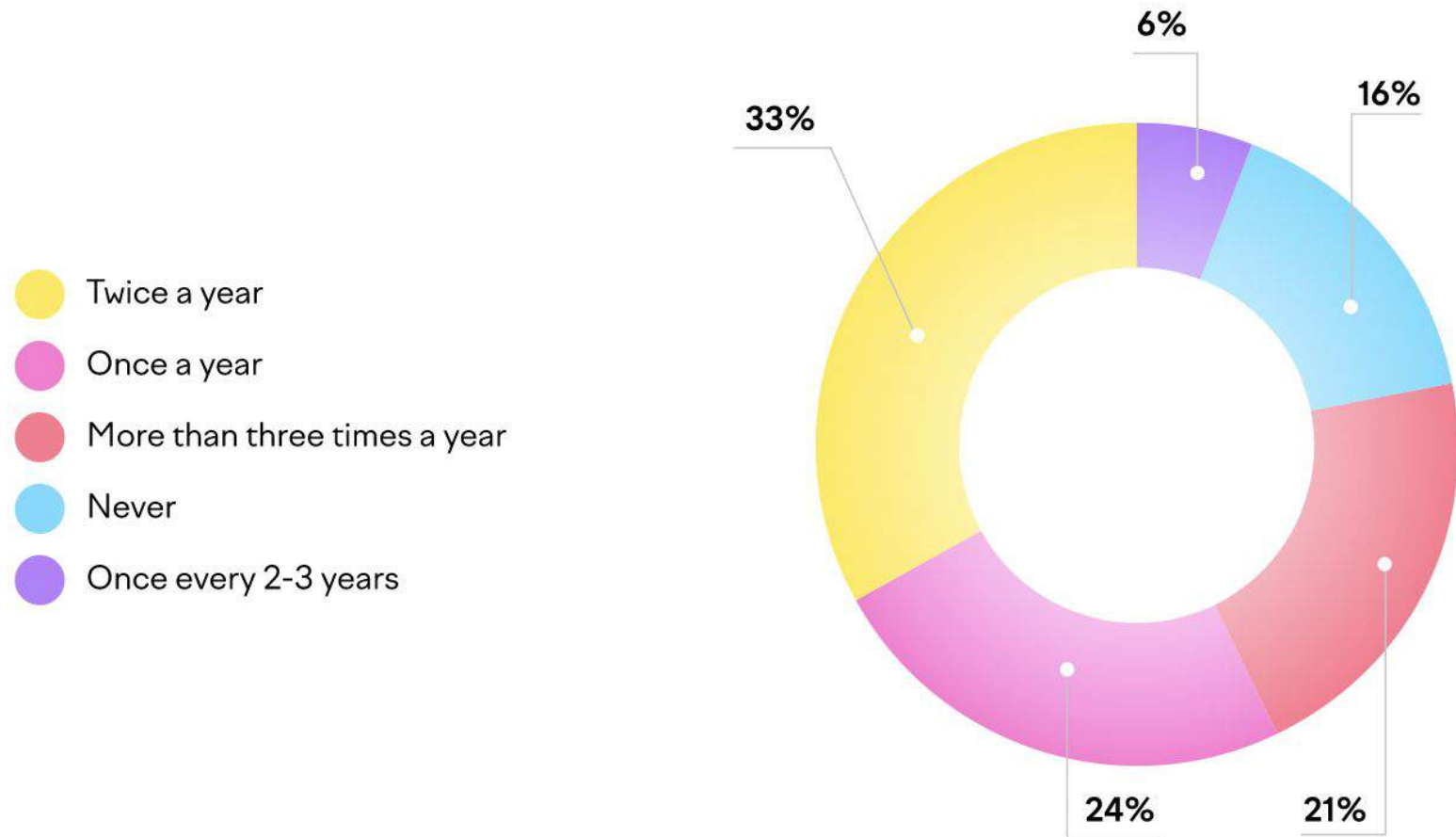


Start with Google for free



final note...

How often do you conduct content audits?

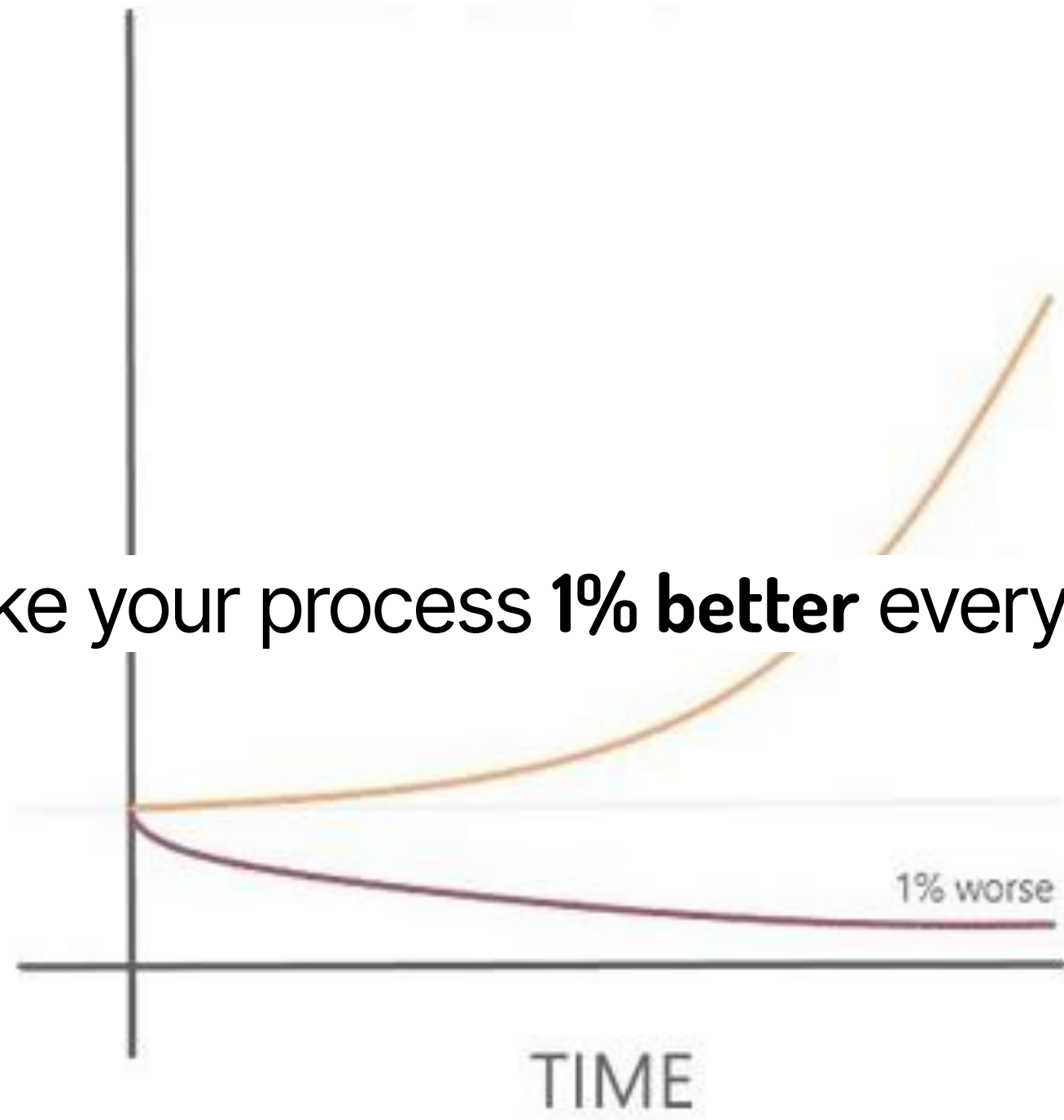


REDUCE
REUSE
RECYCLE

your auditing process

RESULTS

make your process **1% better** every 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





MLforSEO Academy
 real machine learning for Organic Search ✨

Resources | Blog | Contact | Sign up to the academy | Subscribe

Subscribe to our newsletter | Contact us




TRAINING ACADEMY FOR SEARCH MARKETERS

Improve your Organic Strategy with Machine Learning Automation

Unlock the future of organic search with MLforSEO.com, a training platform created by Lazarina Stoy to empower search marketers with machine learning expertise...

About us



20+ Templates, guides, tutorials, and more!
0 ChatGPT "hacks" and more!

The training platform for ambitious search marketers, interested in machine learning

Built on a foundation of expertise and innovation, our platform is designed with beginners in mind, to teach you real machine learning in SEO – everything from training your own models, to implementing ML APIs, but also learning what to do with the data, and how to make the most of the automations.

Radio Miqahed' Cray is the m...
 Lidia Infante
 Radio Miqahed' Cray is the m...
 Lidia Infante

LEARN THE HOW IN MACHINE LEARNING | [View all machine learning tutorials](#)

Machine Learning Tutorials

Straightforward machine learning tutorials, ideal for beginners. Start with all the resources you need, and step-by-step guidance.




Entity analysis

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

Lazarina Stoy · Mar 27, 2024

Text Classification

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

Lazarina Stoy · Mar 07, 2024

Features

- Practical machine learning
- No chatGPT hacks, tips, and tricks
- Training or API implementation
- Beginner-friendly
- Zero-to-hero

Benefits

- Practical skills
- Step by step guidance
- Templates for every tutorial
- Theoretical background
- Complete guidance

[View all our resources](#)

FREE CODE SAMPLES, SHEETS TEMPLATES, AND DATA VISUALIZATION DECISIONS | [View all our resources](#)

MLforSEO Templates

Kickstart machine learning implementation with our collection of templates, featuring Google Sheets templates with AppScript, Looker Studio dashboard templates, and a range of coding scripts and notebooks. Tailored for efficiency and effectiveness, these resources are designed to streamline your workflow and enhance your analysis, ensuring you start your machine learning journey on the right foot.

- Looker Studio Dashboards Templates (1)
- Google Sheets Templates (2)

Features

- SEO automation made easy
- Code samples
- Visualisation samples
- Applied learning
- Expert insights

Benefits

- Start with a template
- Save time
- Quickly understand your data
- Ideal for beginners
- Streamlined processes

MLforSEO.com ✨

(academy, templates & courses)



MLforSEO Templates

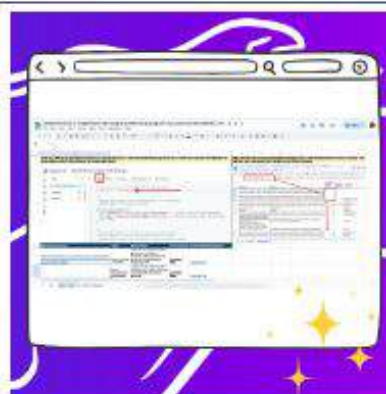
Kickstart machine learning implementation with our collection of templates, featuring Google Sheets templates with AppScript, Looker Studio dashboard templates, and a range of coding scripts and notebooks. Tailored for efficiency and effectiveness, these resources are designed with beginners in mind, ensuring you start your machine learning journey on the right foot.



Content Moderation with Google Cloud Natural Language API (Google Sheets Template and Apps Script)



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



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



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



Text Classification with with Google Cloud Natural Language API (Google Sheets Template with Apps Script)



Text Classification with Google Cloud Natural Language API (Looker Studio Template)

Machine Learning Tutorials

Straightforward machine learning tutorials and how-to guides, ideal for beginners. Learn how to implement an API, or train your own machine learning model from scratch, using popular tools and technologies. Each tutorial includes all the resources you need, plus step-by-step guidance.



Audio Transcription

How to transcribe audio with OpenAI's Whisper API in Google Colab (Python)

Lazarina Stoy. - May 26, 2024



Content Moderation

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

Lazarina Stoy. - May 8, 2024



Syntax analysis

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

Lazarina Stoy. - Apr 23, 2024

✦ academy.mlforseo.com ✦

Thank you for listening. ♡