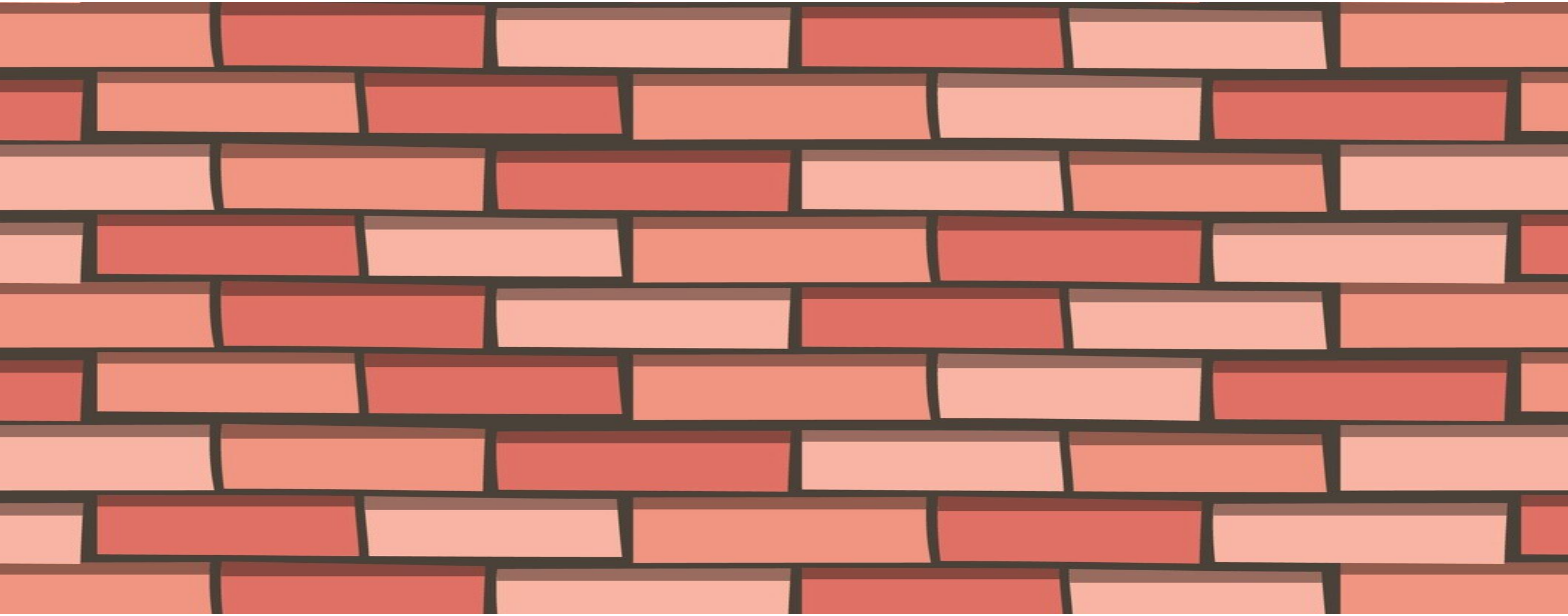
BEGINNER-FRIENDLY GUIDE TO ML-ENABLED AUTOMATION IN ORGANIC MARKETING

LAZARINA STOY.

SEO & Data Science Consultant

SMXL MILAN



the

‘YOU NEED HELP AND HELP IS THERE FOR YOU’

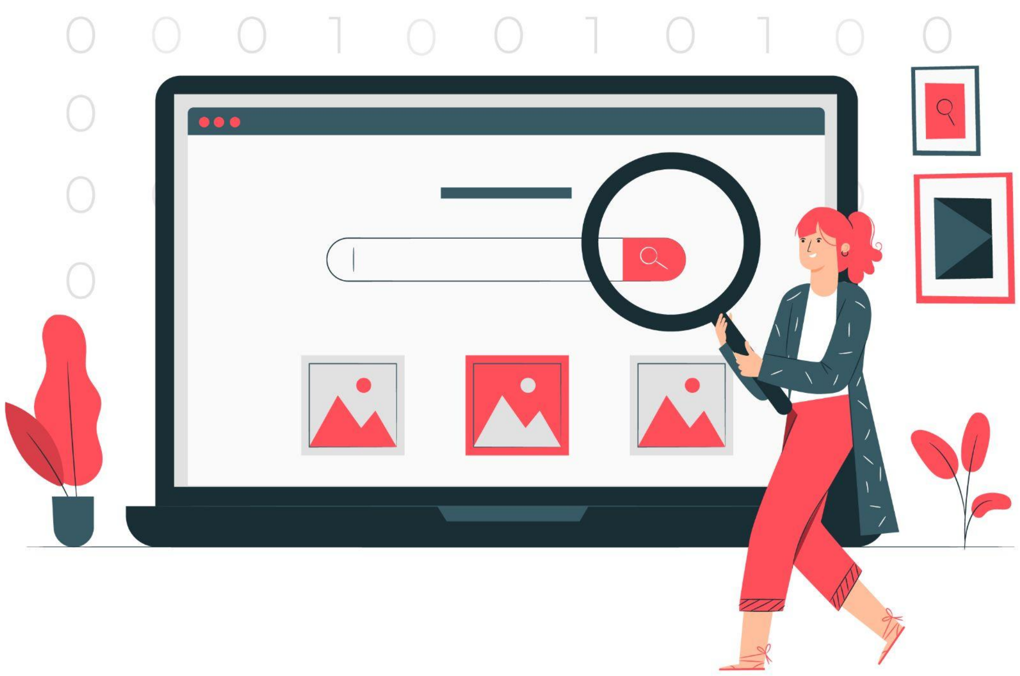
argument

the

‘AUTOMATION IS THE FUTURE’

argument

the 'SOMETHING IS HAPPENING AND IT'S MAKE OR BREAK TIME' argument



the

‘TAKE BACK TIME’

argument

the

‘CHALLENGE YOURSELF BY LEARNING SOMETHING NEW’

argument

I’ll let **you choose**, whichever speaks the most to you

SMXL MILAN

Instead, I'll start with this:

Page A scores for Search Query X Kalicube.pro								
Topicality	Quality	Speed	Entities	RankBrain	Struct. Data	Freshness	A.N.Other 1	A.N.Other 2
2	4	1	8	5	3	2	2	3

There's help in the face of **task automation** for **accelerating your organic performance** in most of these categories.

So, why not **get ahead**?

(before we dive in)

x This is a beginner-friendly talk, meaning you can absolutely do everything mentioned here, if you are a true beginner

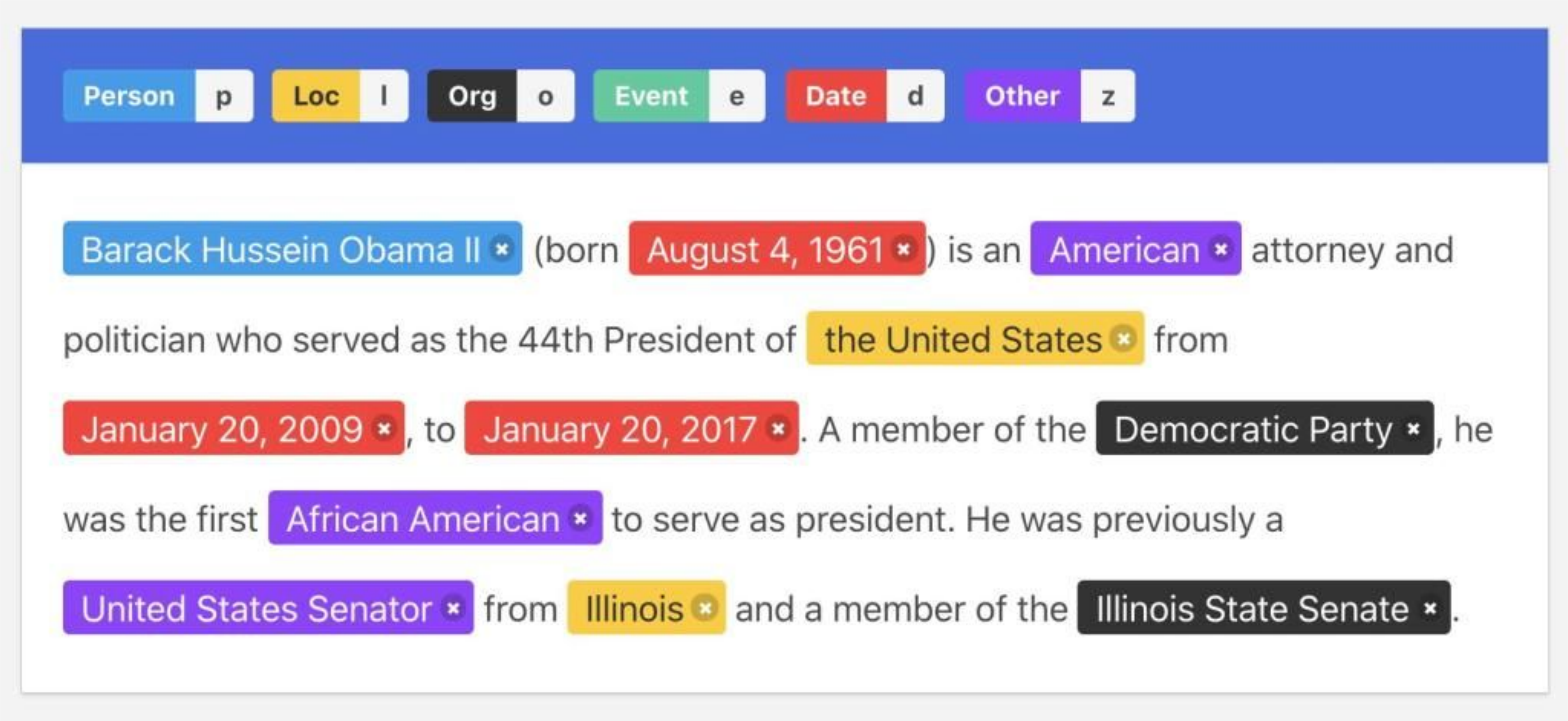
x No advanced coding needed but consider it a starting point - more advanced scripts will likely lead to better, more tailored results

x During this talk you might forget about chatGPT being the "ultimate AI model" (because it is not a 'one size fits all' type of field

let's talk about

entities

SMXL MILAN


Person p
Loc l
Org o
Event e
Date d
Other z
Barack Hussein Obama II (born August 4, 1961) is an American attorney and politician who served as the 44th President of the United States from January 20, 2009, to January 20, 2017. A member of the Democratic Party, he was the first African American to serve as president. He was previously a United States Senator from Illinois and a member of the Illinois State Senate.

SMXL MILAN

entity analysis is a must-have in today’s search landscape

SMXL MILAN

What components should you extract and analyse the entities of?

- your pages' content
- your titles and meta descriptions (SERP data) versus your competitors' titles and meta descriptions
- your competitors' content

★ first-party data
★ UGC
★ social mentions

does it matter how you do it?

yes - use a tailored model (not chatGPT)

SMXL MILAN

Google’s Natural Language API vs ChatGPT

Entities | Sentiment | Syntax | Categories

⟨Lazarina Stoy⟩$_{1}$ (formally known as ⟨Lazarina Stoyanova⟩$_{10}$) is an ⟨SEO⟩$_{4}$ & ⟨Data Science⟩$_{3}$, ⟨Sr. Manager⟩$_{7}$, a freelance ⟨SEO⟩$_{2}$ & ⟨Data Consultant⟩$_{8}$, and a ⟨storyteller⟩$_{5}$. ⟨Lazarina Stoy⟩$_{1}$ creates educational ⟨content⟩$_{13}$ in the ⟨SEO⟩$_{2}$, ⟨data science⟩$_{3}$, and ⟨analytics⟩$_{18}$ ⟨niche⟩$_{21}$, as well as ⟨resources⟩$_{9}$ that can help ⟨SEOs⟩$_{11}$ and digital ⟨analysts⟩$_{19}$ be more efficient with their time. ⟨Lazarina⟩$_{17}$ is a ⟨Conference Speaker⟩$_{16}$, having spoken at world-renowned ⟨conferences⟩$_{31}$ in the [illegible] ⟨topics⟩$_{15}$ that align with her professional ⟨mission⟩$_{36}$ – to make ⟨marketers⟩$_{6}$ ’ ⟨lives⟩$_{12}$ easie[illegible] ⟨automation⟩[illegible] and ⟨tools⟩$_{37}$ ⟨everyone⟩$_{29}$ (regardless of their ⟨tech background⟩$_{34}$) can apply in [illegible] ⟨Lazarina⟩$_{1}$ graduated from the ⟨University of Strathclyde⟩$_{14}$, where she studied to combine her greates[illegible] ⟨passions⟩$_{50}$ – ⟨marketing⟩$_{20}$ and ⟨technology⟩$_{25}$, with the ⟨aim⟩$_{35}$ to work on embedding ML [illegible] ⟨automation⟩$_{27}$ to help ⟨marketers⟩$_{23}$ ’ ⟨lives⟩$_{12}$ become easier. ⟨Lazarina⟩$_{1}$ loves connecting the ⟨dots⟩$_{39}$ between ⟨theory⟩$_{40}$ and ⟨practice⟩$_{33}$, finding ⟨patterns⟩$_{22}$, and discussing ⟨science⟩$_{28}$ in a ⟨way⟩$_{41}$ accessible for ⟨beginners⟩$_{42}$. ⟨Lazarina⟩$_{1}$ tells ⟨stories⟩$_{47}$ about ⟨marketing⟩$_{60}$ and ⟨technology⟩$_{64}$ that educate, inspire, and start ⟨conversations⟩$_{67}$. She has contributed to a ⟨number⟩$_{66}$ of well-known ⟨publications⟩$_{65}$, such as Towards ⟨Data Science⟩$_{62}$, ⟨Better Marketing⟩$_{49}$, as well as to a ⟨number⟩$_{43}$ of ⟨SEO publications⟩$_{44}$ of ⟨companies⟩$_{45}$ like ⟨Oncrawl & Wix⟩$_{61}$. ⟨Character-wise⟩$_{63}$, ⟨Lazarina⟩$_{1}$ is a progress-driven ⟨data⟩$_{26}$ and ⟨automation⟩$_{38}$ ⟨geek⟩$_{46}$. She is always seeking ⟨opportunities⟩$_{52}$ for improving the ⟨efficiency⟩$_{57}$ of ⟨processes⟩$_{51}$. ⟨Lazarina⟩$_{1}$ has a ⟨passion⟩$_{54}$ for spotting ⟨improvement opportunities⟩$_{55}$ in ⟨everything⟩$_{56}$ she does, making her a strong ⟨proponent⟩$_{58}$ of ⟨automation⟩$_{48}$ and ⟨machine learning⟩$_{59}$ in ⟨SEO processes⟩$_{53}$.

53

Entities in the given text are:

- Person: Lazarina Stoy (formally known as Lazarina Stoyanova)
- Job Title: SEO & Data Science, Sr. Manager; Freelance SEO & Data Consultant; Conference Speaker
- Field of Work: SEO, Data Science, Analytics, Marketing, Technology
- Education Institution: University of Strathclyde
- Publications: Towards Data Science, Better Marketing, Oncrawl, Wix

16

SMXL MILAN

Google Natural Language API

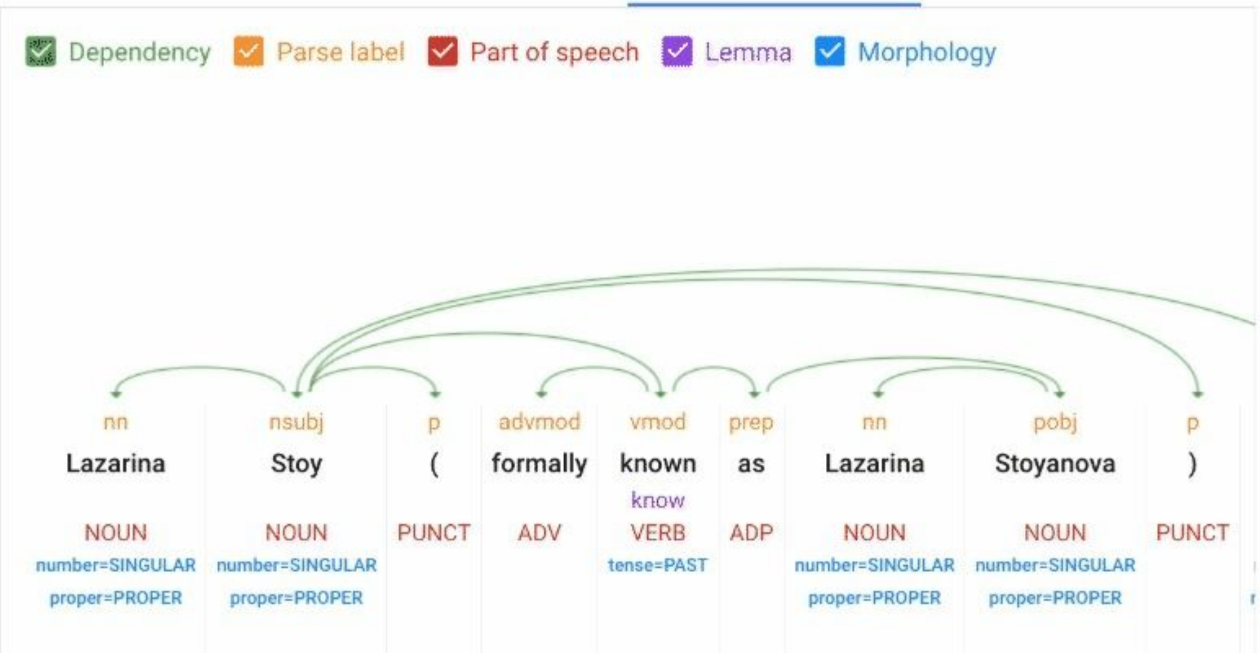
- **Entity =** recognised thing/concept
- **Salience** = importance
- **Sentiment Score**
- **Sentiment Magnitude** = strength
- **# of mentions** = entity prominence

SMXL MILAN

With Google's Natural Language API, you can also do Syntax analysis

- Dependency
- Parse label
- Part of speech
- Lemma
- Morphology

Google Natural Language API



SMXL MILAN

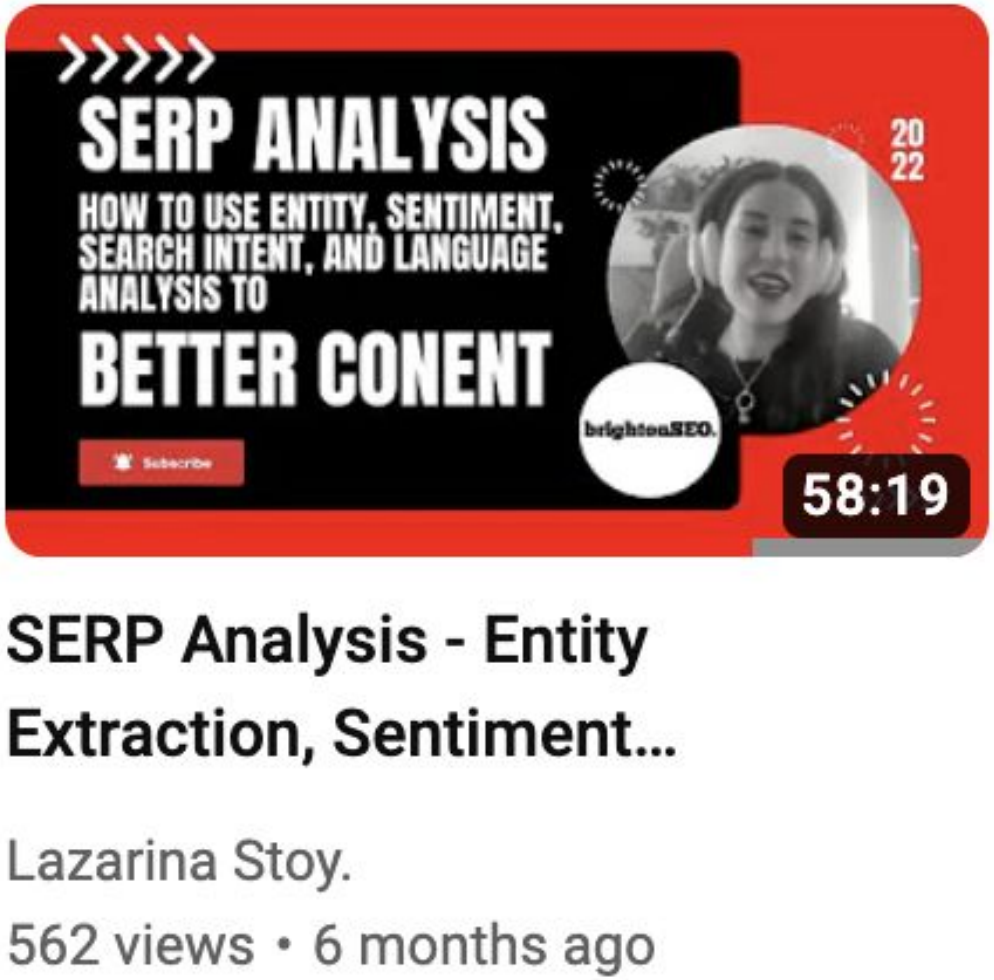
How important getting this right?

Entity analysis work is **central to multiple SEO projects**

SMXL MILAN

(pt.1 - SERP analysis)

- **Entity-driven SERP analysis:**
  - Using your keyword universe as a starting point, collect SERP data in bulk via dataforSEO
  - Run entity extraction on the titles and meta descriptions of results ranking on positions 1-10
  - Analyze the data to inform the content direction



(pt.2 - keyword research):

- **Entity-driven Keyword Research**
  - Using your keyword data as a starting point, validate which keywords contain entities
  - Create content maps, based on closely linked entities
  - Create lists of positive and negative secondary entities to discuss for content writers
  - Create lists of entity attributes to address in content
  - As an additional step, validate which keywords contain entities that are part of the knowledge graph

(pt.3 - internal link analysis):

- **Entity-driven Internal linking anchor text analysis**
    - Using your internal linking anchor text data as a starting point, validate which anchors contain entities
    - As an additional step, validate which keywords contain entities that are part of the knowledge graph 1: Analyze anchor text for entities

How to Implement Machine Learning in Your Internal...

Lazarina Stoy.

395 views • 7 months ago

SMXL MILAN

(pt.3 - internal link analysis):

- **Entity-driven content analysis for internal link identification analysis**
  - Using your website content (scrape the content via crawling), combined with internal links and traffic report, identify pages with entities mentioned and relationship with internal links and traffic.

How to Implement Machine Learning in Your Internal...

Lazarina Stoy.

395 views • 7 months ago

(pt.4 - content analysis, etc..):

- **Entity-driven site content analysis**
  - Starting with your website content (scraped), extract entities from the text, titles and meta descriptions
  - If an entity, appears more than X times on a page of Y words, highlight it as the prominent entity for the article
  - How many articles with prominent entities have them in the title/meta description
  - How many articles with prominent entities have internal links incoming from other articles, where the same entities are mentioned → internal link opportunities
  - Highlight articles that don't contain any entities, map with traffic → Content enhancement/ consolidation opportunities
- **Entity-driven competitor content analysis**
  - Process is same as above, only with competitor website content
- **Entity-driven social comments analysis, etc...**
  - Scrape social comments from YouTube, TikTok, Twitter, Facebook
  - Analyse for entities
  - Map against site content

the **quality of your analysis** will dictate the **quality of your content strategy**

(so, pretty important to get it right)

SMXL MILAN

let’s talk **process**

SMXL MILAN

✨ copy the Sheets templates ✨

brightonSEO demo sheet | Lazarina Stoy (lazarinastoy.com)

File Edit View Insert Format Data Tools Extensions Help Save Values

Entity Recognition & Entity Sentiment | brightonSEO demo

File Edit View Insert Format Data Tools Extensions Help Sentiment Tools

Sentiment Analysis | brightonSEO demo

File Edit View Insert Format Data Tools Extensions Help Sentiment Tools

1. Copy the templates. Get your API key. Enter it in the AppScript tab.

your api key set-up guide

Cloud Natural Language > Documentation > Guides

Quickstart: Setup the Natural Language API

This guide provides all required setup steps to start using Natural Language.

your sheets template

Entity Recognition & Entity Sentiment | brightonSEO demo

File Edit View Insert Format Data Tools Extensions Help Sentiment Tools

2. Input your data (URL (yours or competitors') + page content, paragraphs, titles, MDs, etc.)

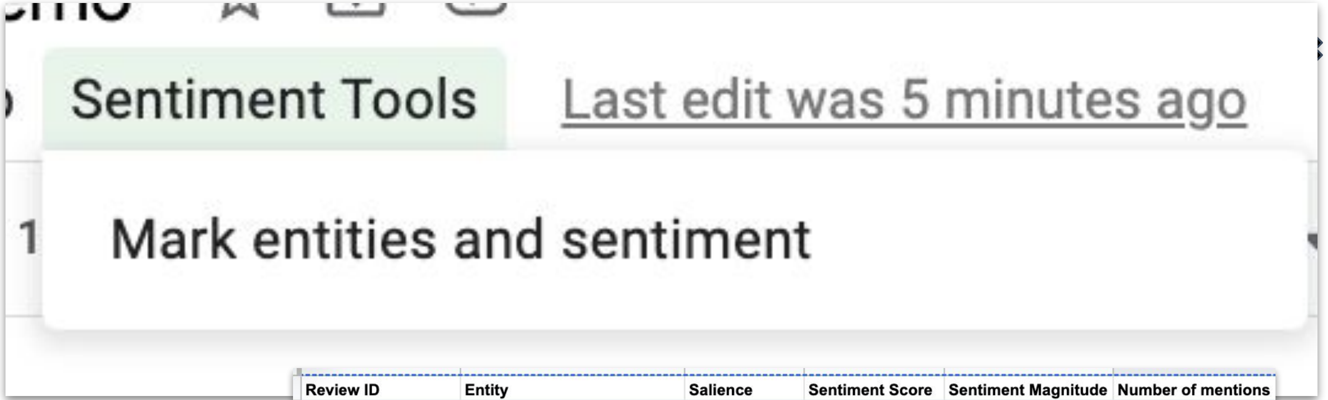
Entity Recognition & Entity Sentiment | brightonSEO demo

File Edit View Insert Format Data Tools Extensions Help Sentiment Tools Last edit was seconds ago

URLs **Meta Descriptions**

id	comments
https://en.wikipedia.org/wiki/Enterprise_application_integration	Enterprise application integration (EAI) is the use of software and computer sy
https://www.oracle.com/database/what-is-database/	-
https://www.oracle.com/database/what-is-a-relational-database/	A relational database is a type of database that stores and provides access to
https://www.mongodb.com/databases/non-relational	A non-relational database stores data in a non-tabular form, and tends to be m
https://www.cloudflare.com/learning/cdn/what-is-a-cdn/	A content delivery network (CDN) refers to a geographically distributed group
https://aws.amazon.com/free/	Deploy Secure, Reliable, & Scalable Websites, Apps or Processes with Free C
https://www.pcmag.com/picks/the-best-cloud-storage-and-file-sharing-services	The Best Cloud Storage and File-Sharing Services for 2022 · Our Top 7 Picks
https://www.containersystems.com/	Container Systems, Inc. ... If you are searching for a source for pallet rack, she
https://www.gartner.com/en/information-technology/glossary/mobile-web-applications	Mobile Web applications refer to applications for mobile devices that require o
https://aws.amazon.com/serverless/	Serverless computing allows you to build and run applications and services wi
https://aws.amazon.com/free/machine-learning/	Access the Broadest & Deepest Set of Machine Learning Services for Your Bu
https://www.scnsoft.com/services/analytics	-
https://www.gartner.com/en/information-technology/glossary/application-integration	-
https://www.bdc.ca/en/articles-tools/business-strategy-planning/manage-business/3-strategies	-
https://www.lwtech.edu/academics/computing-software-development/	Located just outside of Seattle, Washington, the Computing and Software Dev
https://www.gominis.com/new-orleans/-/new-orleans/moving-containers/	We Provide Low-Cost Prices Our Customers Want & High-Quality Service The
https://www.intercom.com/drlp/customer-engagement-software	Deliver eye-opening marketing experiences to maximize conversions and capt
https://www.dbsinfo.com/	Works offline so you can serve anywhere. Notifications in real time. Upload ph
https://developer.chrome.com/docs/devtools/	Chrome DevTools is a set of web developer tools built directly into the Google
https://mitratech.com/products/clusterseven/end-user-computing-risk-management/	The End User Computing (EUC) definition is 'a system in which individuals are
https://developer.mozilla.org/en-US/docs/Learn/Front-end_web_developer	Here we provide you with a structured course that will teach you all you need t
https://www.goodfirms.co/glossary/front-end	The mobile app development process involves cooperation between front end
https://aws.amazon.com/gametech/	Build faster, operate smarter, and create innovative games using Amazon Wel
https://en.wikipedia.org/wiki/Internet_of_things	The Internet of things (IoT) describes physical objects with sensors, processin
https://en.wikipedia.org/wiki/Machine_learning	Machine learning (ML) is a field of inquiry devoted to understanding and buildi
https://www.g2.com/categories/cloud-management-platforms	Cloud management platforms (CMPs) are toolsets that help companies monit
https://www.imperva.com/learn/data-security/cloud-governance/	Cloud governance is a set of rules and policies adopted by companies that rur
https://www.mediaservices.com/	Entertainment payroll, accounting and residuals. Get W2 payroll service for cre
https://aws.amazon.com/products/migration-and-transfer/	AWS offers the most complete migration solutions to help you assess, mobiliz
https://www.vmware.com/topics/glossary/content/cloud-networking.html	Cloud networking is a type of IT infrastructure in which some or all of an organ

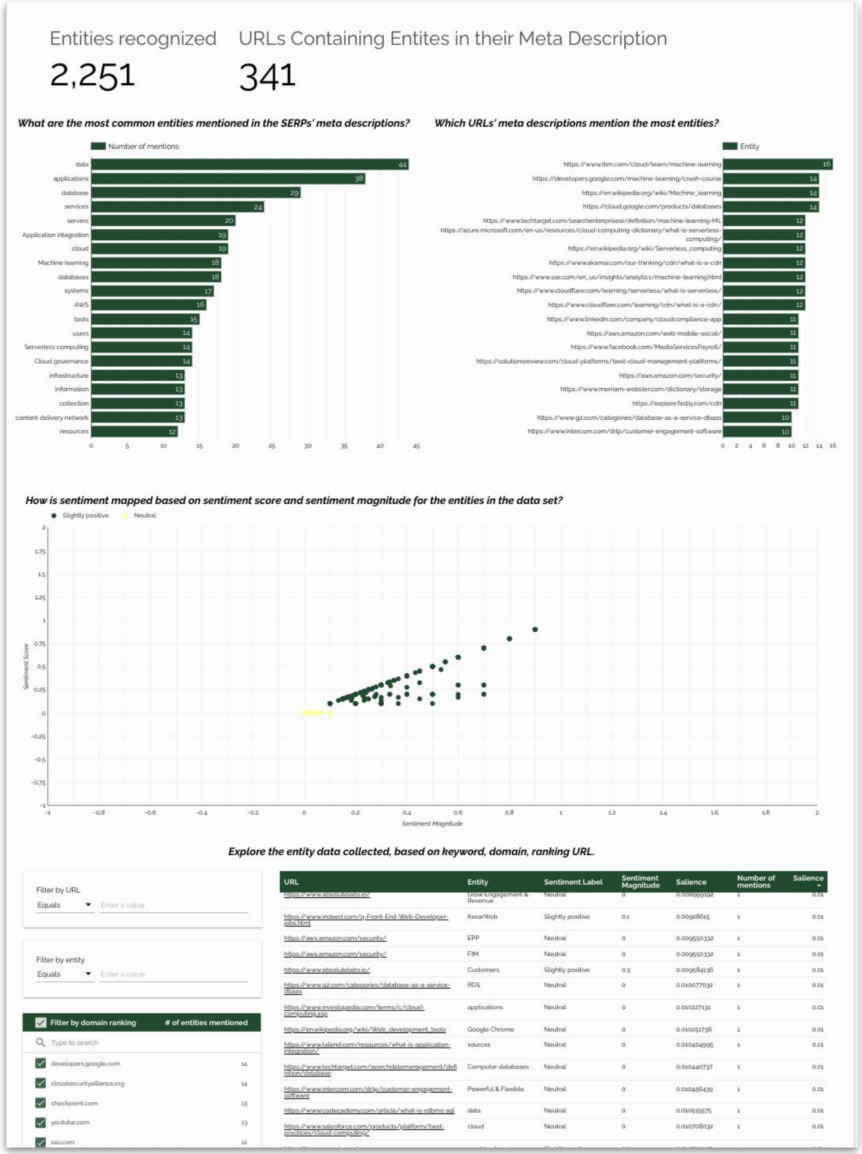
3. Get the entities and sentiment scores for each of your analyzed data points (pages content, titles, feedback forms, ...)



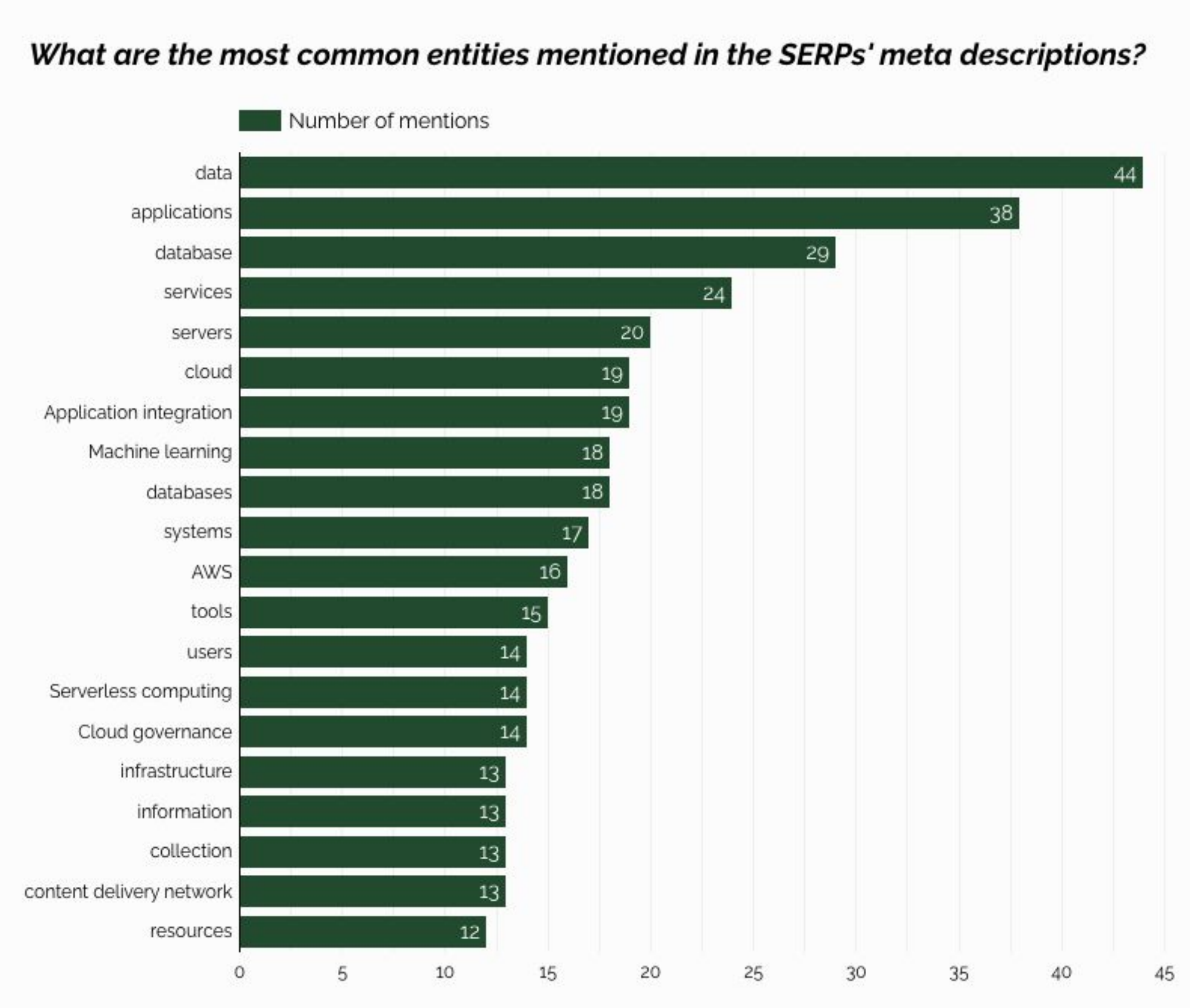
Review ID	Entity	Salience	Sentiment Score	Sentiment Magnitude	Number of mentions
https://en.wikipedia.or	Enterprise application integratio	0.6282338	0	0	2
https://en.wikipedia.or	EAI	0.107579336	0	0	1
https://en.wikipedia.or	software	0.07536966	0	0	1
https://en.wikipedia.or	computer systems	0.06946866	0	0	1
https://en.wikipedia.or	set	0.053013343	0	0	1
https://en.wikipedia.or	enterprise computer	0.042577475	0	0	1
https://en.wikipedia.or	principles	0.023757715	0	0	1
https://www.oracle.cor	database	0.70133245	0.1	0.3	2
https://www.oracle.cor	data points	0.098208494	0	0.1	1
https://www.oracle.cor	database	0.08552961	0.1	0.1	1
https://www.oracle.cor	stores	0.04493161	0.1	0.1	1
https://www.oracle.cor	access	0.035053052	0	0	1
https://www.oracle.cor	databases	0.02072558	0	0	1
https://www.oracle.cor	another	0.0142192	0	0	1
https://www.mongodb.	database	0.3604171	0	0	1
https://www.mongodb.	stores	0.30372602	0	0	1
https://www.mongodb.	data	0.16117582	0	0	1
https://www.mongodb.	form	0.09173636	0	0	1
https://www.mongodb.	relational database structures	0.065426126	0	0	1
https://www.mongodb.	SQL	0.017518587	0	0	1
https://www.cloudflare	content delivery network	0.3859859	0	0	1
https://www.cloudflare	group	0.26281422	0	0	1
https://www.cloudflare	servers	0.10876744	0	0	1
https://www.cloudflare	CDN	0.10294777	0	0	1
https://www.cloudflare	delivery	0.07263352	0	0	1
https://www.cloudflare	Internet content	0.06685114	0	0	1
https://aws.amazon.cc	Secure	0.31401837	0.4	0.4	1
https://aws.amazon.cc	Websites	0.2091138	0.4	0.4	1
https://aws.amazon.cc	Apps	0.14441134	0.3	0.3	1

SMXL
MILAN

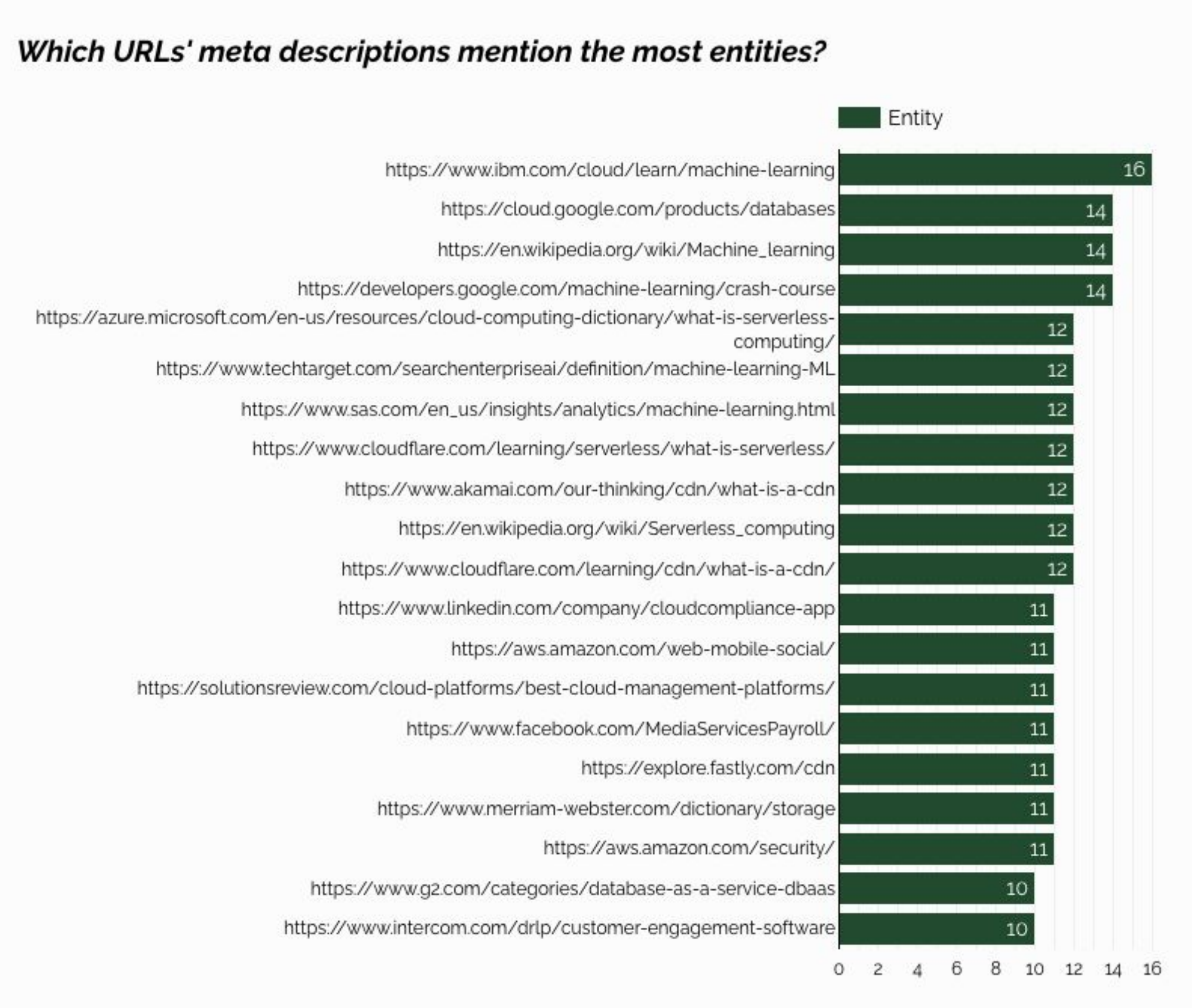
4. Visualize in Looker Studio and explore the data



SMXL MILAN



What are the most common entities mentioned in the SERPs' meta descriptions?
Number of mentions
data 44
applications 38
database 29
services 24
servers 20
cloud 19
Application integration 19
Machine learning 18
databases 18
systems 17
AWS 16
tools 15
users 14
Serverless computing 14
Cloud governance 14
infrastructure 13
information 13
collection 13
content delivery network 13
resources 12
0 5 10 15 20 25 30 35 40 45



Which URLs' meta descriptions mention the most entities?
Entity
https://www.ibm.com/cloud/learn/machine-learning 16
https://cloud.google.com/products/databases 14
https://en.wikipedia.org/wiki/Machine_learning 14
https://developers.google.com/machine-learning/crash-course 14
https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-serverless-computing/ 12
https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML 12
https://www.sas.com/en_us/insights/analytics/machine-learning.html 12
https://www.cloudflare.com/learning/serverless/what-is-serverless/ 12
https://www.akamai.com/our-thinking/cdn/what-is-a-cdn 12
https://en.wikipedia.org/wiki/Serverless_computing 12
https://www.cloudflare.com/learning/cdn/what-is-a-cdn/ 12
https://www.linkedin.com/company/cloudcompliance-app 11
https://aws.amazon.com/web-mobile-social/ 11
https://solutionsreview.com/cloud-platforms/best-cloud-management-platforms/ 11
https://www.facebook.com/MediaServicesPayroll/ 11
https://explore.fastly.com/cdn 11
https://www.merriam-webster.com/dictionary/storage 11
https://aws.amazon.com/security/ 11
https://www.g2.com/categories/database-as-a-service-dbaas 10
https://www.intercom.com/drlp/customer-engagement-software 10
0 2 4 6 8 10 12 14 16

Explore the entity data collected, based on keyword, domain, ranking URL.

Filter by URL

Equals ▾ Enter a value

Filter by entity

Equals ▾ Enter a value

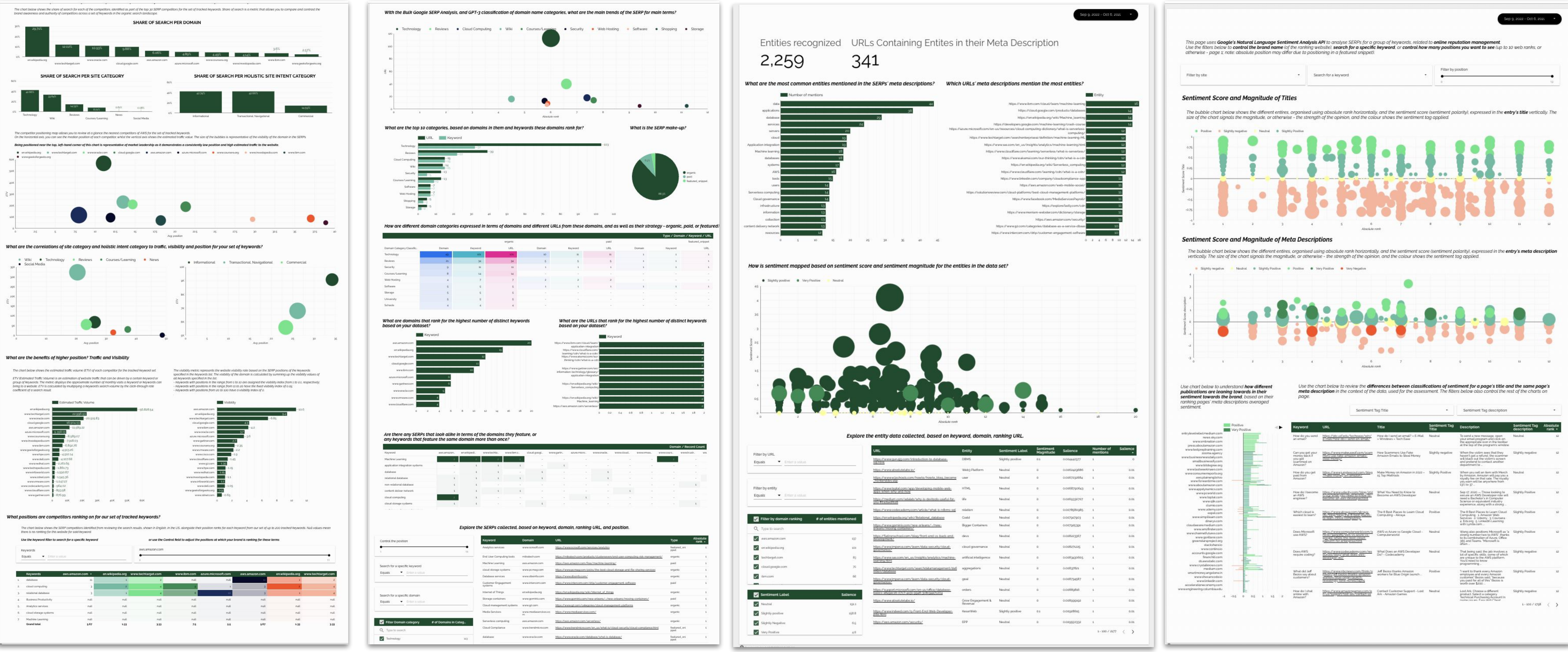
Filter by domain ranking	# of entities mentioned
developers.google.com	14
cloudsecurityalliance.org	14
checkpoint.com	13
youtube.com	13
sas.com	12

Sentiment Label (Exclude 1)	Salience
Neutral	191.1
Slightly positive	158.8
Slightly Negative	6.5
Very Positive	4.6

URL	Entity	Sentiment Label	Sentiment Magnitude	Salience	Number of mentions	Salience
	Revenue					
https://www.indeed.com/q-Front-End-Web-Developer-jobs.html	KesarWeb	Slightly positive	0.1	0.00928615	1	0.01
https://aws.amazon.com/security/	EPP	Neutral	0	0.009550332	1	0.01
https://aws.amazon.com/security/	FIM	Neutral	0	0.009550332	1	0.01
https://www.absolutelabs.io/	Customers	Slightly positive	0.3	0.009584136	1	0.01
https://www.g2.com/categories/database-as-a-service-dbaas	RDS	Neutral	0	0.010077032	1	0.01
https://www.investopedia.com/terms/c/cloud-computing.asp	applications	Neutral	0	0.010227131	1	0.01
https://en.wikipedia.org/wiki/Web_development_tools	Google Chrome	Neutral	0	0.01025173B	1	0.01
https://www.talend.com/resources/what-is-application-integration/	sources	Neutral	0	0.010404995	1	0.01
https://www.techtarget.com/searchdatamanagement/definition/database	Computer databases	Neutral	0	0.010440737	1	0.01
https://www.intercom.com/drlp/customer-engagement-software	Powerful & Flexible	Neutral	0	0.010456439	1	0.01
https://www.codecademy.com/article/what-is-rdbms-sql	data	Neutral	0	0.010515575	1	0.01
https://www.salesforce.com/products/platform/best-practices/cloud-computing/	cloud	Neutral	0	0.010708032	1	0.01
https://www.mysql.com/	machine learning workloads	Slightly positive	0.2	0.010735927	1	0.01
https://www.blackforestmktg.com/	Black Forest Container Systems	Neutral	0	0.010848048	1	0.01
https://www.imperva.com/learn/data-security/cloud-governance/	data security	Neutral	0	0.0115666855	1	0.01
https://martinfowler.com/articles/serverless.html	databases	Slightly positive	0.1	0.011604427	1	0.01
https://www.usg.edu/galileo/skills/unit04/primer04_01.phtml	number	Neutral	0	0.01169396	1	0.01
https://en.wikipedia.org/wiki/Front-	CSS	Neutral	0	0.011733394	1	0.01

1 - 100 / 2169 ‹ ›

✨ get the Looker Studio template ✨



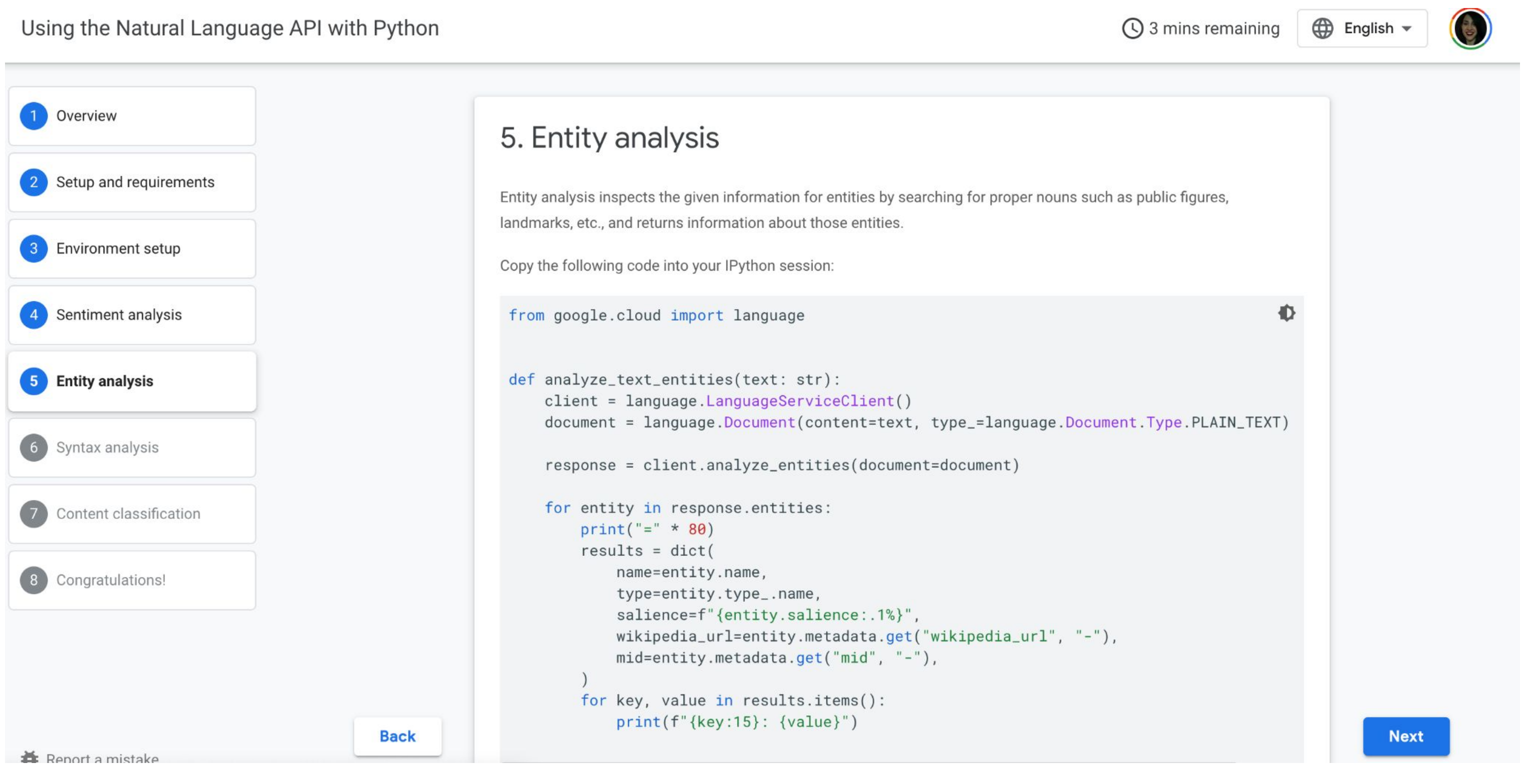
If you're working with a multi-million page site, or massive dataset, choose Python & Pandas library, not Google Sheets

Google releases handy code labs (practice environments) with sample code.

Tip: Prompt an LLM or code buddy to rework the code so it works in:

- Google colab
- With an API key
- Takes input a pandas dataset (csv)

Test, troubleshoot, iterate as needed for your dataset.



Then what?

- Audit if you’re stuffing your pages with unnecessary entities that are not related to your main topic
- Audit if you’re providing enough context about the topics and entities you’re discussing for Google to know what you’re talking about
- Audit the use of entities between you and other, better-performing competitors

a step further:

understanding language used within the high-performing articles for your terms can be beneficial for building content briefs.

Playground

Load a preset... Save

write an appscript formula using javascript for a function identifying the main n-grams (words) of a text input using the text as input and the number of words to be included in a gram

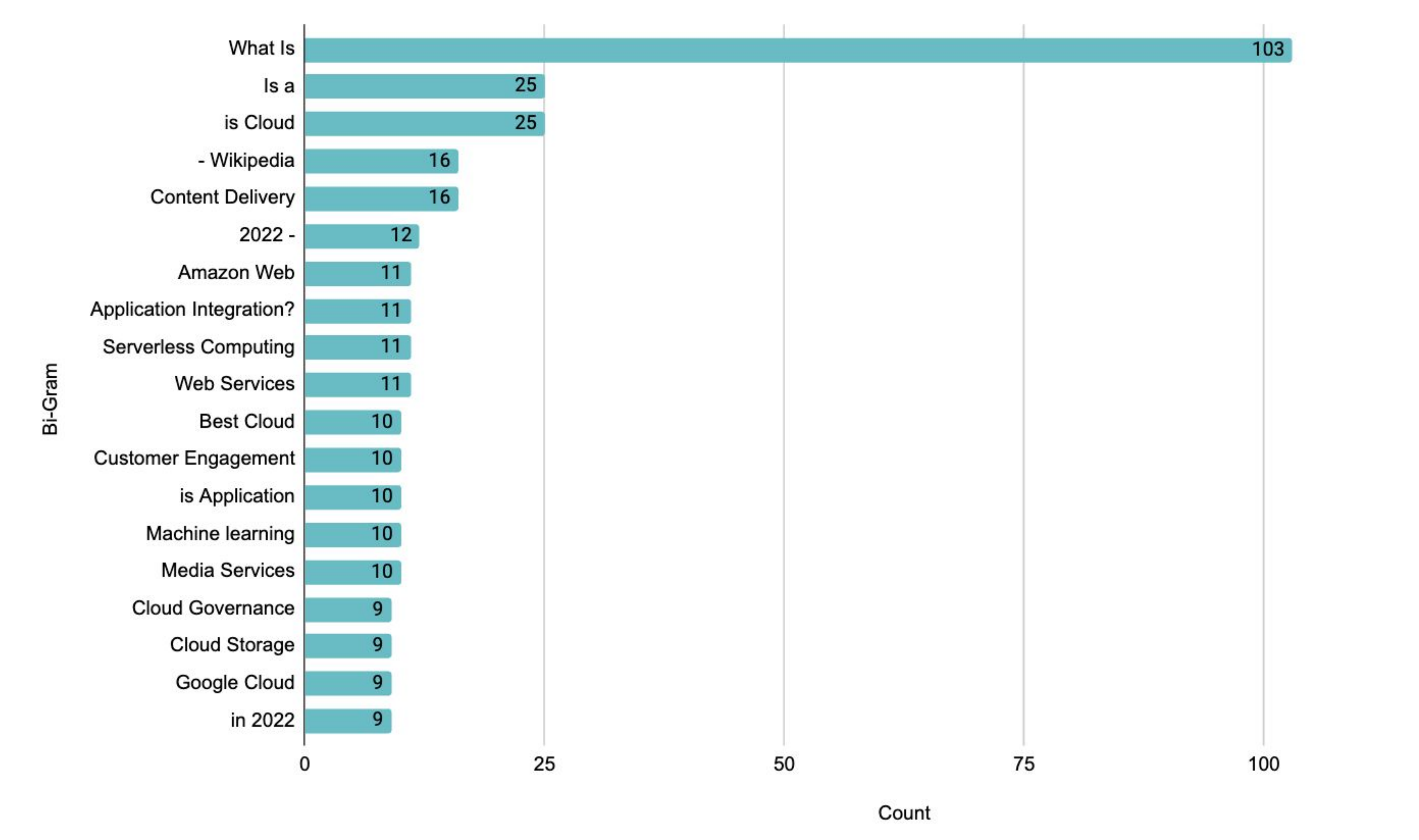
```
function getNgrams(text, n) {
 var words = text.split(" ");
 var ngrams = [];
 for (var i = 0; i < words.length - (n - 1); i++) {
  ngrams.push(words.slice(i, i + n).join(" "));
 }
 return ngrams;
}
```

GPT3

SMXL MILAN

```
=transpose(getngrams(F2, 2))
```

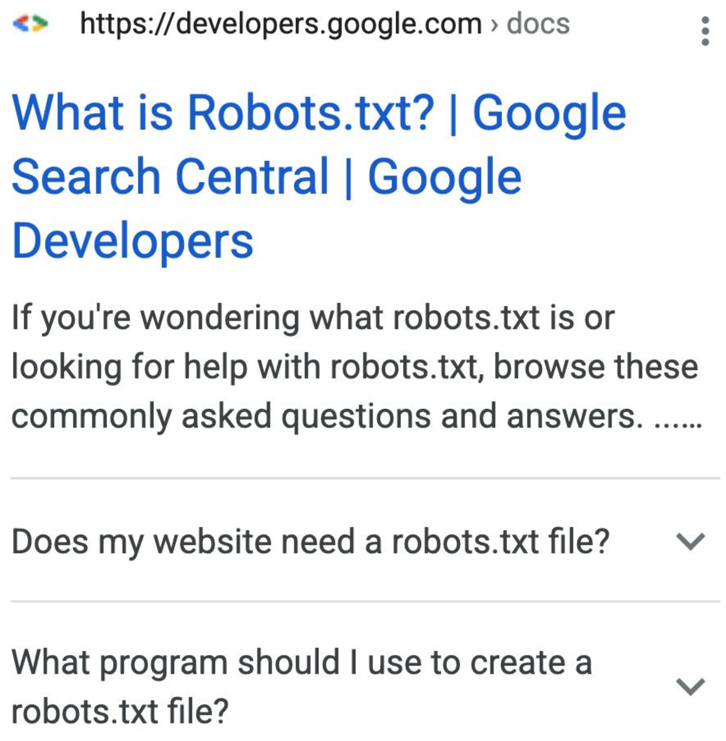
I	J	K	L	M	
bi-grams in title					
Enterprise application	application integration	integration -	- Wikipedia		
What Is	Is a	a Database	Database -	- Oracle	
What is	is a	a Relational	Relational Database	Database (RDBMS)?	(RDBMS)? -
What Is	Is A	A Non-Relational	Non-Relational Database?	Database? -	- MongoDB
What is	is a	a CDN?	CDN? |	| How	How do


What Is
Is a
is Cloud
- Wikipedia
Content Delivery
2022 -
Amazon Web
Application Integration?
Serverless Computing
Web Services
Best Cloud
Customer Engagement
is Application
Machine learning
Media Services
Cloud Governance
Cloud Storage
Google Cloud
in 2022
103
25
25
16
16
12
11
11
11
11
10
10
10
10
10
9
9
9
9
Bi-Gram
0
25
50
75
100
Count

Language use analysis can also be useful for scaling generation of or identifying opportunities for Structured Data (schema) mark-up

SMXL MILAN

Here’s a demo with perhaps the simplest of solutions & schema types…




```
<script type="application/ld+json">
 {
  "@context": "https://schema.org",
  "@type": "FAQPage",
  "mainEntity": [{
    "@type": "Question",
    "name": "What is the return policy?",
    "acceptedAnswer": {
      "@type": "Answer",
[illegible]
    }
  }, {
    "@type": "Question",
    "name": "How long does it take to process a refund?",
    "acceptedAnswer": {
      "@type": "Answer",
[illegible]
    }
  }, {   }
 </script>
```

Python Extract a Q&A pair from content and Turn into FAQ Schema.ipynb

File Edit View Insert Runtime Tools Help All changes saved

How it works:

- You provide the content (currently - copy-paste; of course, you can edit the script so it reads from a column from a pandas dataset (csv) if working in bulk)
- Script tokenizes the text, discovers the questions, and pulls the answers
- Script organises these into a schema dictionary, which is saved as a JSON file

```python
import nltk
nltk.download('punkt')
import re
import json

# Function to check if a sentence is a question that ends with a question mark
def is_question(sentence):
    question_words = ['what', 'when', 'where', 'which', 'who', 'whom', 'whose', 'why', 'how', 'is', 'will', 'should']
    if sentence.split()[0].lower() in question_words and sentence.strip().endswith('?'):
        return True
    else:
        return False


# Read text file
text =input("Paste the webpage content: ")

# Tokenize text into sentences
sentences = nltk.sent_tokenize(text)

# Initialize FAQ Schema dictionary
faq_schema = {'@context': 'https://schema.org', '@type': 'FAQPage', 'mainEntity': []}

# Loop through sentences to find questions and their answers
for i in range(len(sentences)):
    if is_question(sentences[i]):
        # Extract question
        question = re.sub(r'[^\w\s]', '', sentences[i]).strip()
        # Extract answer
        j = i+1
        answer = ''
        while j < len(sentences) and not is_question(sentences[j]):
            answer += ' ' + sentences[j]
            j += 1
        answer = answer.strip()
        # Add question-answer pair to FAQ Schema
        faq_schema['mainEntity'].append({'@type': 'Question', 'name': question, 'acceptedAnswer': {'@type': 'Answer', 'text': answer}})

# Save FAQ Schema to JSON file
with open('output.json', 'w') as f:
    json.dump(faq_schema, f, indent=4)
print(faq_schema)
```

Here's an example blog post

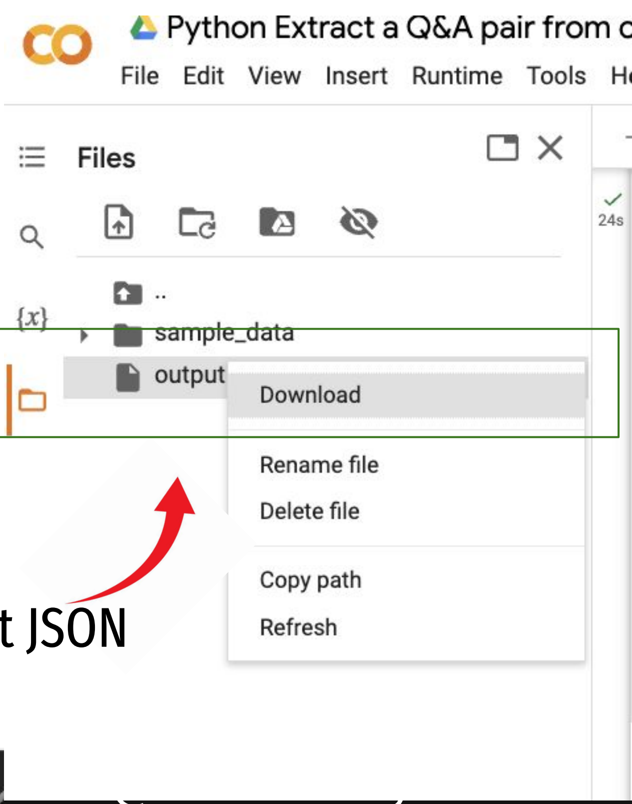
Table Of Contents

\+ Code + Text

```python
# Tokenize text into sentences
sentences = nltk.sent_tokenize(text)

# Initialize FAQ Schema dictionary
faq_schema = {'@context': 'https://schema.org', '@type': 'FAQPage', 'mainEntity': []}

# Loop through sentences to find questions and their answers
for i in range(len(sentences)):
    if is_question(sentences[i]):
        # Extract question
        question = re.sub(r'[^\w\s]', '', sentences[i]).strip()
        # Extract answer
        j = i+1
        answer = ''
        while j < len(sentences) and not is_question(sentences[j]):
            answer += ' ' + sentences[j]
            j += 1
        answer = answer.strip()
        # Add question-answer pair to FAQ Schema
        faq_schema['mainEntity'].append({'@type': 'Question', 'name': question, 'acceptedAnswer': {'@type': 'Answer', 'text': answer}})

# Save FAQ Schema to JSON file
with open('output.json', 'w') as f:
    json.dump(faq_schema, f, indent=4)
print(faq_schema)
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
Paste the webpage content: ch engines and visitors.
```

Run the cell, paste the content



Download the output JSON

SMXL MILAN

Review & edit, if needed


```json
{
    "@context": "https://schema.org",
    "@type": "FAQPage",
    "mainEntity": [
        {
            "@type": "Question",
            "name": "Why should you link your blog from the main menu",
            "acceptedAnswer": {
                "@type": "Answer",
                "text": "The main navigation is typically the primary point of entry for website users and should contain all significant sections, includi
            }
        },
        {
            "@type": "Question",
            "name": "Why should you link your blog from the footer menu",
            "acceptedAnswer": {
                "@type": "Answer",
                "text": "Placing a link to your blog in the footer ensures accessibility from every page of your website, providing a seamless user experie
            }
        },
        {
            "@type": "Question",
            "name": "What should you choose  hosting the blog on a subdomain or subfolder",
            "acceptedAnswer": {
                "@type": "Answer",
                "text": "One option is to integrate your blog into your website\u2019s domain by creating a subdirectory or otherwise -subfolder, such as y
            }
        },
        {
            "@type": "Question",
            "name": "Should you include tags and category names in your blog URLs",
            "acceptedAnswer": {
                "@type": "Answer",
                "text": "Including tags and category names in your blog URLs can have some benefits, but it may not be necessary in all cases. Including ta
            }
        },
        {
            "@type": "Question",
            "name": "What are the benefits of using a mix of broad and specific topic tags in a blog",
            "acceptedAnswer": {
```

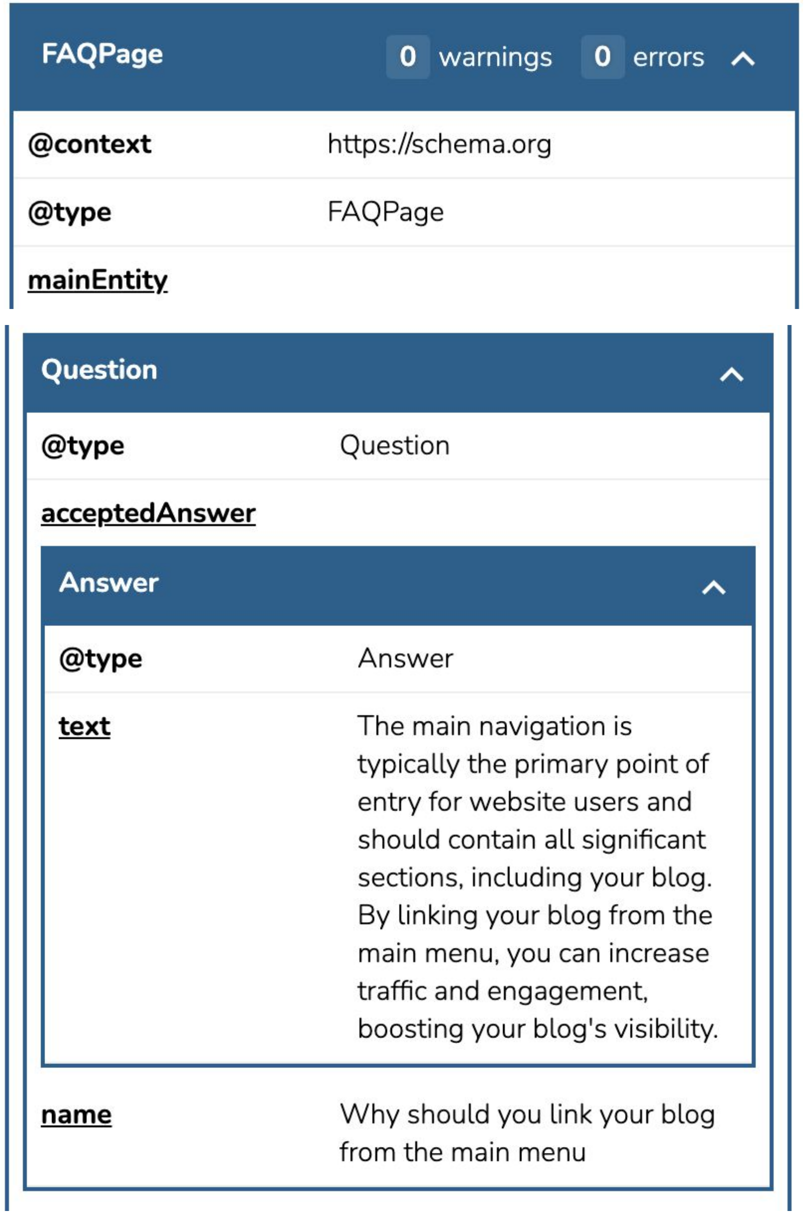

Review if the replies are accurate - typically it might include more than one paragraph, so you might want to shorten them. (you can also edit the script so it only pulls the first 1 or two sentences after the question)

TIP: If using for multiple content pieces, crawl your site, export the content, and change the method to list-based based on the column where the content is exported, as opposed to user input. (using python pandas library)

Upload.

What took minutes even with the most straightforward methods (e.g. using a schema markup generator, and copy-pasting individual questions into it), now took seconds.

With a bit of testing and script manipulation, you can put this process into production.



SMXL MILAN

This approach is great for websites and schema types that are niche-specific

(eg. VehicleListing, PropertyListing, Recipe, etc.) where you:

- You know the page format used for all pages and the elements you should pull schema data into
- you know the structure of the schema
- AND if you have a website with hundreds or thousands of pages, you can also use keyword identification to flag pages for structured data or migration to a particular template

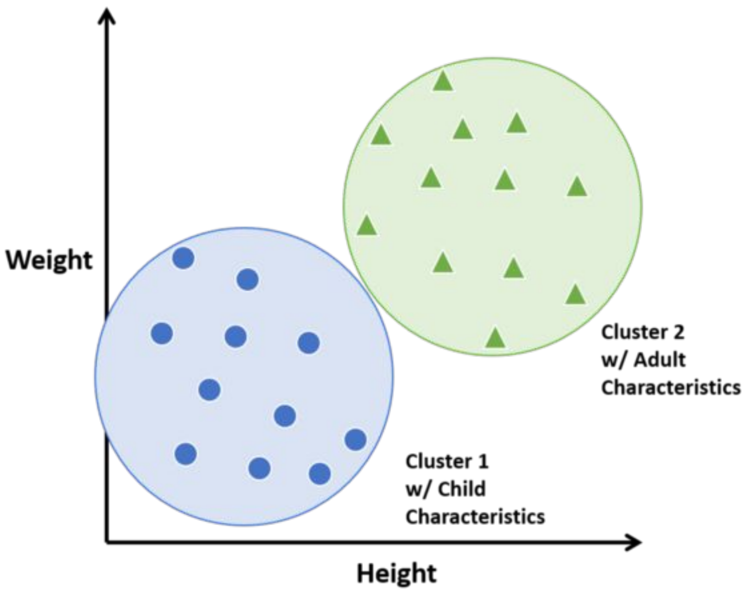
let’s talk about

clustering and classification

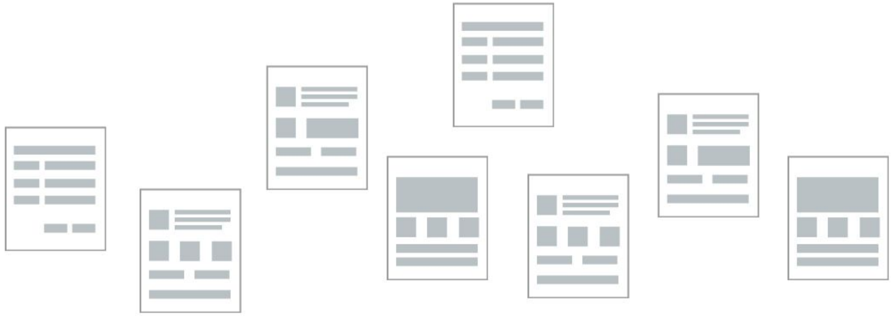
Classification vs Clustering - What's the difference?

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

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

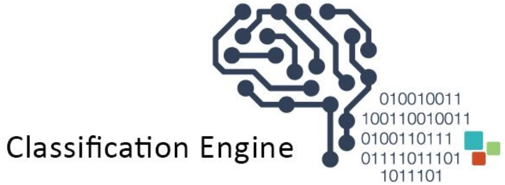


How important is classification, really?



Content understanding

quickly understand what topics the site is covering with the content



Content Gap analysis

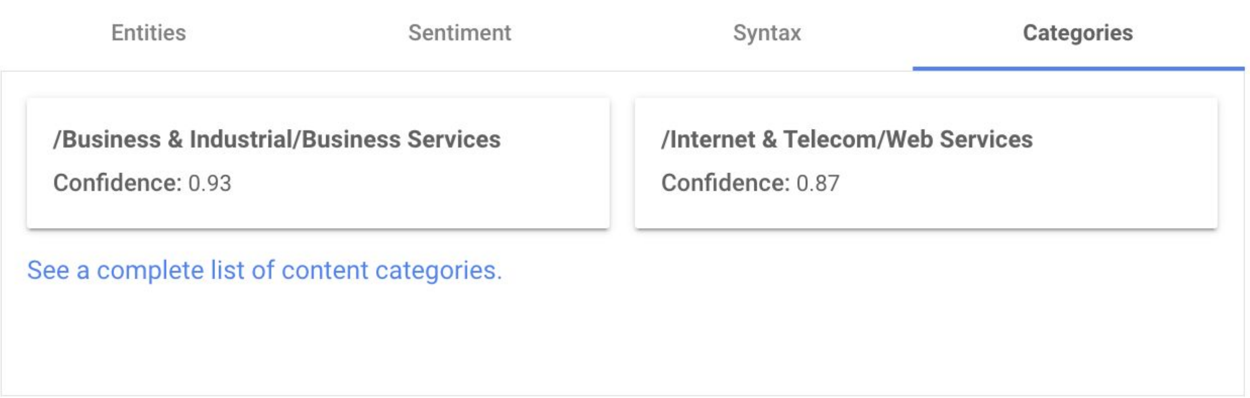
check whether the content topics on the site aligns with the business direction desired

Competitor analysis

quickly understand the topics that competitors' content talks about

With Google's Natural Language API, you can **classify** documents in **700+** predefined categories

(out of the box, can be custom-trained, too with AutoML)



With Open AI's GPT-3 or with ChatGPT, you can do both, but results are a hit or miss.

Choose a genre category for each book 1. The Hunger Games, 2. The Kite Runner 3. A Wrinkle in Time ("fiction", "young adult", "science fiction", "fantasy", "other") and make a list of the book and its genre:

```
1. The Hunger Games: young adult, fiction
2. The Kite Runner: fiction, young adult
3. A Wrinkle in Time: science fiction, fantasy, other
```

SMXL MILAN

Can you guess what can go great? (and what - horribly wrong)

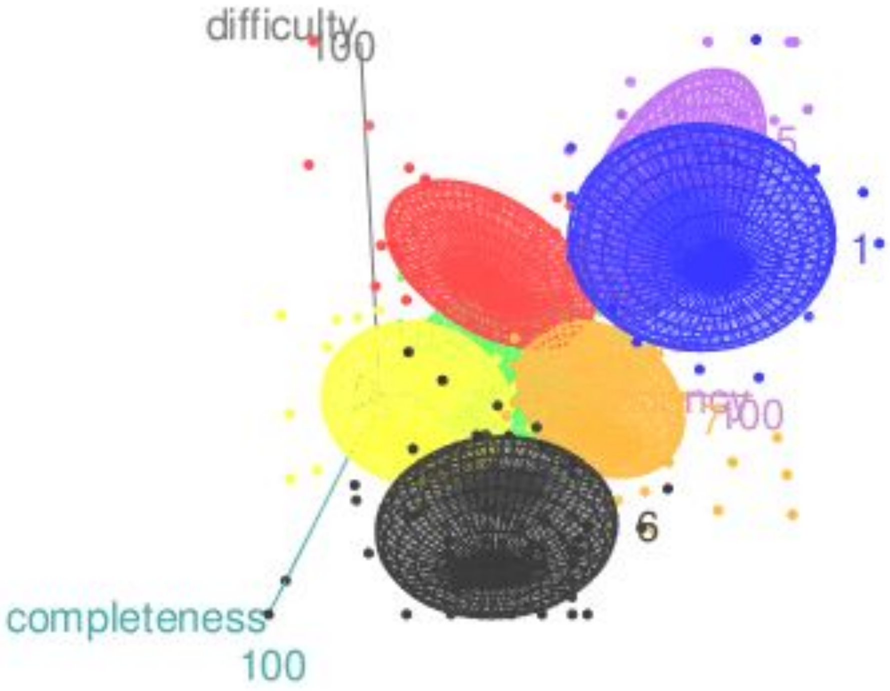
- ✅Predictable categories
- ✅Controlled training of model
- ✅Accuracy indicated
- ✅Great for scale and beckmarking

- ✅Can map the information to a label or assign a plausible such, provided it has this information in its training set
- ✅Very adaptive
- ✅Great for small projects, one offs

- ❌ Can't be used for uses outside of the main task
- ❌ Can't be given custom lists (...unless)
- ❌ Requires time and data for custom training models with AutoML

- ❌ Non-predictable results
- ❌ Direction might not followed
- ❌ Model not trained for task
- ❌ Limited knowledge
- ❌ Unsuitable for niche industries

What about clustering?



Topic understanding

quickly understand what topics the keyword universe you have consists of

Keyword clustering

quickly understand how other parameters of keyword research relate to the clusters identified the goal:

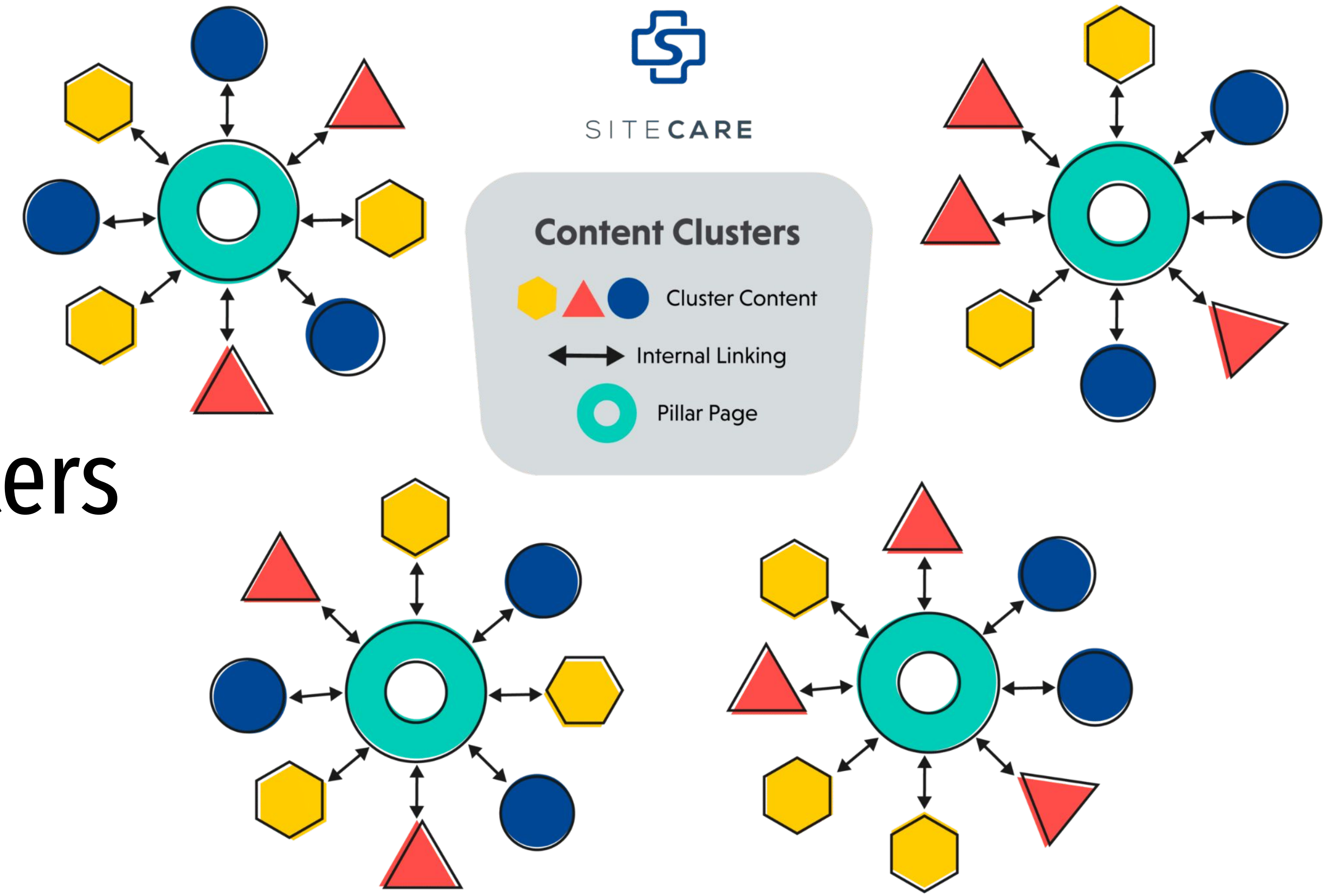
see your content and links

the way search engines see it

to understand how to improve it.

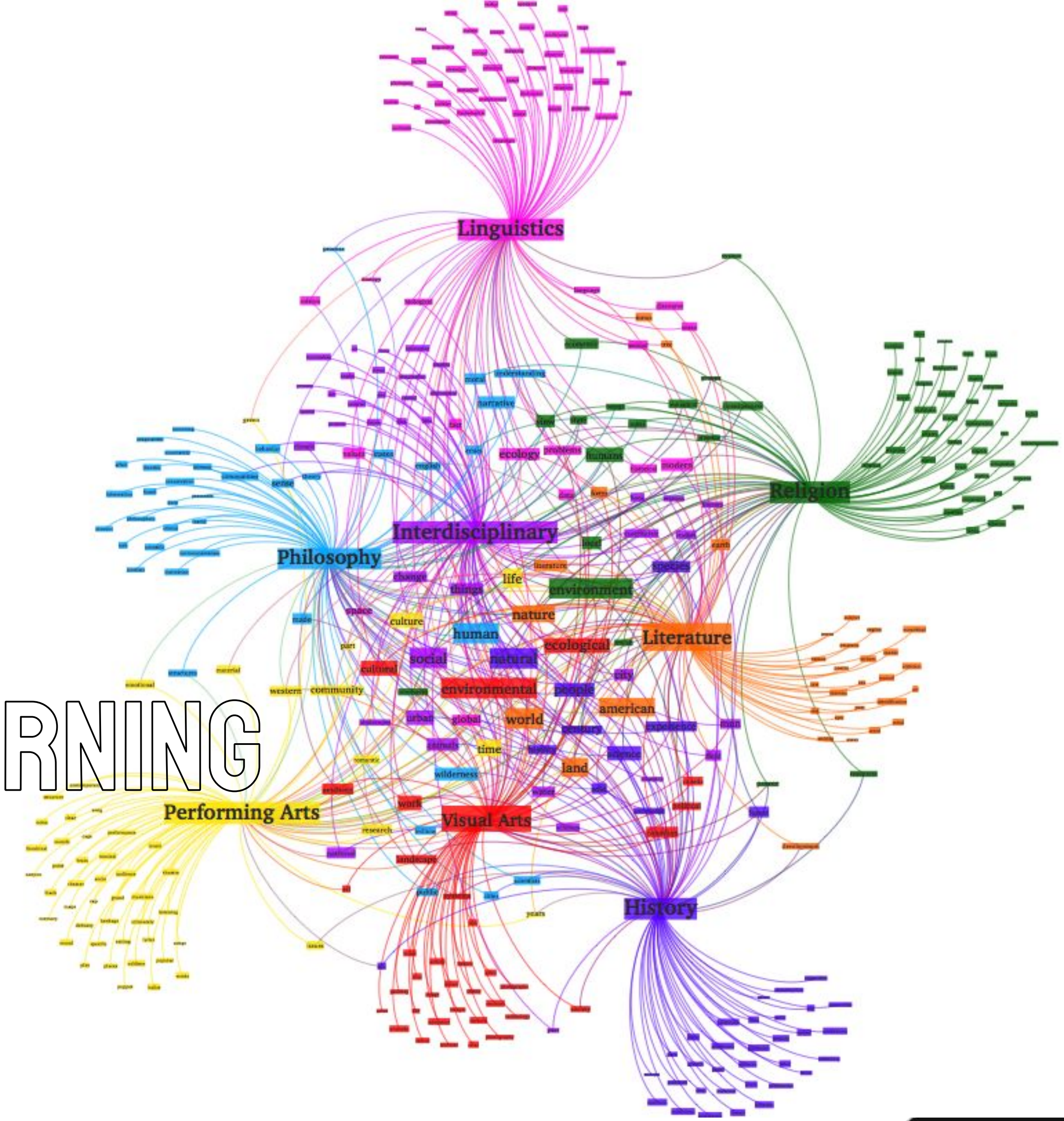
SMXL MILAN

Content clusters
IN SEO



SMXL
MILAN

Topic Models
IN MACHINE LEARNING



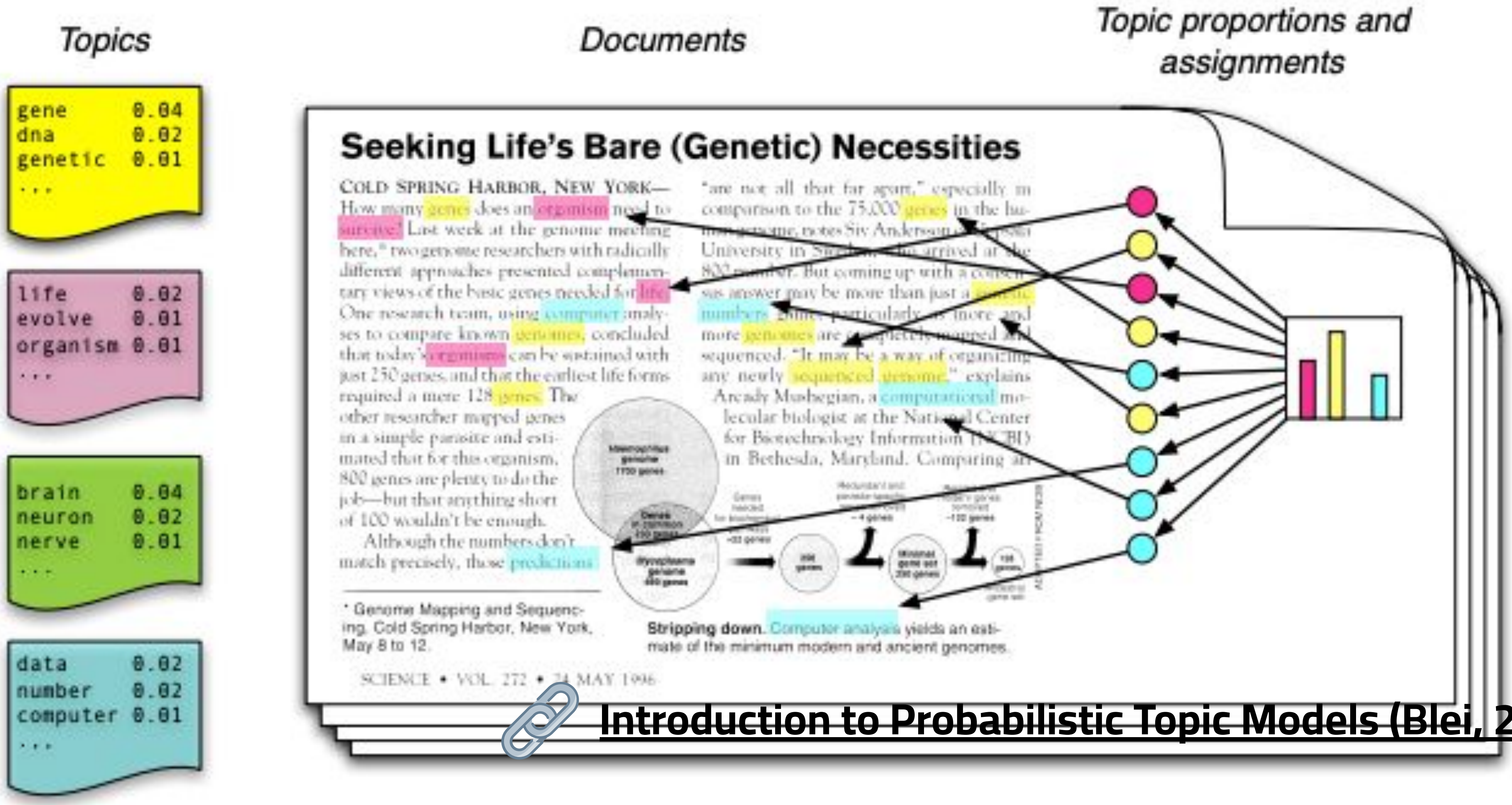
LDA emerged to:

- **remove dependency on links** and introduce the "things" concept and topic/term understanding
- introduce **computational understanding** of topics and terms and their importance
- highlight the assumption that **each page will have multiple different topics or subtopics addressed**, which might be of value to different people and should be understood and surfaced in results

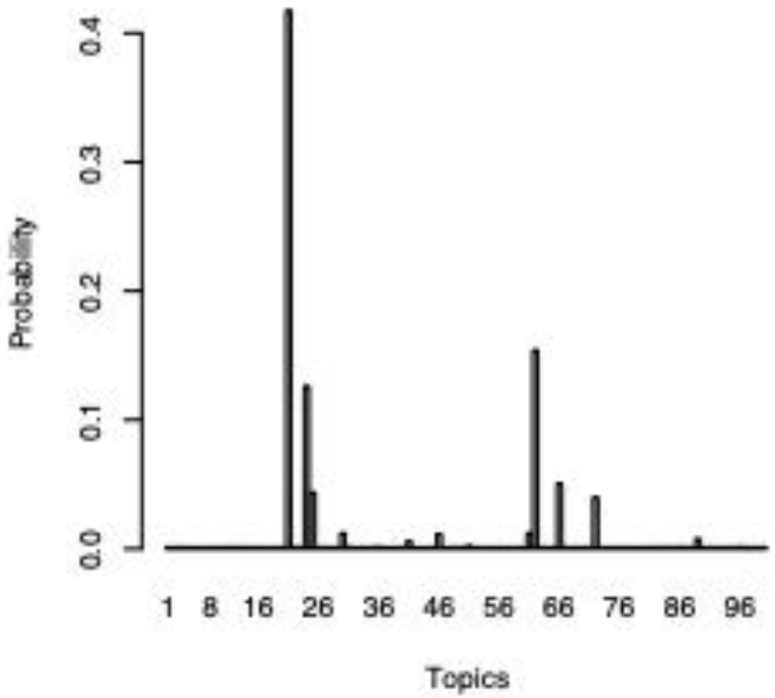
Needless to say, the field has evolved a lot since then!

TOPIC MODELING IS
PATTERN RECOGNITION
IN LARGE, TEXT-BASED CORPUSES OF DATA.

SMXL MILAN



🔗 [Introduction to Probabilistic Topic Models (Blei, 2012)]



"Genetics"	"Evolution"	"Disease"	"Computers"
human	evolution	disease	computer
genome	evolutionary	host	models
dna	species	bacteria	information
genetic	organisms	diseases	data
genes	life	resistance	computers
sequence	origin	bacterial	system
gene	biology	new	network
molecular	groups	strains	systems
sequencing	phylogenetic	control	model
map	living	infectious	parallel
information	diversity	malaria	methods
genetics	group	parasite	networks
mapping	new	parasites	software
project	two	united	new
sequences	common	tuberculosis	simulations

Introduction to Probabilistic Topic Models (Blei, 2012)

SMXL MILAN

WATCH the details LATER

I've recorded a step-by-step tutorial on doing **topic modelling** using a no-code, publicly-available, web-based app using LDA.

IDENTIFY TOPICS

CRAWL

Export the content from the website

UPLOAD + FINETUNE

Upload the files to the web app + finetune

DOWNLOAD

Download all files

EXPLORE + BUILD

Explore the outcome and build your deliverable

Topic models

Topic models

A	B	C	D	E	F	G	H	I	J	K
	health mental staff workplace home support group working ehl students	corporate volunteering giving purpose social responsibility grawehr stéphanie	nonprofit volunteers time content media support volunteer nonprofits share form	alaya platform data user services users policy information conditions general	csr business social companies strategy responsibility corporate initiatives	volunteering program employees corporate giving programs matching benefits	people company time it's back feel start make mission that's	employees impact engage purpose activities make platform community	engagement employees work team engaged good teams find make virtual	nonprofits season donors carmen amell nonprofit fundraising make strategy story
[illegible]	0.00%	17.65%	-44.77%		-16.00%	-104.73%	-4.67%	9.29%	-5.99%	
[illegible]	17.65%	0.00%	-21.13%	-88.64%	-8.61%	-62.86%	-36.37%	-23.42%	27.18%	6.80%
[illegible]	-44.77%	-21.13%	0.00%	9.88%	-26.01%	-58.78%	12.52%	16.87%	-56.55%	70.50%
[illegible]		-88.64%	9.88%	0.00%	-30.66%		-157.96%	-53.92%		
[illegible]	-16.00%	-8.61%	-26.01%	-30.66%	0.00%	60.18%	54.64%	26.11%	-0.59%	-218.54%
[illegible]	-104.73%	-62.86%	-58.78%		60.18%	0.00%	35.22%	30.17%	-0.59%	-77.01%
[illegible]	-4.67%	-36.37%	12.52%	-157.96%	54.64%	35.22%	0.00%	28.20%	27.18%	-46.26%
[illegible]	9.29%	-23.42%	16.87%	-53.92%	26.11%	30.17%	28.20%	0.00%	-1.53%	-40.31%
[illegible]	-5.99%	27.18%	-56.55%		-0.59%	-0.59%	27.18%	-1.53%	0.00%	-29.35%
[illegible]		6.80%	70.50%		-218.54%	-77.01%	-46.26%	-40.31%	-29.35%	0.00%

Topic to Topic Similarity ▾ Topic Modelling per Page ▾

Topic to Topic Similarity ▾

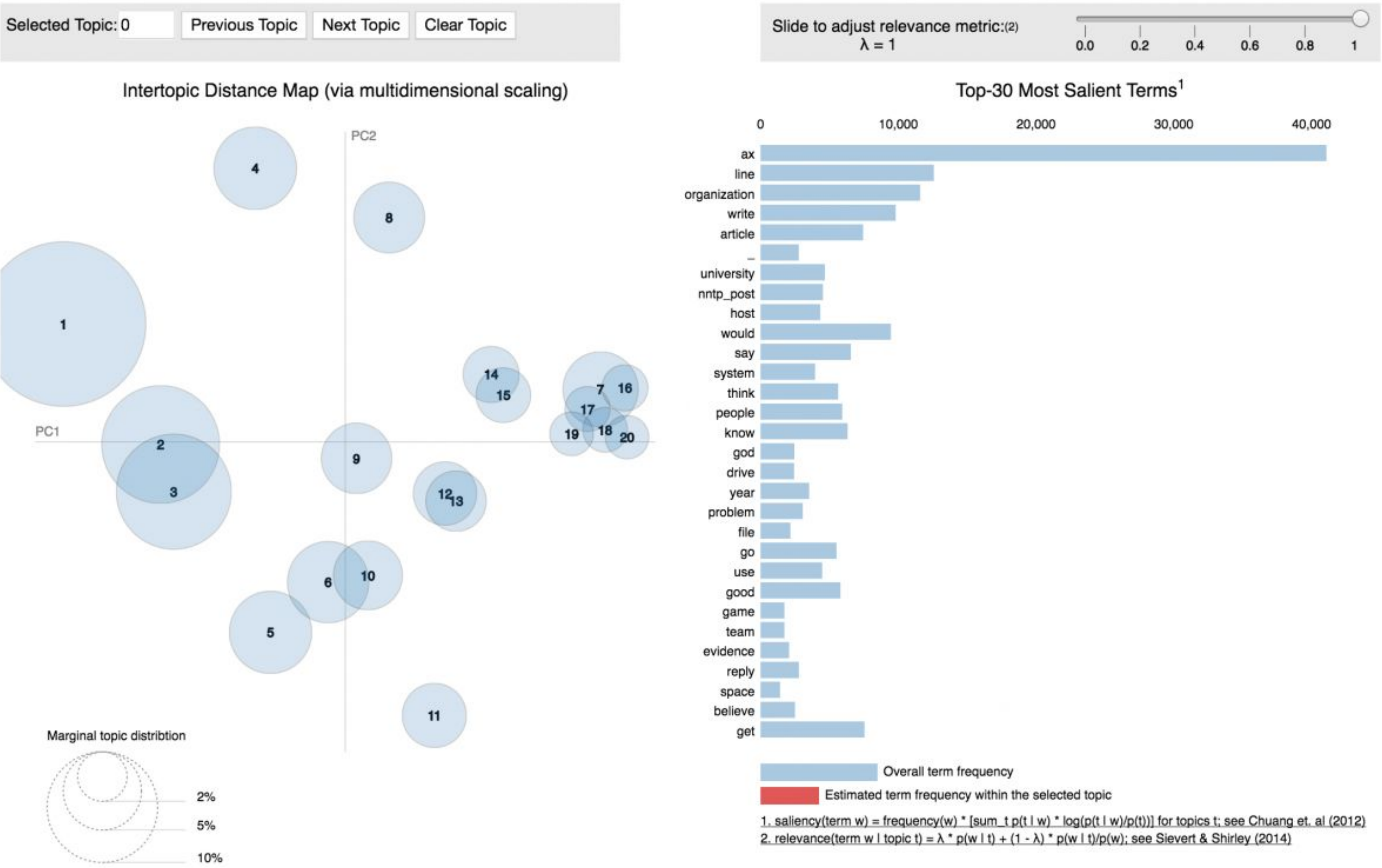
Page Info

Topic models

Content Export	Address	normalised title										
			0.00%	0.00%	50.00%	0.00%	0.00%	50.00%	0.00%	0.00%	0.00%	0.00%
			0.00%	0.00%	1.49%	0.37%	0.00%	20.07%	1.86%	33.09%	0.00%	0.00%
			1.33%	0.00%	2.00%	0.36%	5.74%	18.55%	6.20%	10.51%	7.64%	0.00%
			0.00%	0.00%	0.00%	0.00%	0.60%	14.83%	1.40%	38.48%	1.20%	0.00%
			0.00%	48.53%	0.74%	0.00%	0.00%	0.00%	0.00%	0.00%	8.09%	4.04%
			0.00%	0.00%	9.38%	0.00%	0.00%	0.00%	0.00%	3.13%	0.00%	53.13%
			0.00%	3.51%	15.59%	2.10%	0.86%	2.73%	6.63%	5.22%	0.00%	8.96%
			2.99%	0.48%	4.31%	1.08%	4.55%	1.20%	7.19%	6.23%	15.81%	0.96%
			6.50%	3.58%	3.17%	0.41%	6.81%	7.16%	7.98%	4.35%	15.05%	0.10%
			9.84%	7.81%	9.12%	1.74%	3.91%	0.00%	5.64%	9.41%	2.03%	0.00%
			0.09%	1.29%	0.76%	0.09%	25.68%	5.38%	8.90%	5.47%	4.09%	0.00%
			4.12%	1.17%	9.48%	1.58%	1.37%	2.68%	5.36%	4.53%	0.76%	15.52%
			1.40%	0.97%	7.97%	2.37%	3.47%	4.50%	5.36%	4.81%	1.58%	7.61%
			0.00%	2.53%	27.09%	2.61%	2.61%	0.00%	3.15%	6.45%	2.23%	8.21%
			0.00%	49.14%	0.00%	0.00%	0.00%	0.00%	0.69%	0.00%	9.62%	5.84%
			0.00%	5.18%	8.83%	6.91%	7.49%	1.92%	4.41%	15.36%	0.58%	0.58%
			0.00%	0.00%	43.24%	0.00%	0.00%	0.00%	2.70%	0.00%	0.00%	18.92%
			0.00%	19.63%	8.89%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	32.59%
			0.00%	62.77%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
			7.07%	51.52%	0.00%	0.00%	0.00%	0.00%	0.34%	0.00%	5.39%	1.68%
			4.34%	1.42%	5.54%	0.78%	2.35%	0.28%	10.73%	3.98%	0.57%	13.15%
			0.00%	54.29%	2.86%	2.86%	2.86%	0.00%	0.00%	0.00%	0.00%	0.00%
			1.93%	4.16%	12.01%	1.97%	6.22%	13.46%	4.72%	4.25%	2.19%	0.00%
			2.59%	4.25%	3.22%	0.00%	5.97%	3.14%	7.63%	12.74%	12.50%	0.39%
			2.03%	1.92%	1.05%	0.17%	23.55%	8.31%	14.42%	2.56%	3.43%	0.00%

Search Intent Matching | Topic to Topic Similarity | Topic Modelling per Page | Other Linking Opportunities (3N | Cluster 1 | Cluster 2 | Cluster 3

Topic Modelling per Page


Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Marginal topic distribtion
2%
5%
10%
Top-30 Most Salient Terms[1]
0
10,000
20,000
30,000
40,000
ax
line
organization
write
article
_
university
nntp_post
host
would
say
system
think
people
know
god
drive
year
problem
file
go
use
good
game
team
evidence
reply
space
believe
get
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

for less than 30 minutes:

- baseline overview of topics and relationships
- understanding of main terms that define topics
- understanding of term and topic dominance per page

What you can use this for
(literally, you're only limited by imagination here)

- Categorise/ discover patterns and topics on site content
  - → Identify opportunities for internal linking
  - → identify what your site is about and whether it aligns with business positioning
  - → Identify the topics that your competitor site tackles
- Categorise/ discover patterns and topics on YouTube titles or video catalogs
  - → Quickly understand competitive landscape in hundreds or thousands of videos in a niche
- Categorise/ discover patterns and topics in first-party data (any kind of user forms)
  - → Quickly see what topics your feedback is centred upon

SMXL MILAN

KEYWORD CLUSTERING IS
SUB-TOPIC KEYWORD EXTRACTION
IN TEXT-BASED DOCUMENTS.

KeyBERT

SMXL MILAN

Although there are already many methods available for keyword generation (e.g., Rake, YAKE!, TF-IDF, etc.) I wanted to create a very basic, but powerful method for extracting keywords and keyphrases. This is where **KeyBERT** comes in! Which uses BERT-embeddings and simple cosine similarity to find the sub-phrases in a document that are the most similar to the document itself.

SMXL MILAN

2.1. Installation

Installation can be done using pypi:

```
pip install keybert
```

You may want to install more depending on the transformers and language backends that you will be using. The possible installations are:

```
pip install keybert[flair]
pip install keybert[gensim]
pip install keybert[spacy]
pip install keybert[use]
```

2.2. Usage

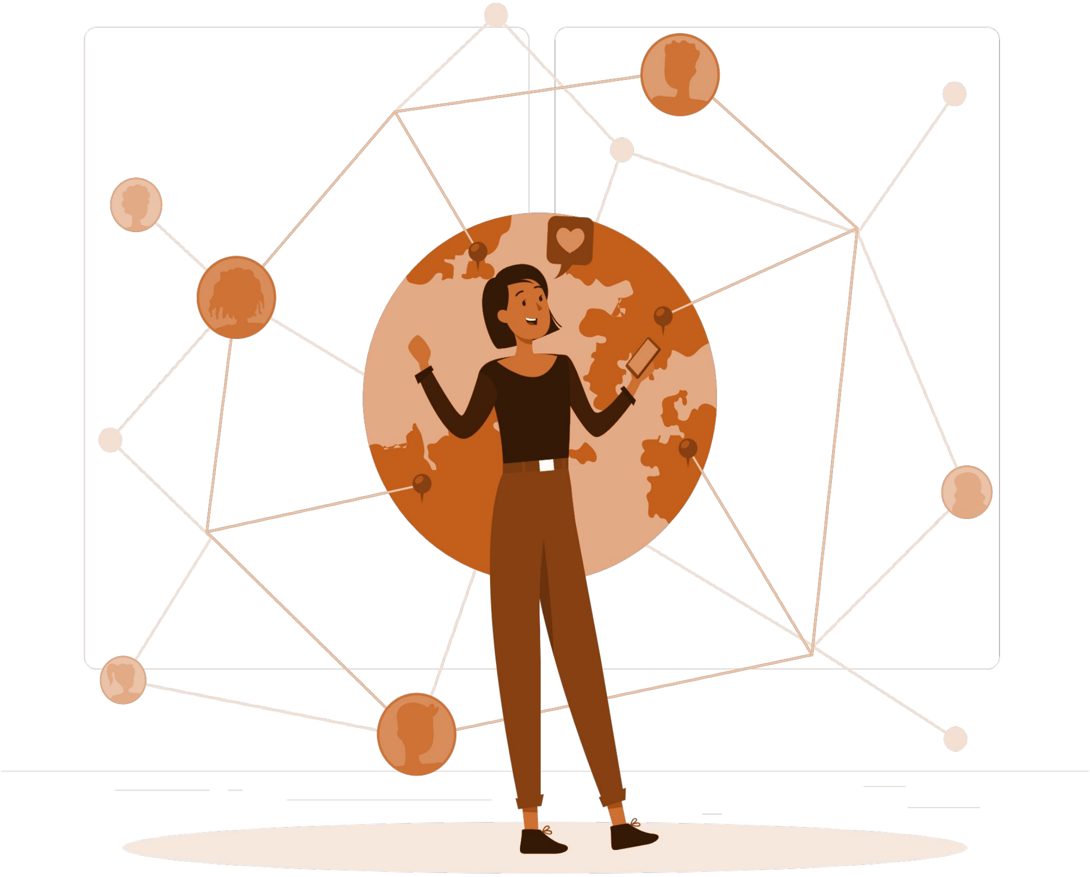
The most minimal example can be seen below for the extraction of keywords:

```
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 functic
         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)
```

great if you want to:

- do a **competitor-informed content strategy**
- **quickly understand the content** of a massive website

SMXL MILAN



let’s talk about

fuzzy matching

A quick and dirty way
for calculating
similarity between two strings

SMXL MILAN

ASSESS SIMILARITY

CRAWL + MODEL

Export the content from the website

ORGANIZE

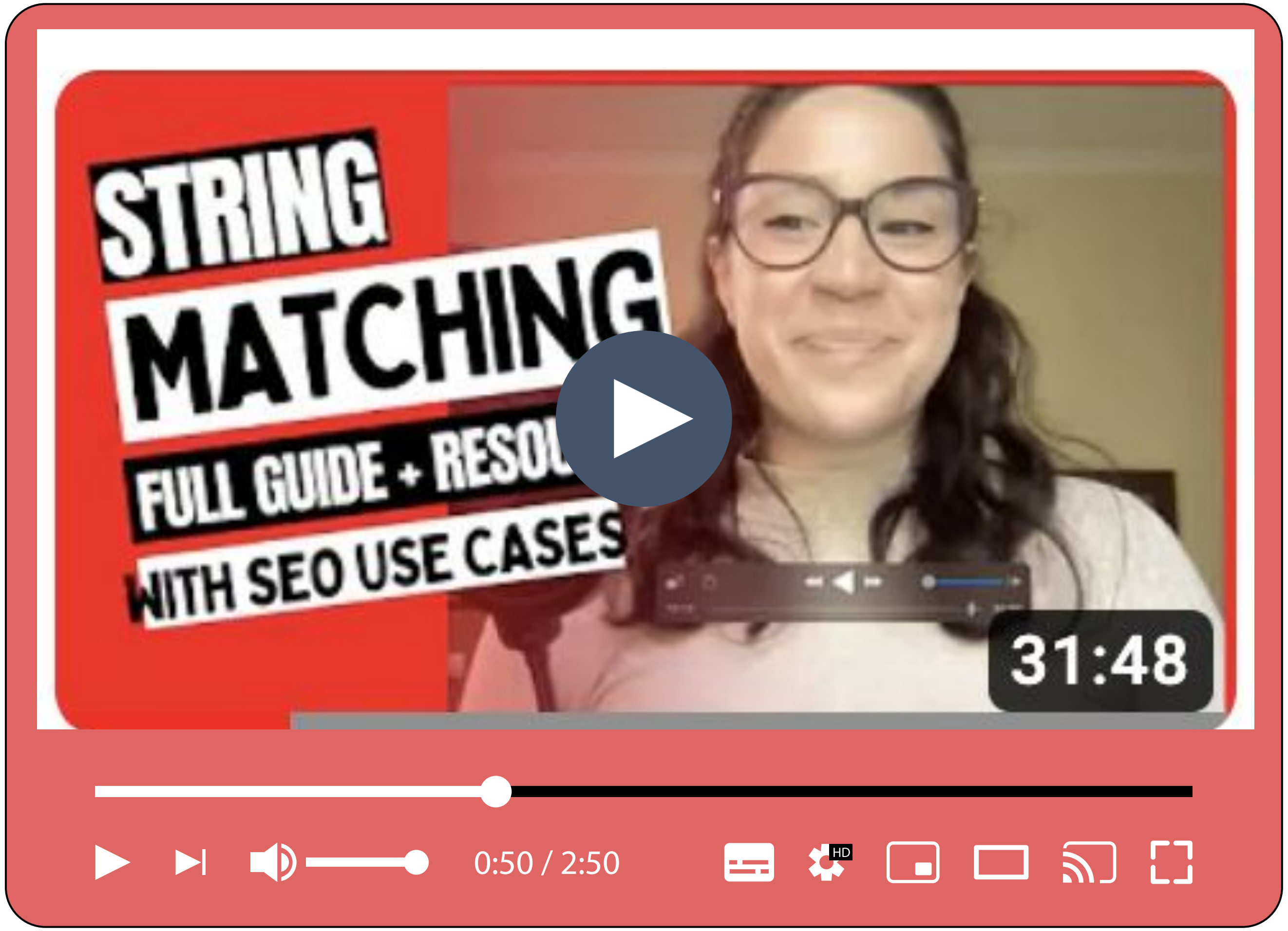
Choose what you want to assess for similarity - queries, anchor text, titles, paragraphs, URLs

ASSESS

Human-check similarity scores. Is the output good?

ACT

Implement or recommend



WATCH the details LATER

I've recorded a step-by-step tutorial on using **fuzzy matching** for things like:

- Identifying link opportunities
- String Similarity Analysis
- redirect mapping of URLs

✨ copy the template ✨

Google Sheets

Copy document

⚠ The attached Apps Script file and functionality will also be copied

Would you like to make a copy of **Fuzzy Matching for SEO Use Cases [Template by LazarinaStoy.com]**?

Make a copy | View Apps Script file

SMXL MILAN

let's talk about something a bit more advanced (but still beginner-friendly)

Content Transformation

Omni-presence and Accessibility

Both users and search engines want to se
multi-modal presence for high-value site

Meaning:

- Text to video
- Videos to text
- Text to audio
- Audio to text
- Text summaries for longer texts (e
headings)

"First of all we don't do text analysis of the videos and then map them to web pages. If your video has the same content as your blog post it's still something different. People sometimes go to Google with the intent to read something, and sometimes they go to Google with the intent to watch something or to listen to something, and those are very different things.

We wouldn't not say the text in this video is exactly the same as a blog post therefore we don't show either of them or we only show one of them. So if you have a video that matches your blog post I think that's perfectly fine.

That's a great way to spread your information in different channels. I would definitely not stop doing that. I would not take the video down or take the blog post down. If the blog post is not ranking in google then that would be very specific to the blog post and not specific to the video blog post combination."

SMXL MILAN

Dear Olivia,

The solar energy conference went great. New Horizon Manufacturing wants to meet with us to talk about our photovoltaic window system we're about to launch.
I think it would be great to talk this Tuesday.

Best,
Allison

Lengthen ✔

Dear Olivia,

I hope this email finds you well. I wanted to give you a quick update on the solar energy conference that I attended last week. It was a huge success! We received a lot of interest and positive feedback on our photovoltaic window system that we're about to launch.

In fact, I'm excited to let you know that New Horizon Manufacturing has expressed interest in meeting with us to discuss our product further. They were impressed with our technology and are eager to explore potential partnerships with us.

Given the great opportunity that this presents, I was hoping we could schedule a meeting with them as soon as possible. Specifically, I was thinking this Tuesday would work well. Please let me know if that works for you, and we can coordinate a time that works for all of us.

I'm looking forward to hearing back from you soon and hopefully, we can take advantage of this exciting opportunity to grow our business.

Best regards,
Allison

Shorten ✔

Allison is excited to meet with New Horizon Manufacturing to discuss their photovoltaic window system.

Answer ✔
(with some limitations)

Answer the following questions about this customer email:

I'm wondering if you could provide me with information about your cloud services for companies. I'm CTO for Moon Door, a movie production company and want to know if it can help us manage our different projects that are in production.

Questions:
1. What industry is the company
2. What is the name of the company
3. What are they interested in
4. What is the customer's position

Answers:

```
1. The company is in the movie production industry.
2. The company is Moon Door.
3. The customer is interested in cloud services for companies.
4. The customer is the CTO.
```

SMXL MILAN

Multi-modal Transformation ✔

Speech-to-Text

Speech recognition and transcription supporting 125 languages.

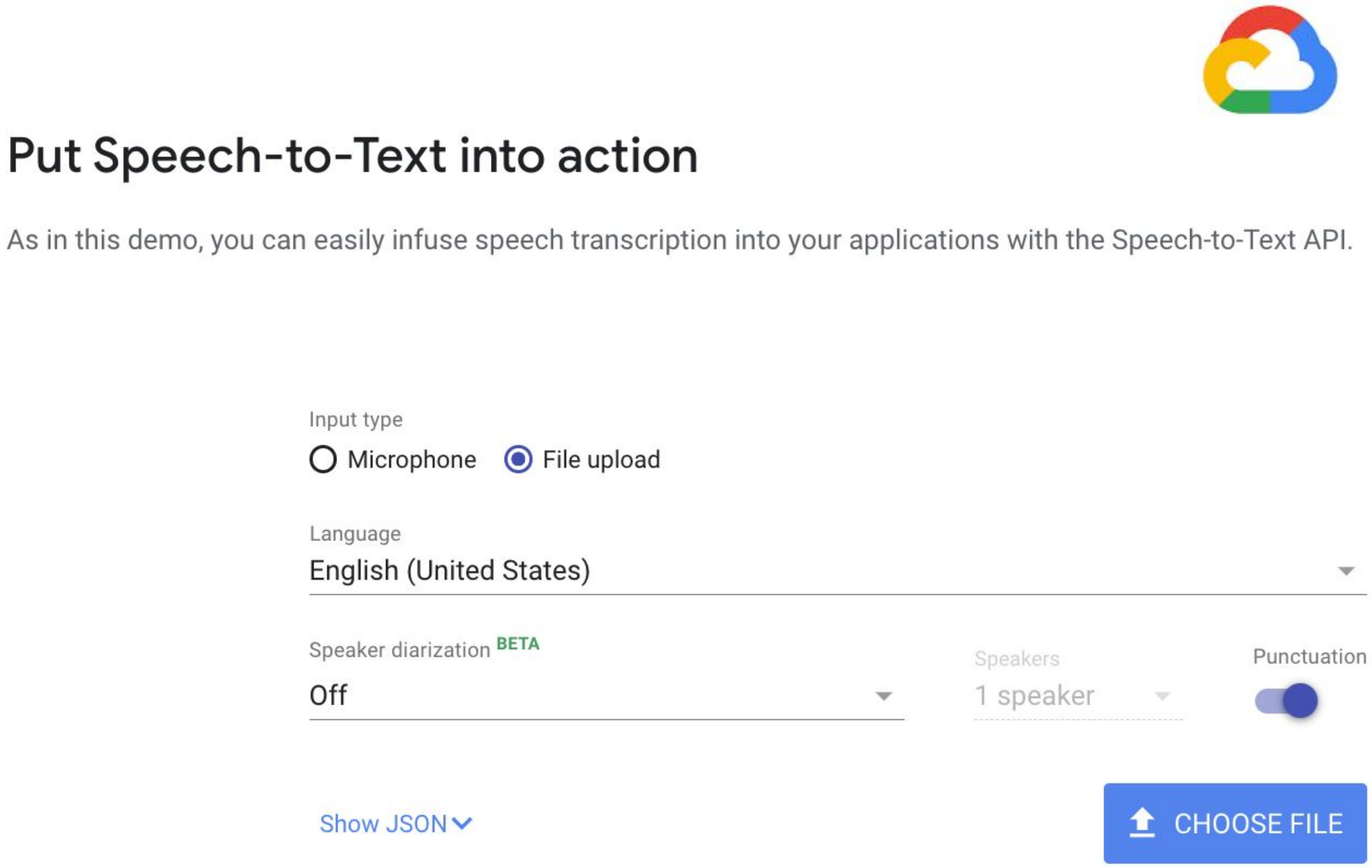
Text-to-Speech

Speech synthesis in 220+ voices and 40+ languages.

Needless to say scaling production here is pretty significant for organic growth.

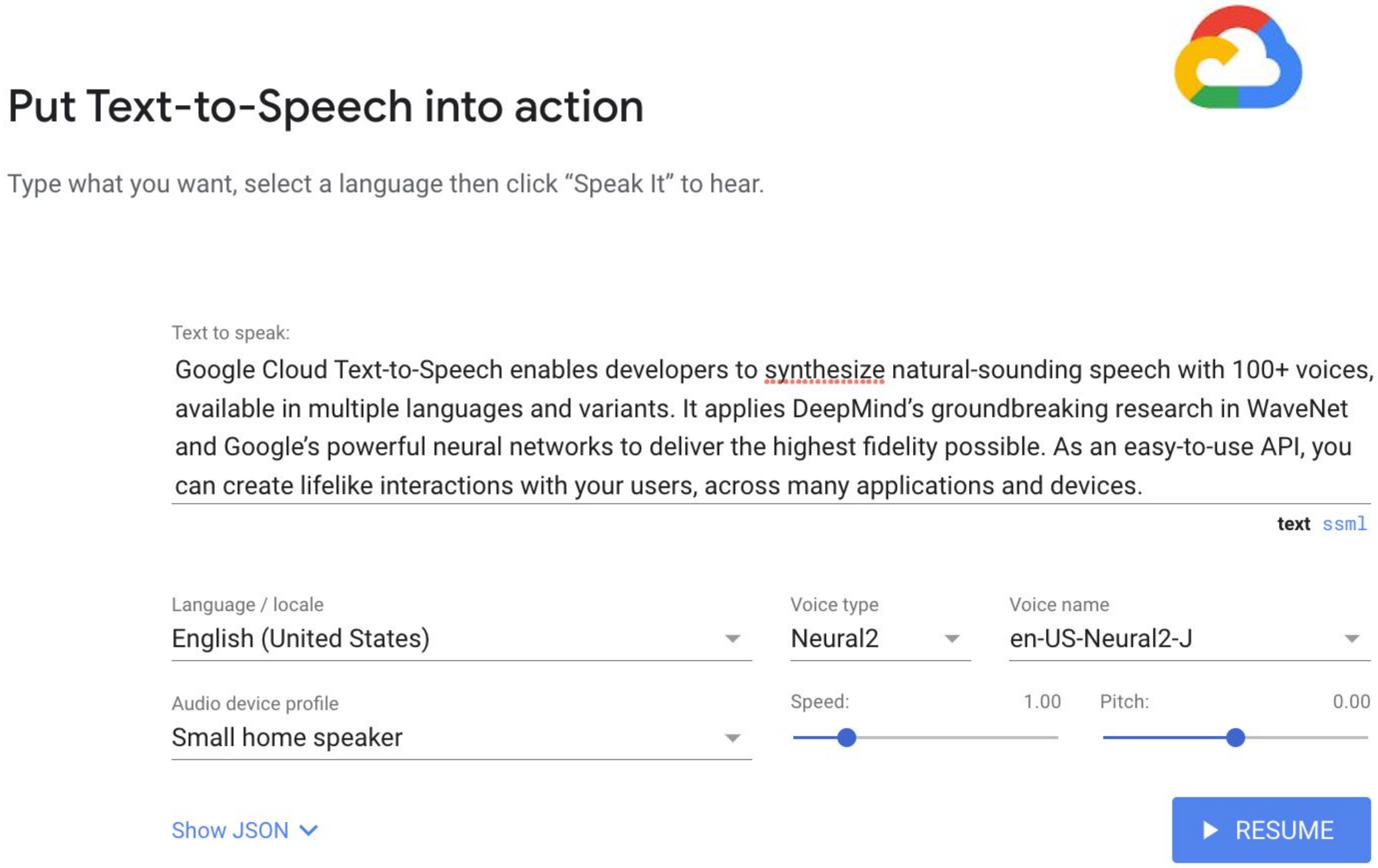
SMXL MILAN

You have a library of videos on YouTube but no blog?
→ **Scale it's launch quickly through transcription.**



You have a library of high-performing blog posts but no presence on YouTube/TikTok?

→**Scale production with text to speech.**



SMXL MILAN

Want to optimize for high-intent visitors or improve accessibility of content?

→ **Provide summaries of sections, improve headings, add FAQs, at scale.**

generate three possible frequently asked question and answer pairs from the following text: (...)

1. What is string matching in machine learning?
   String matching in machine learning is a process that involves measuring the similarity between two strings and calculating a score based on that. It can be used to identify patterns in a large corpus of data, or retrieve information based on a specified input. The goal is to either classify the strings as equivalent, similar, or distant.
2. What are the reasons for implementing string matching algorithms?
   There are two main reasons for implementing string matching algorithms: error correction and information retrieval. Error correction involves identifying patterns in a large corpus of data and correcting errors, while information retrieval is about providing an input that best describes the information we are trying to retrieve from the dataset.
3. How is similarity measured in string matching?
   In string matching, similarity is measured by calculating the distance between two strings. This can involve measuring errors in spelling and typing, such as mistaking a letter for another letter, omitting a letter, or inserting another by mistake. The goal is to approximate how alike the two strings are, and classify them accordingly.

half of the people in the room might be thinking:

" I got this."

the next part of the presentation is for those that are thinking:

"I could never."

“I am not technical enough”

"I don't know enough about ML to do these things"

SMXL MILAN

"I sucked at math in high school, so this just goes above my head"

“I simply don’t have the time to do these things”

SMXL MILAN

Recognise any of these?

Your limiting beliefs about what it takes to implement automation and ML in your tasks and processes **might be holding you back.**

Waiting To Get Started?

Search machine learning in 10 minutes. Follow along.

Start small but do something today.

Awaiting Perfect Conditions?

Build a habit and track your progress.

Start small and remain consistent.

Struggling or Tried and Failed?

Cut scope or change direction. **Start small** to get back into it.

Or, maybe you lack context?

SMXL MILAN

”

Like hearing:

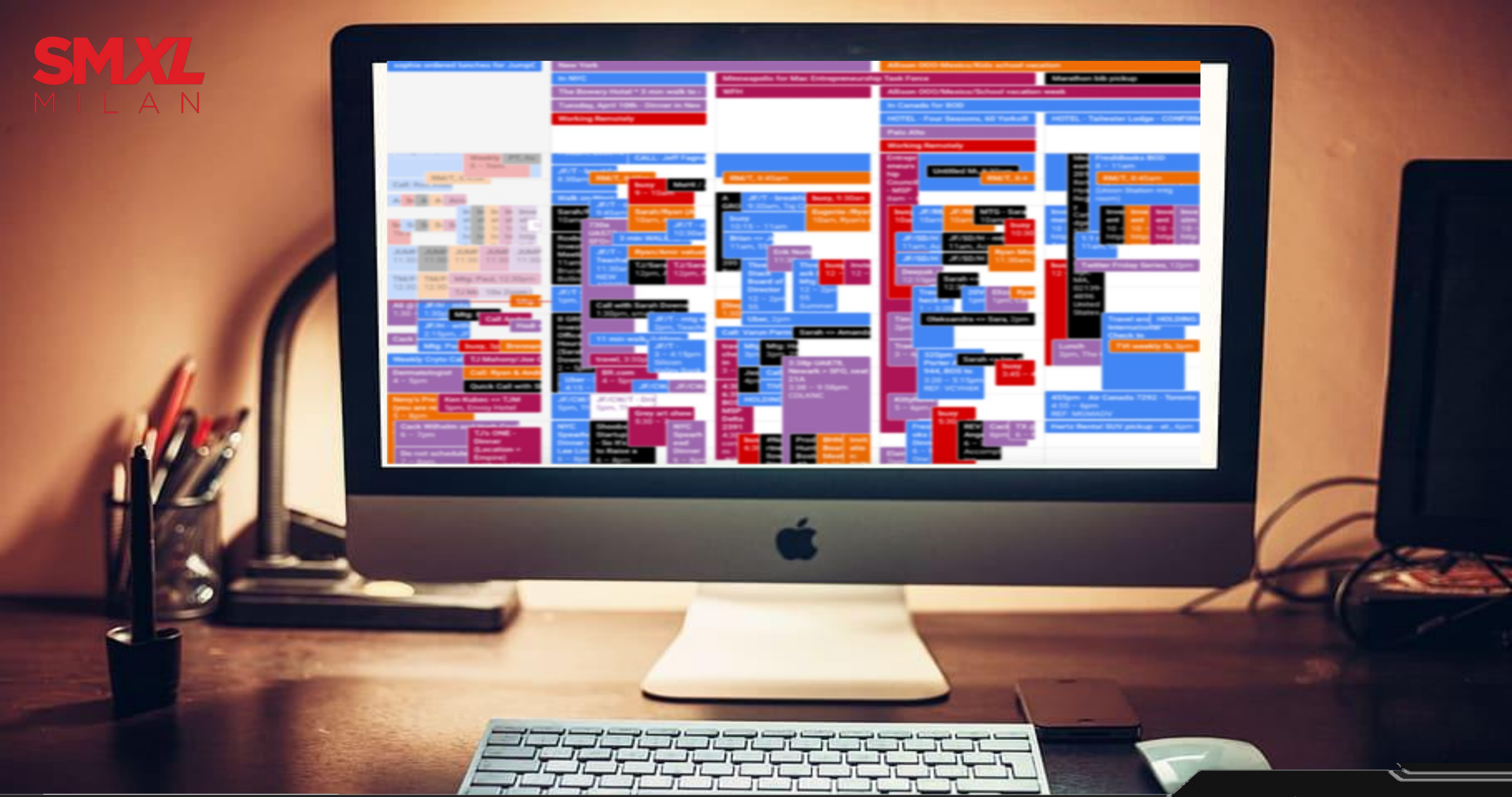
Just think about things to automate.

SMXL MILAN

”

and thinking:

I don’t even know what’s possible



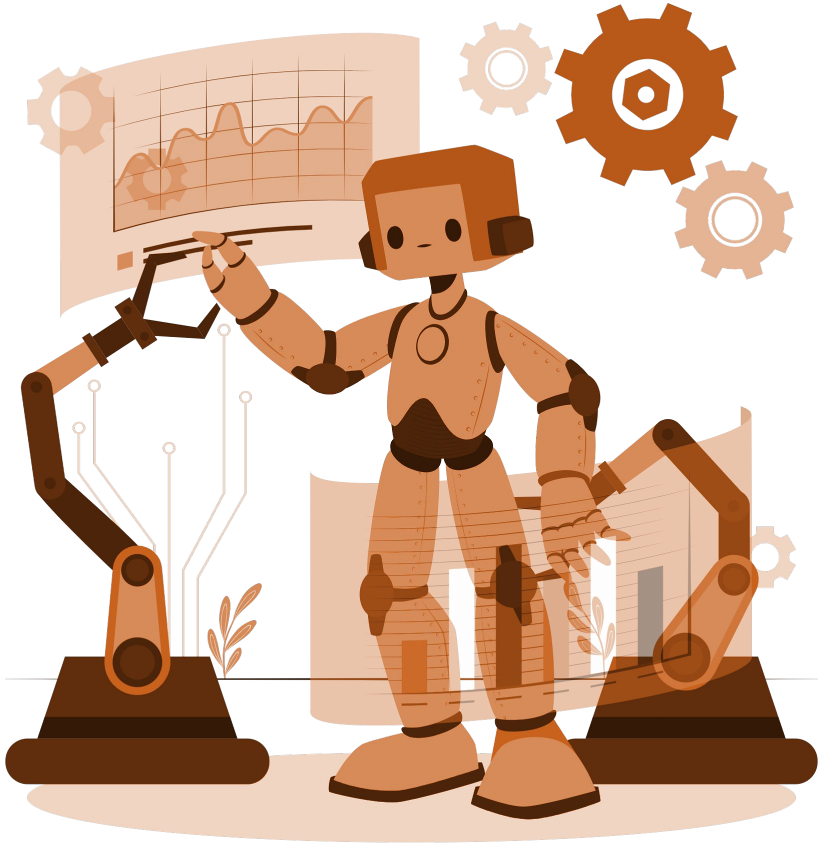
SMXL MILAN

Often, there is no need to reinvent the wheel, especially as **a beginner**.

SMXL MILAN

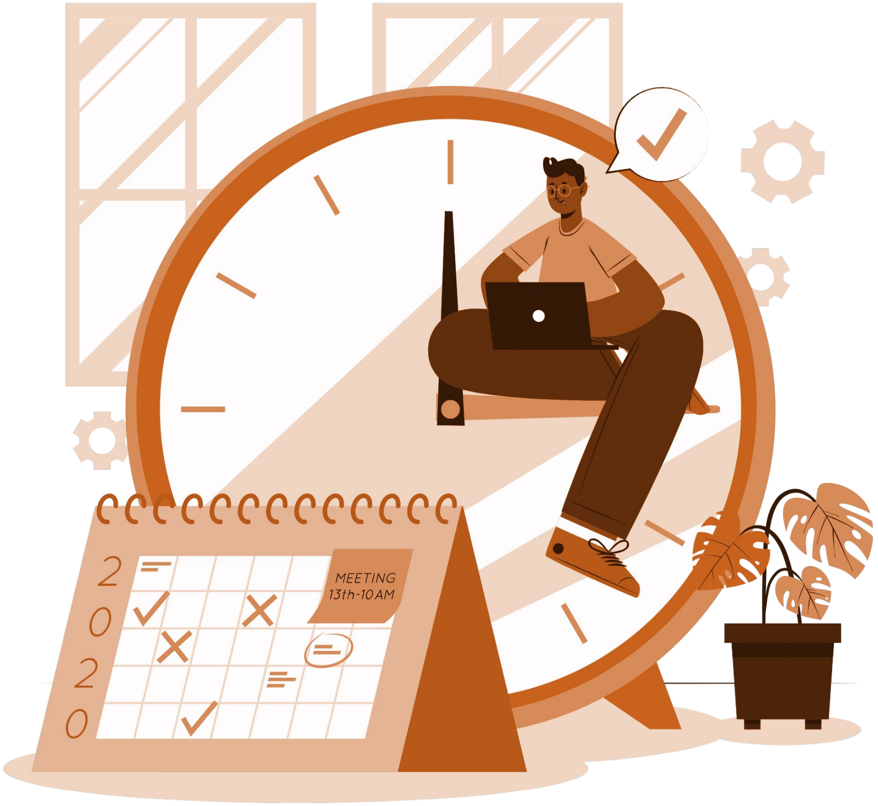
You don't need to know a ton of theory.

SMXL MILAN



You don't need end-to-end, custom-built, automated solutions.

SMXL MILAN



You don’t need to spend a ton of time.

Aim to **drive value** and **have fun** with the process.

SMXL MILAN

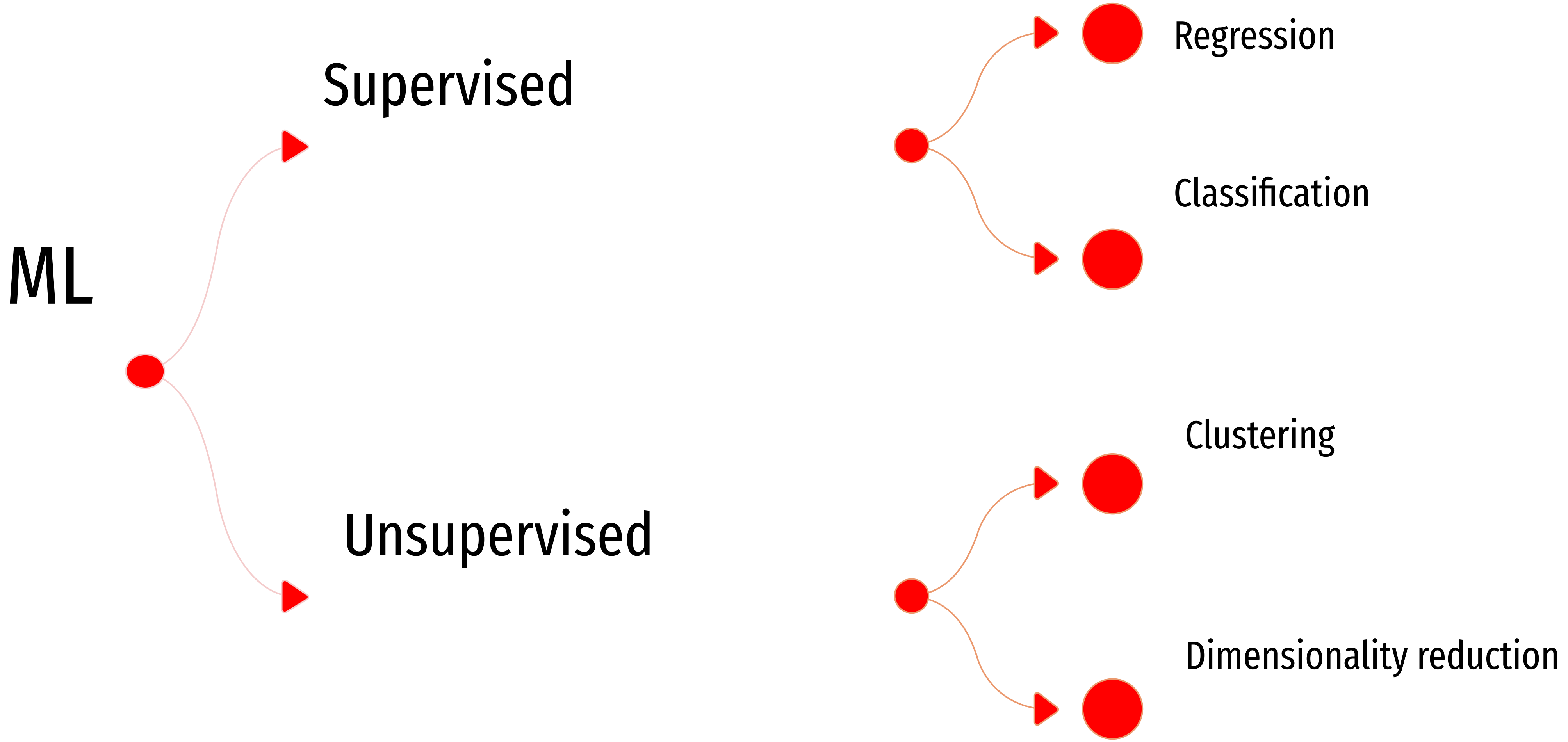
Knowing

- what model to use
- how to find and implement it quickly
- how to drive value via ML

is the perfect way to start.

Let's start with the basics.
For each potential project consider **three aspects**.

Task characteristics


ML
Supervised
Unsupervised
Regression
Classification
Clustering
Dimensionality reduction

SMXL MILAN

ML

You have labelled data to validate results

Make predictions

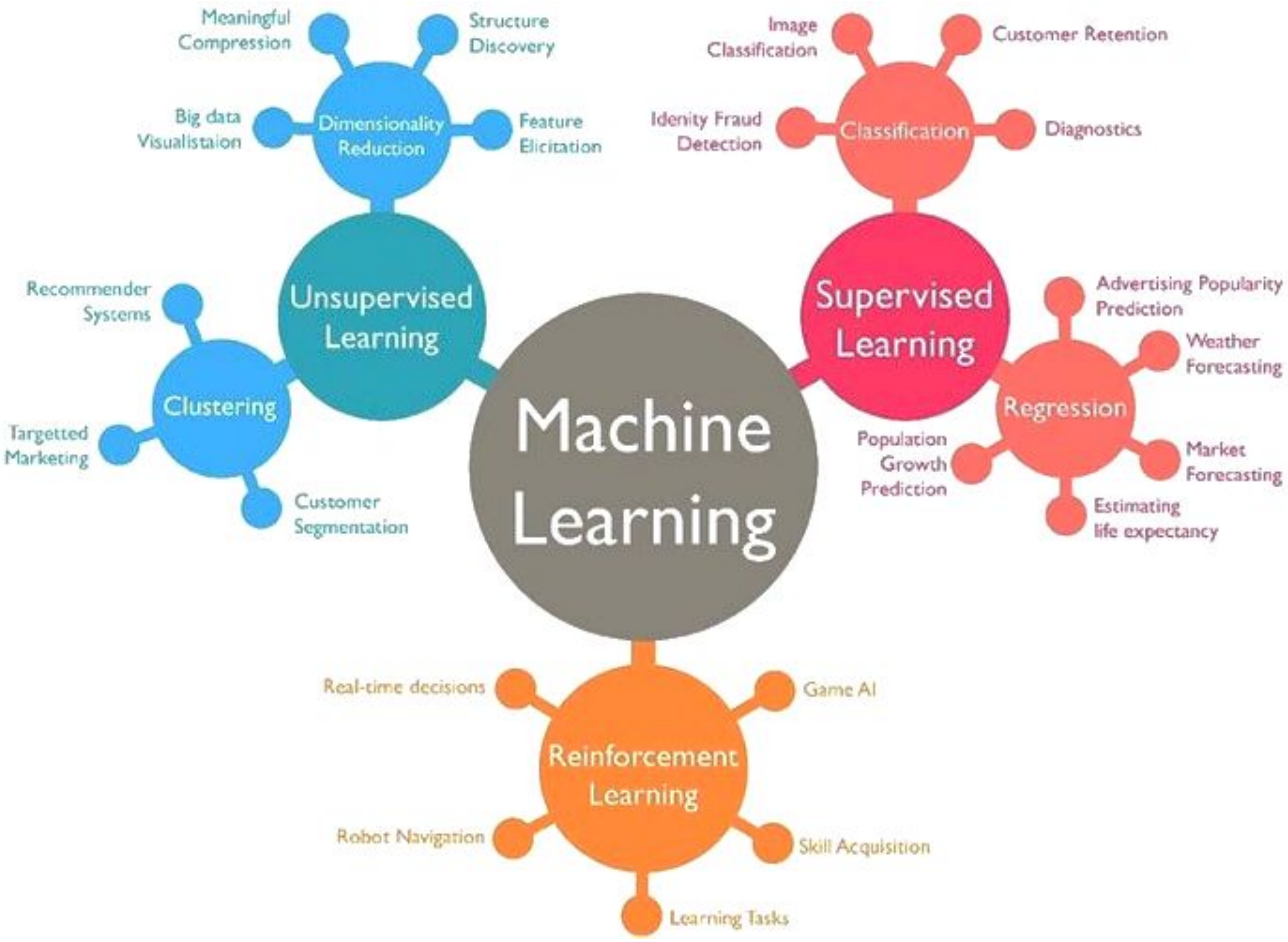
Split into groups, based on existing classes

You don't have a way to validate results

Find patterns and group based on similarity

Simplify or transform your data

SMXL MILAN



Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

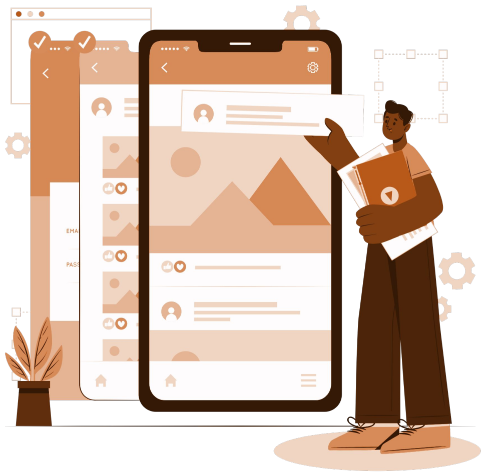
Data characteristics

Is your input data:

Textual

Numeric

Image-based



Solution characteristics

Is this task mission critical?

✅ Don't rely on AI.

(seriously)

❌

Must the results remain consistent every time?

✅ → Don’t rely on AI.

(yeah, really)

↓ ❌

Must the results remain easy to understand and relay to the stakeholder?

✅ Skip deep learning.

❌

Is it okay that simply on average the output outperforms existing methods?

✅ ➡

Okay, then.

Take a look at ML options.

SMXL MILAN

Assess using multiple factors:

- insights
- complexity
- accuracy
- scalability
- assets (data in = data out)
- Resources
- bottom line

Let's put this into context...

How to find suitable ML buddies (scripts, models, tools)?

SMXL MILAN

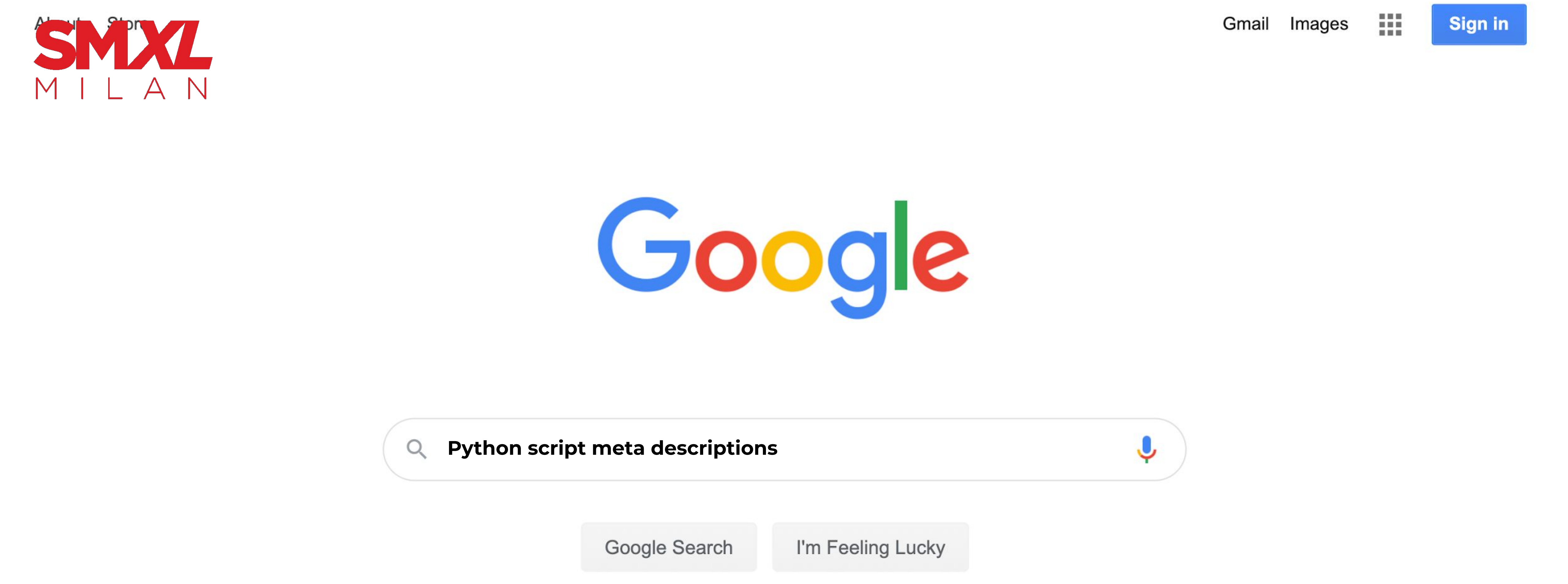
Keep **queries** **specific to data, task, and solution.**

SMXL MILAN

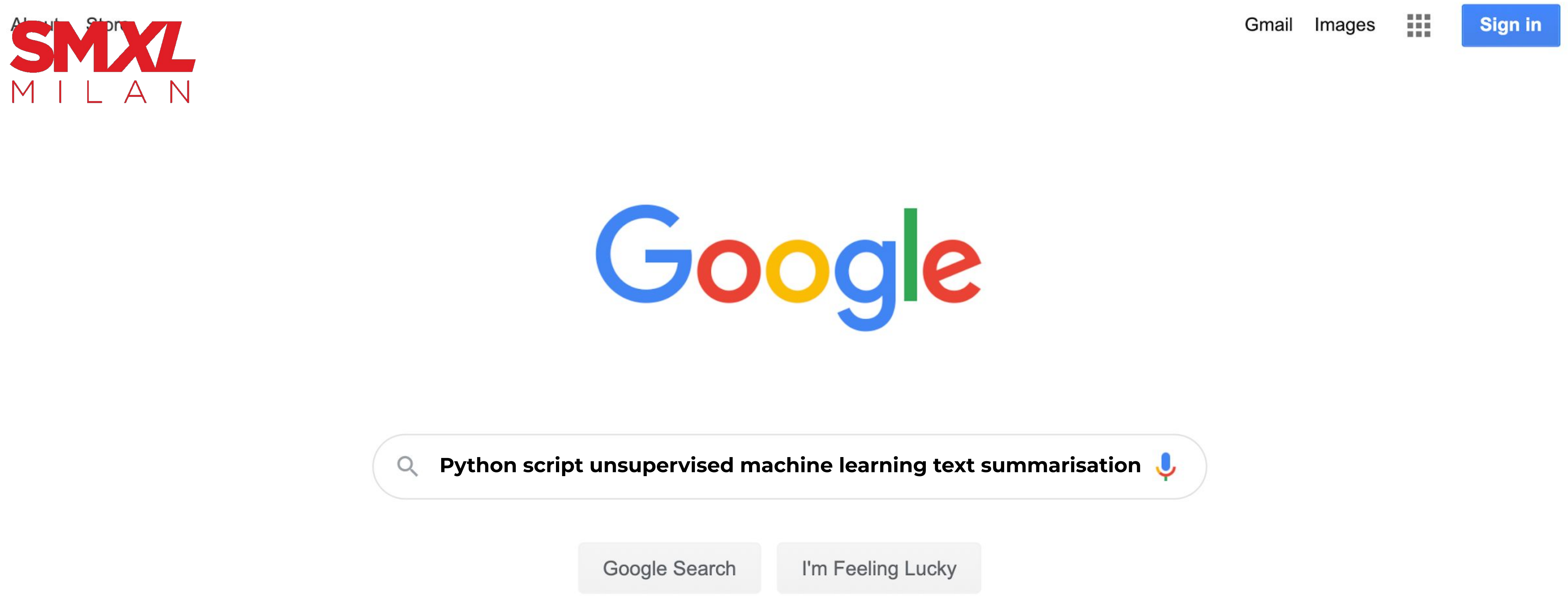
A typical on-page optimisation project might include mini-projects like...

1. Writing Meta Descriptions

- Input data? **Textual** (page content).
- Supervised or unsupervised? **Unsupervised**
    - It is **transformational** (Page content to Page Summary in less than 160 characters)
    - It can also be **generative** (write them from scratch)
- Is it mission critical? **No.**
- Different results okay? **Yes.**
- Explanation of process needed? **Not really.**
- Outperforms average methods? Yes, much **faster.**

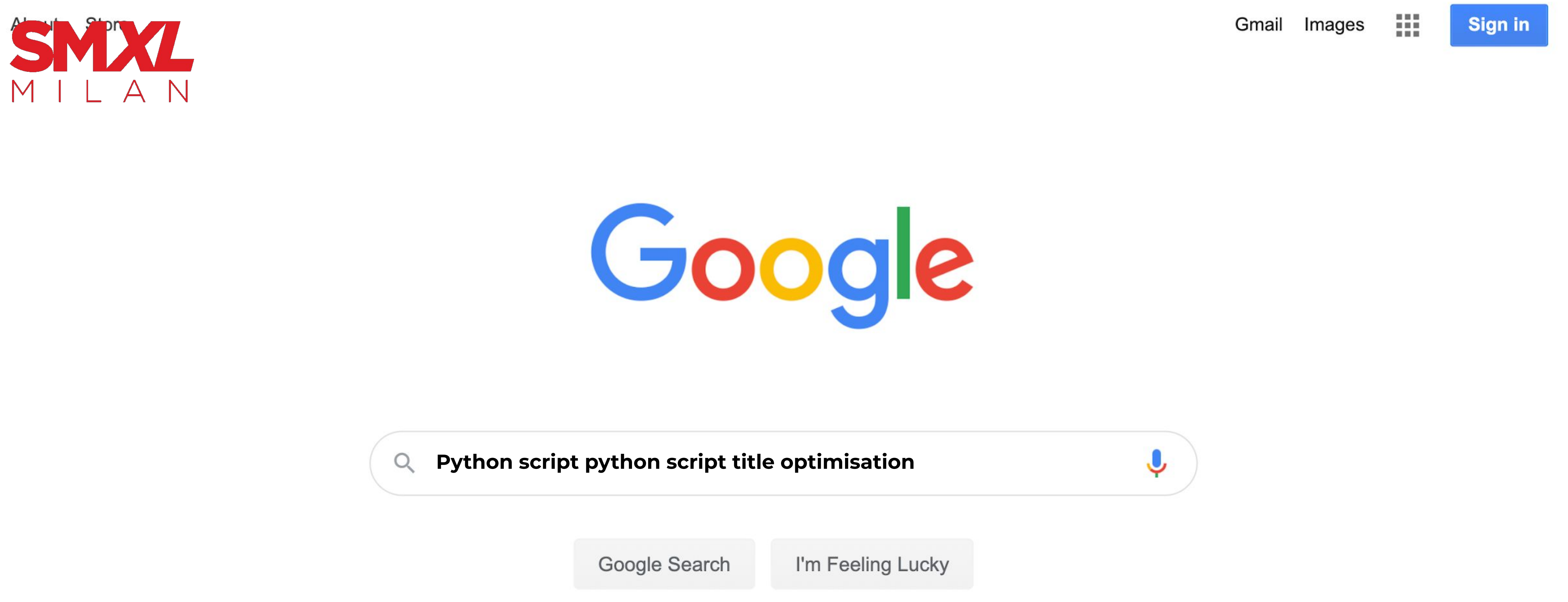

Gmail
Images
Sign in
Google
Python script meta descriptions
Google Search
I'm Feeling Lucky

SMXL MILAN


About Store
Gmail
Images
Sign in
Google
Python script unsupervised machine learning text summarisation
Google Search
I'm Feeling Lucky

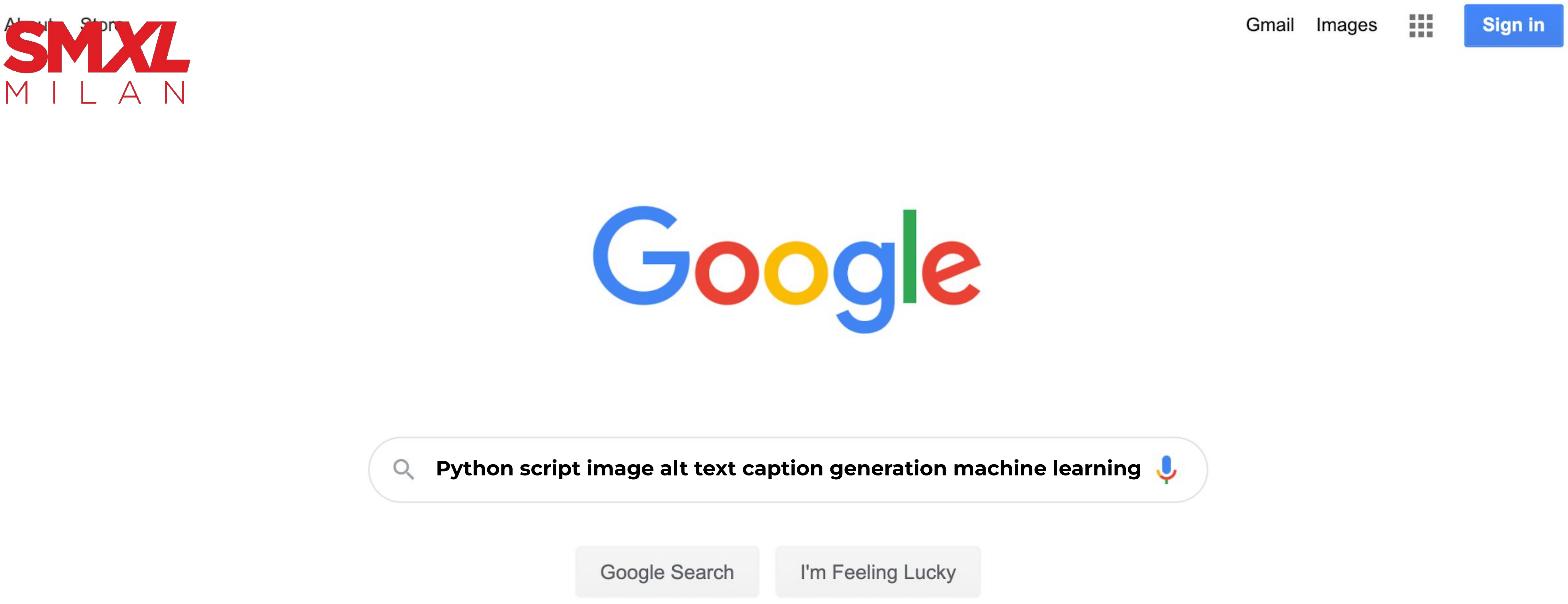
2.Title / H1 Optimisations

- Input data? **Textual** (page content).
- Supervised or unsupervised? **Unsupervised**
  - It is **transformational** (Page content to Page Summary in less than 60 characters)
  - It can also be **generative** (write them from scratch)
- Is it mission critical? **Hm, debatable (critical for YMYL)**
- Different results okay? **Again, debatable (critical for certain industries)**
- Explanation of process needed? **Kind of.**
- Outperforms average methods? Yes, much **faster. Not better.**


Gmail
Images
Sign in
Google
Python script python script title optimisation
Google Search
I'm Feeling Lucky

3. Image Alt tag Generation

- Input data? **Image**
- Supervised or unsupervised? **Unsupervised**
- **Generative AI / Image recognition**
- Is it mission critical? **No**
- Different results okay? **Yes.**
- Explanation of process needed? **Not really.**
- Outperforms average methods? Yes, much **faster.**


Gmail
Images
Sign in
Google
Python script image alt text caption generation machine learning
Google Search
I'm Feeling Lucky

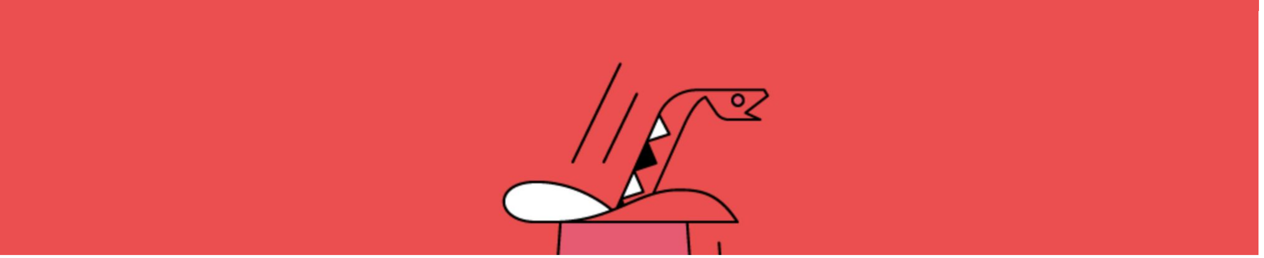
But might also involve...

- **Predicting traffic / revenue** based on presence of keyword in the title/ h1 to get buy in on proposed changes
- Updating **internal links**
- **Researching keywords** for new content updates
- Schema implementation

“Surely a script can’t do all that?!”

Not yet, anyway.

But a few can.



Predict SEO Organic CTR Based on Position Using Machine Learning

Know if your metatags and CTR% are good or bad.

Michael Van Den Reym Following
Dec 27, 2021 · 4 min read

Run 50 iterations Iterations: 0 Train with 25 topics

[0] government american federal peace war america country economic made security

[1] government america country war federal national american work peace united

[2] government federal american country america states made security peace national

[3] government war america nation states security made peace work national

Topic Documents | Topic Correlations | Time Series | Vocabulary | Downloads

Documents are sorted by their proportion of the currently selected topic, biased to prefer longer documents.

Use a different collection:
Documents Choose file No file chosen
Stoplist Choose file No file chosen
Upload

[1953-53/5.5%] Because the building of a completely impenetrable defense against attack is still not possible, total defensive strength must include civil defense preparedness. Because we have incontrovertible evidence that Soviet Russia possesses atomic weapons, this kind of protection becomes sheer necessity...

[1946-444/5.3%] Contract settlement and surplus property disposal.--The winding up of war procurement is the second most important liquidation job. By the end of November a total of 301,000 prime contracts involving commitments of 64 billion dollars had been terminated. Of

.P

String Matching with BERT, TF-IDF, and more!

Automate Ke with Google

Michael Van Den Reym
Jun 5, 2021 · 3 min read

SEO keyword research is time automating this process!

Charly Wargnier · Jul 25, 2021 · 1 min read

Generate FAQs for your pages automatically with What The FAQ!

Updated: Jul 26, 2021

Adding value doesn't necessarily mean a fully-automated, autonomous solution.

Incremental improvements can lead to a **compounding effect**.

SMXL MILAN

what I do my website

seo@lazarinastoy.com

social handle

best way to get in touch

Thank you for listening.