

AI Content Prediction for Large Websites

Mike Moran

June 2020



**Think of a company with an *awesome* customer experience.
When you go to their website, you get what you want.**

Was it...



amazon

The word 'amazon' is written in a bold, black, lowercase sans-serif font. Below the word is the iconic orange curved arrow, which starts under the 'a' and points towards the 'n'.

What's Amazon's secret?



Site Search

**When you've visited Amazon,
have you ever *not* searched?**

Related Products

**Customers who viewed this
item also viewed...**

Both techniques are *content prediction*



Site Search

**When you've visited Amazon,
have you ever *not* searched?**

Related Products

**Customers who viewed this
item also viewed...**



What results from a strong digital CX?

- **Higher sales.** More leads and shorter sales cycles
- **Higher profit.** Lower cost of sales and customer acquisition costs
- **Higher growth.** More to invest to scale faster—with digital



**Increase your
marketing yield**

The challenge for websites



Who wants to be like Amazon?

Every website



How many websites succeed?

Few, if any



Why is it hard for them?



They need personal
information to fuel
traditional tech

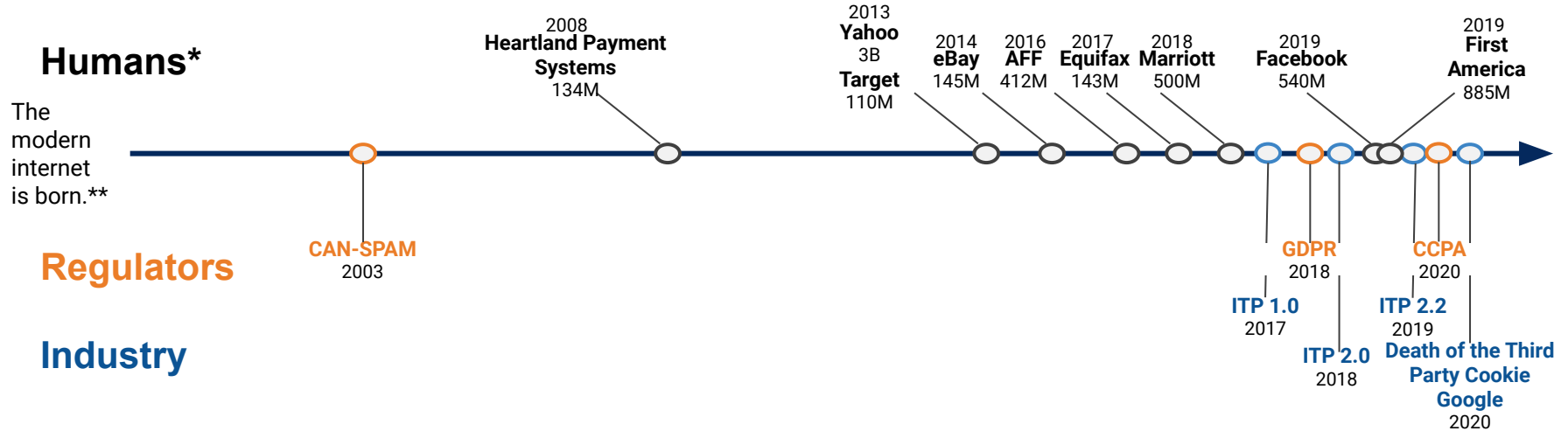
The vast majority of
their visitors are
anonymous

The world has
turned against
personal data

Why now? The world has changed



Personal data—the traditional fuel for personalization—is drying up. Regulators and industry are responding to customer frustration over the use (and misuse) of their personal information.



* Accounts compromised. Only incidents >100M Accounts

** 1994ish



147 Conversations

with large enterprise executives responsible for marketing or customer experience

Competitive software
doesn't deliver enough
value

Content demands are
costly and difficult to
meet

Long time to value—it
takes the IT team
forever to implement

Source: SoloSegment market research interviews, 1H2019

What makes CX so difficult for large sites?



Anonymous visitors

**Few websites know anything
about who visits**

Massive content

**Tens of thousands to even
millions of pages**

So what do you have to work with?



Anonymous visitors

Visitor behavior analysis



Massive content

Content topic analysis



Relevance is at the intersection



The secret of CX is to provide what customers want with the content you have



Relevance is at the intersection

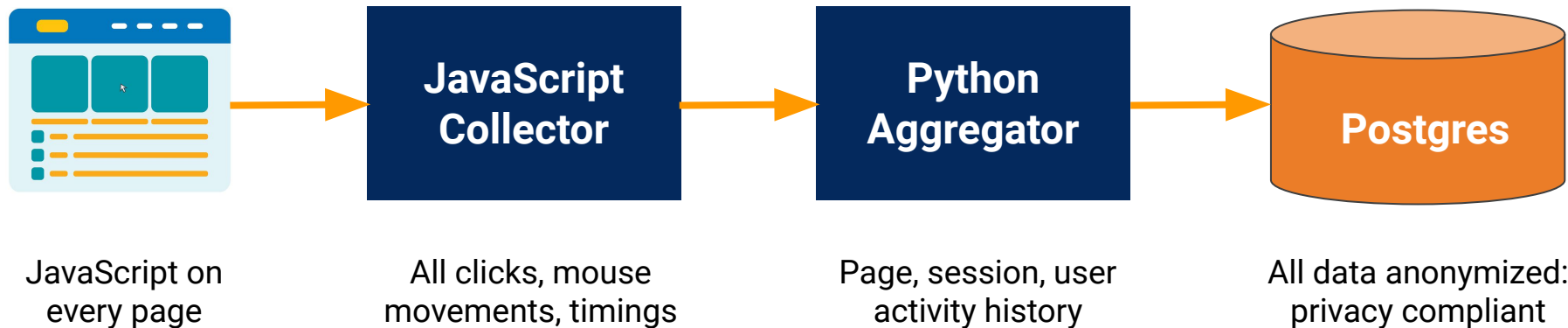


The secret of CX is to provide what customers want with the content you have



AI Using Behavioral Data

How do you collect the behavioral data?

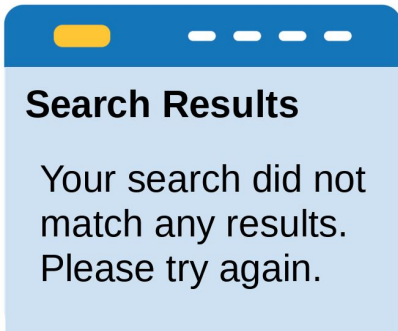


Are site searchers finding what they look for?



Sometimes it's easier to identify failures

No results



The search shows no results for the keyword

No clicks



The search shows results but the searcher doesn't click on any

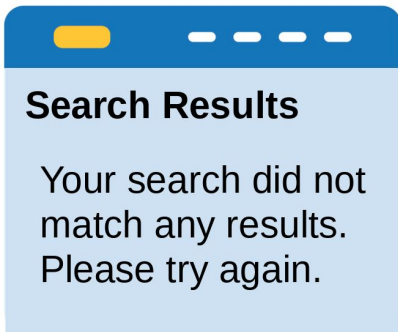
- It's easy to agree that these outcomes are failed searches
- But is every other search a success?
- Hardly...

Some searches fail, even when clicked...

...but how do you know which ones?



No results



The search shows no results for the keyword

No clicks



The search shows results but the searcher doesn't click on any

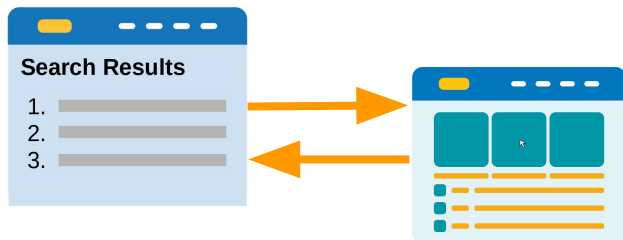
Failed Clicks



The searcher clicks but interacts with content in a manner that indicates failure.

Exhibit A: Pogosticks

Is a pogostick always a failed click?



The searcher clicks but returns to the search results page

- **YES** it's a failure

If the searcher found the right page, why would he return to the results page?

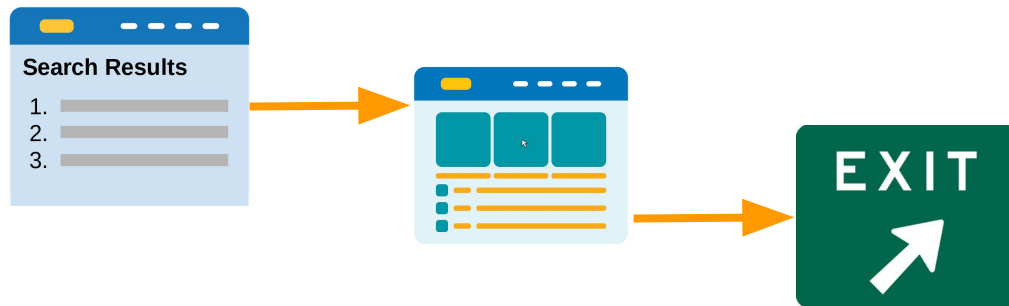
- **NO** it's not a failure

Maybe the searcher is researching a deep topic and needs to review several pages.

Exhibit B: Exits



Is an exit always a failed click?

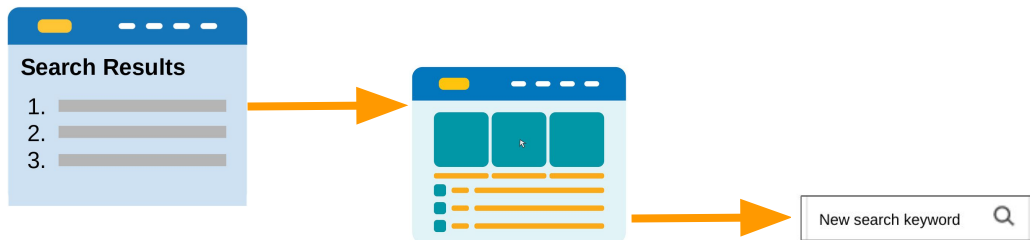


The searcher clicks but exits the site from the next page.

- **YES** it's a failure
If the searcher found the right page, why would she exit the site when she saw it?
- **NO** it's not a failure
Maybe the searcher found her answer and has completed her task successfully.

Exhibit C: New search

Is a new search always a failed click?



The searcher clicks but searches for something else from the next page.

- **YES** it's a failure

If the searcher found the right page, why would she search again?

- **NO** it's not a failure

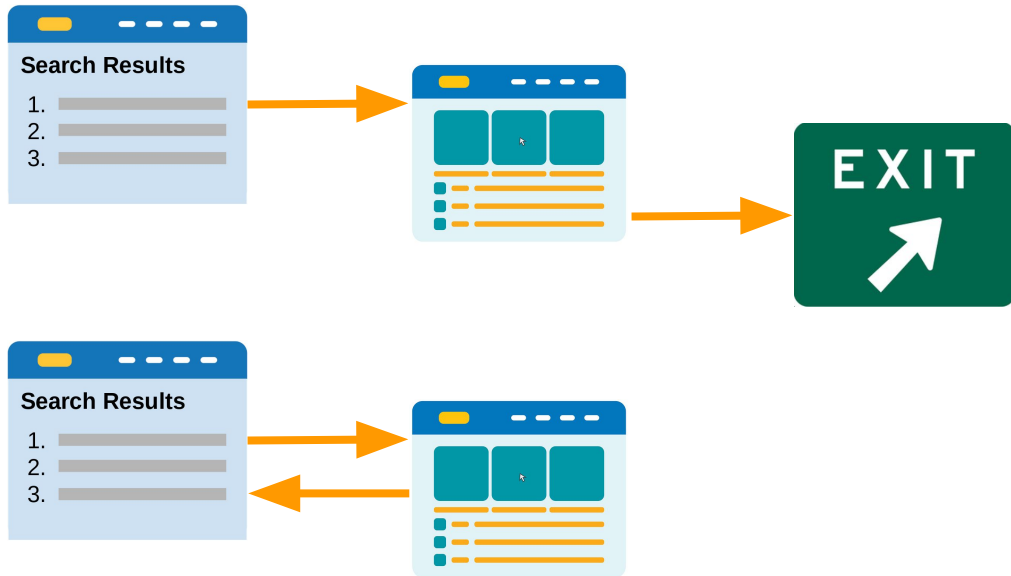
Maybe the searcher found her answer and is looking for something else now.

With enough data, AI can detect “intent shifts”

Search experiences are not black and white



The measurements can't be black and white, either



- **MAYBE** it's a failure

The searcher found the wrong page and pogo-sticked or exited.

- **MAYBE** it's not

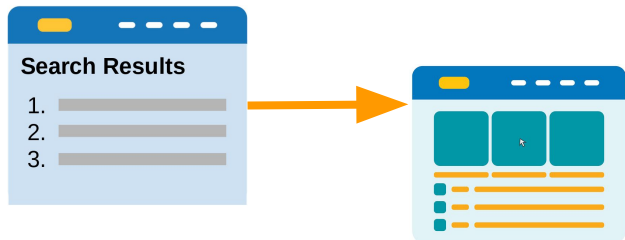
The searcher found the answer, or is looking for more than one answer.

How do we tell the difference?

It's all about reading the signals



What happens on the journey after the search?



Success Signals

- Completed a task
- Clicked a content link
- Spent time on the page
- Scrolled down the page

Failure Signals

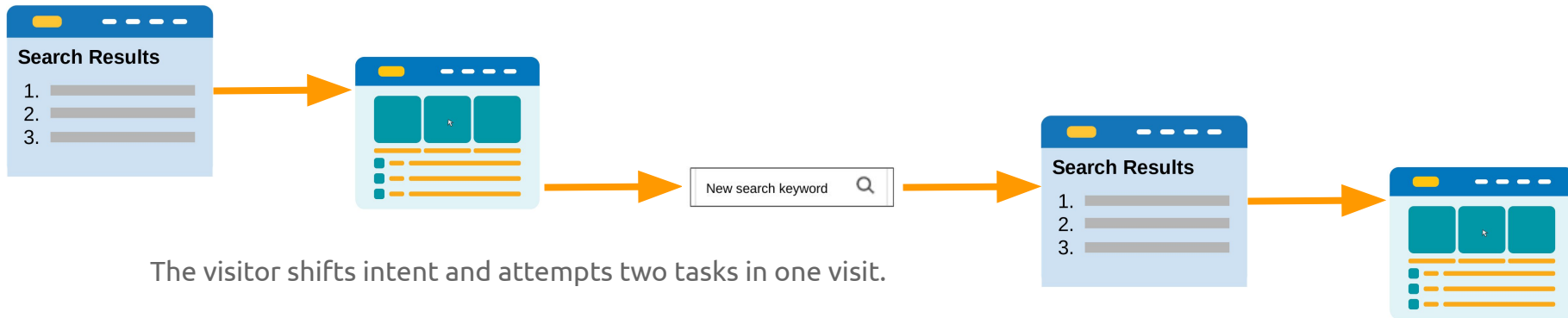
- Took no action
- Clicked on site navigation
- Short time on the page
- Never scrolled

AI can weigh all of these signals to task completion

More success involves task completion



Detecting form submissions augments knowing when answers are found

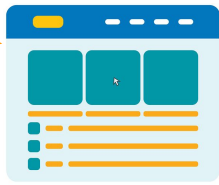


Search data reveals intent

That intent can be generalized to navigational data

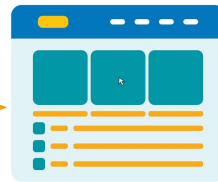
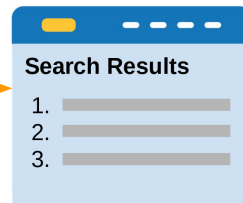
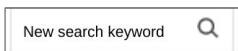


Visitor 1

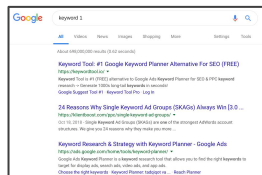


Page 1

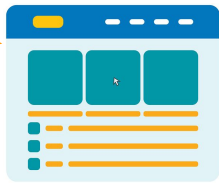
One session on your site, attempting two tasks using site search



Page 2

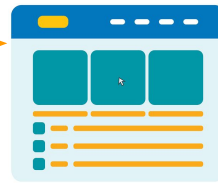
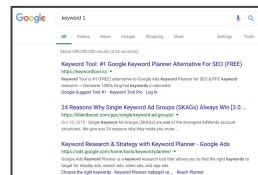


Visitor 2



Page 1

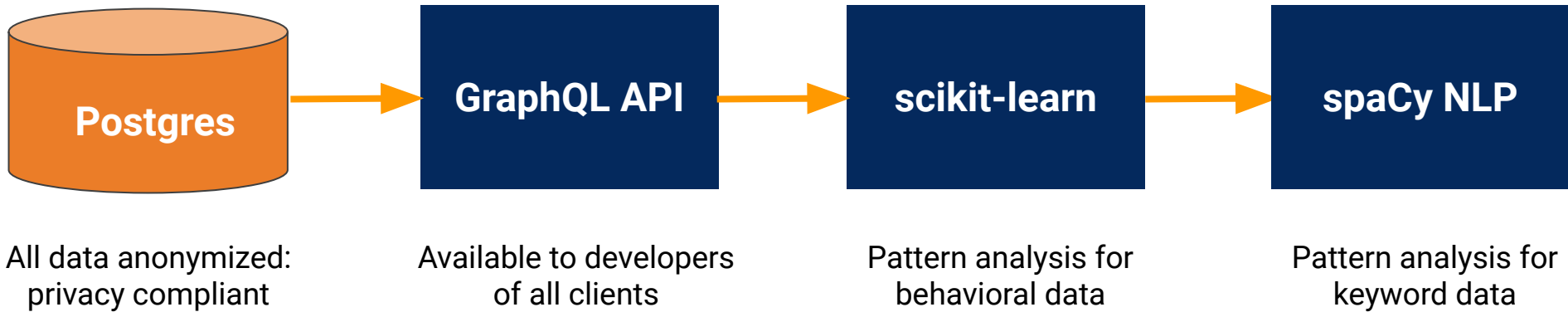
Two sessions on your site, attempting the same two tasks using Google



Page 2

How can that be done?

Predictive Analytics and AI



A GraphQL API provides behavioral data



Site Searches

- Originating URL
- Keyword
- Facets
- Results flag
- Click flag
- Success flag

Site Search Results

- Originating URL
- Keyword
- Facets
- Results URL
- Results URL rank
- Success flag

Page Navigation

- Session ID
- URL viewed
- Page events
(scroll, resize, etc.)
- Time on page
- Channel type

Using the data to suggest better keywords

Success-based search suggestions yield more success



- Instead of searching for *phone*, they can immediately select what they really want.
- Most suggestion tools suggest frequent or recent keywords—we can suggest frequent and *successful* keywords

The screenshot shows a search bar with the text 'phone' entered. Below the search bar, a list of suggestions is displayed, each starting with 'phone' followed by a keyword. The suggestions are: 'phone car charger', 'phone testing', 'phone number for ul melville', 'phone chargers', 'phone', 'phone number united states', 'phone number to reach ul', 'phone directory', and 'phone number for help'. The interface has a red header with navigation links: 'SERVICES', 'STANDARDS', and 'DASHBOARD'. A search icon is visible in the top right corner of the header.

phone

- phone **car charger**
- phone **testing**
- phone **number for ul melville**
- phone **chargers**
- phone
- phone **number united states**
- phone **number to reach ul**
- phone **directory**
- phone **number for help**

Easier than typing

Better search results

Works with any search engine

Use the data to “auto-curate” the right results

Prior successes drive the right results to the top

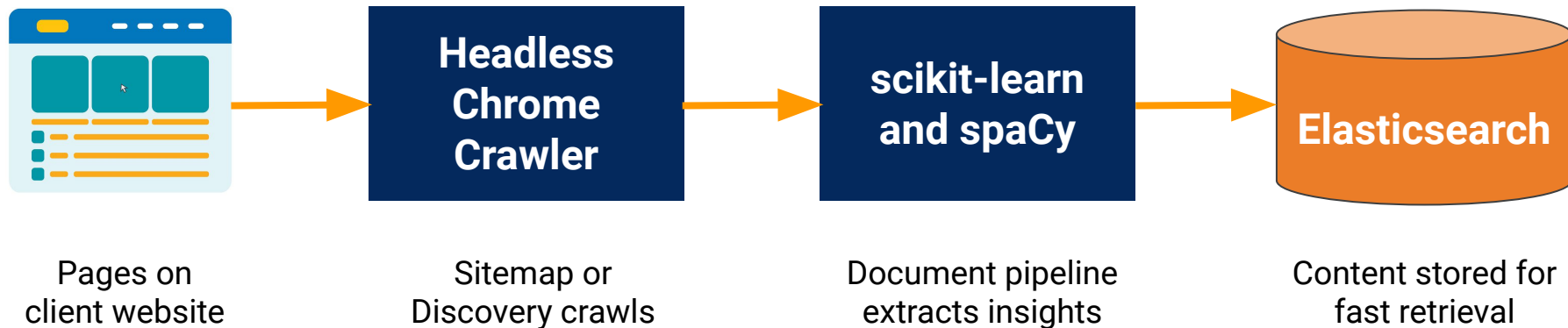


- When your search engine doesn't find the right answer, you can *manually curate* the right answer, but only for the most popular keywords
- Success data API allows your search engine to *automatically* curate *all* keywords with the successful pages

The screenshot shows a search engine interface with the search term 'silverpop' entered in the top search bar. Below the search bar, there are navigation tabs: 'All' (selected), 'Products', 'Services', 'Support', and 'For developers'. The search results show '1,165 results for silverpop'. On the right, there is a 'View' dropdown menu set to '20'. The first result is 'Silverpop Hub', which is highlighted with a large orange '#1' and an arrow pointing to it. Below this, there is a result for 'IBM Marketing Cloud and MediaMath TerminalOne - Overview - United States', which is enclosed in an orange rectangular box. The text of this result states: 'IBM and MediaMath are powering the way brands interact with consumers by helping marketers execute optimized campaigns across paid and owned channels.'

AI for Website Topics

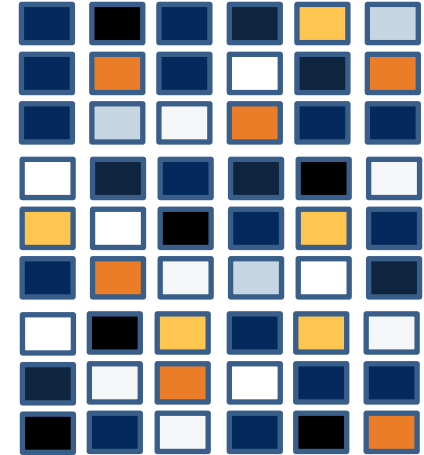
How do you collect the content data?



A multi-purpose insight is the content's topics



- **Site Search:** Facets improve site search success by allowing easy drill-down to find needed content
- **Personalization:** Content can be dynamically selected based on the visitor and the context
- **CMS and SEO:** Coverage of topics, industries, and more can be constantly assessed and adjusted in content strategy, related content
- **Standardization of Nomenclature.** It can help to have one central group naming terms of art



Topics allows searchers to drill down



Faceted search
is powered by
topics

The screenshot displays the ASME website's search interface. At the top, the ASME logo and navigation links are visible. A search bar contains the text 'pressure vessel'. Below the search bar, a horizontal menu allows filtering by category: All, Articles News Media, CS Committees, Community, Courses, Events, Publications, and Standards. The 'All' category is selected. On the left, a 'Refine Search' sidebar is highlighted with an orange border. It contains two sections: 'Topic' and 'Industry', each with an 'expand all' link. Under 'Topic', there are three checkboxes: 'ASME Information', 'Products, Services, and Publications', and 'Engineering'. Under 'Industry', there is one checkbox: 'Energy and Utilities'. The main content area shows the search results for 'pressure vessel', indicating 667042 results. Two results are visible: 'ASME BPV Code, Section VIII, Division 2: Design & Fabrication of Pressure Vessels - ASME' and 'Boiler and Pressure Vessel Certification | ASME - ASME'.

Refine Search

Topic expand all

- ☐ ASME Information
- ☐ Products, Services, and Publications
- ☐ Engineering

Industry expand all

- ☐ Energy and Utilities

Your search for 'pressure vessel' returned **667042** results

[ASME BPV Code, Section VIII, Division 2: Design & Fabrication of Pressure Vessels - ASME](https://www.asme.org/learning-development/find-course/bpv-code-section-viii-division-2-alternative-rules-construction-pressure-vessels)

<https://www.asme.org/learning-development/find-course/bpv-code-section-viii-division-2-alternative-rules-construction-pressure-vessels>

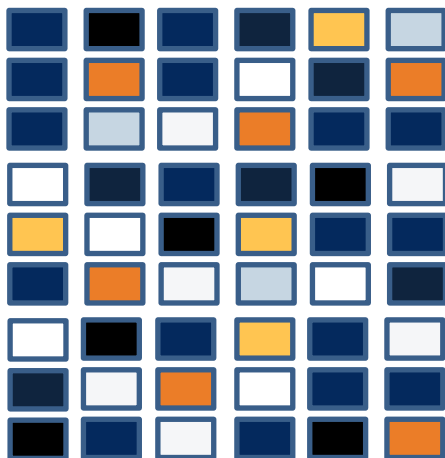
This course describes the use of alternative rules for the design and fabrication of pressure vessels given in ASME BPV Code, Section VIII, Division 2.

[Boiler and Pressure Vessel Certification | ASME - ASME](https://www.asme.org/certification-accreditation/boiler-and-pressure-vessel-certification)

<https://www.asme.org/certification-accreditation/boiler-and-pressure-vessel-certification>

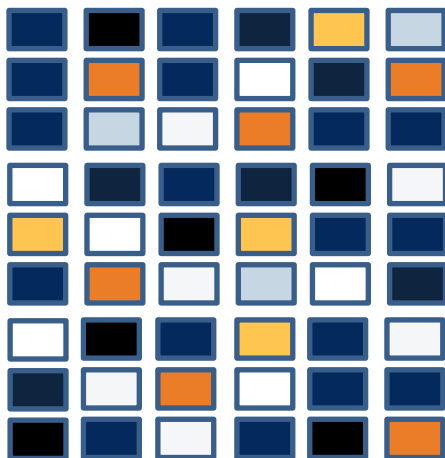
For a certification in accordance with ASME Boiler & Pressure Vessel Code (BPVC) Sections I, IV, VIII, X and/or XII of manufacturer's or assembler's quality control system, [click here!](#)

What's wrong with existing topic modeling?



- **Clustering isn't good enough.** Machines can automatically group topics, but the resulting topics often don't make sense to humans. Some are OK, but others cause head-scratching.
- **Manual taxonomies take time.** If you lay five SMEs end to end, they all point in different directions. It can take months to get agreement on the “right” topics.
- **Taxonomies must reflect the documents.** Often, manually-created taxonomies aspirational topics that are important, but they hinder retrieval because documents about those topics are rare.
- **Manual taxonomies are hard to classify.** We'd like to automatically label all documents, but taxonomies with lots of topic overlap are hard to label accurately.

What can you do instead? Human-in-the-loop

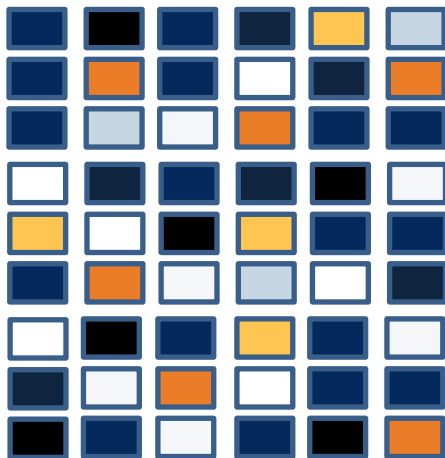


- **First, the machine identifies topics.** It starts with messy, raw data that the machine has grouped together. The topics look awkward because they haven't yet had human grouping or feedback.
- **SMEs analyze the topics.** SMEs look at the language and help answer questions about what the topics should look like and how they might be grouped together using their expertise.
- **SMEs provide feedback.** SMEs provide information about which topics are appropriate and desired, and which ones to remove, as well as which ones might be merged with others.

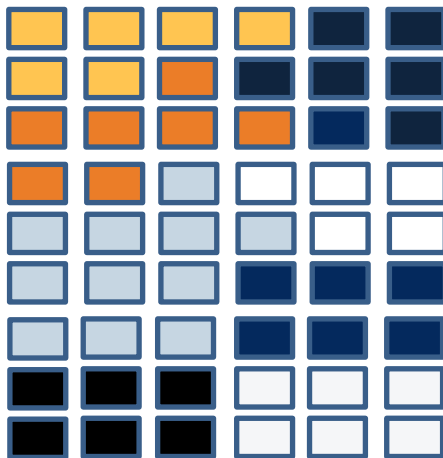
The SME identifies the right topics



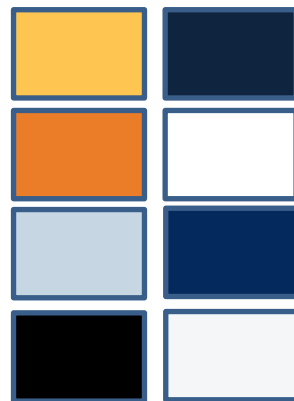
**Identify topics
from documents**



**Re-run the model
to refine topics**



**Finalize
the topics**

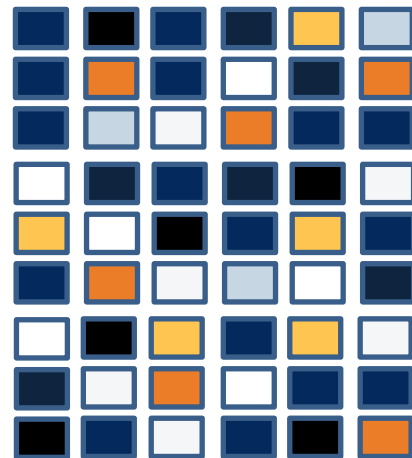


Why shouldn't we tag documents by hand?



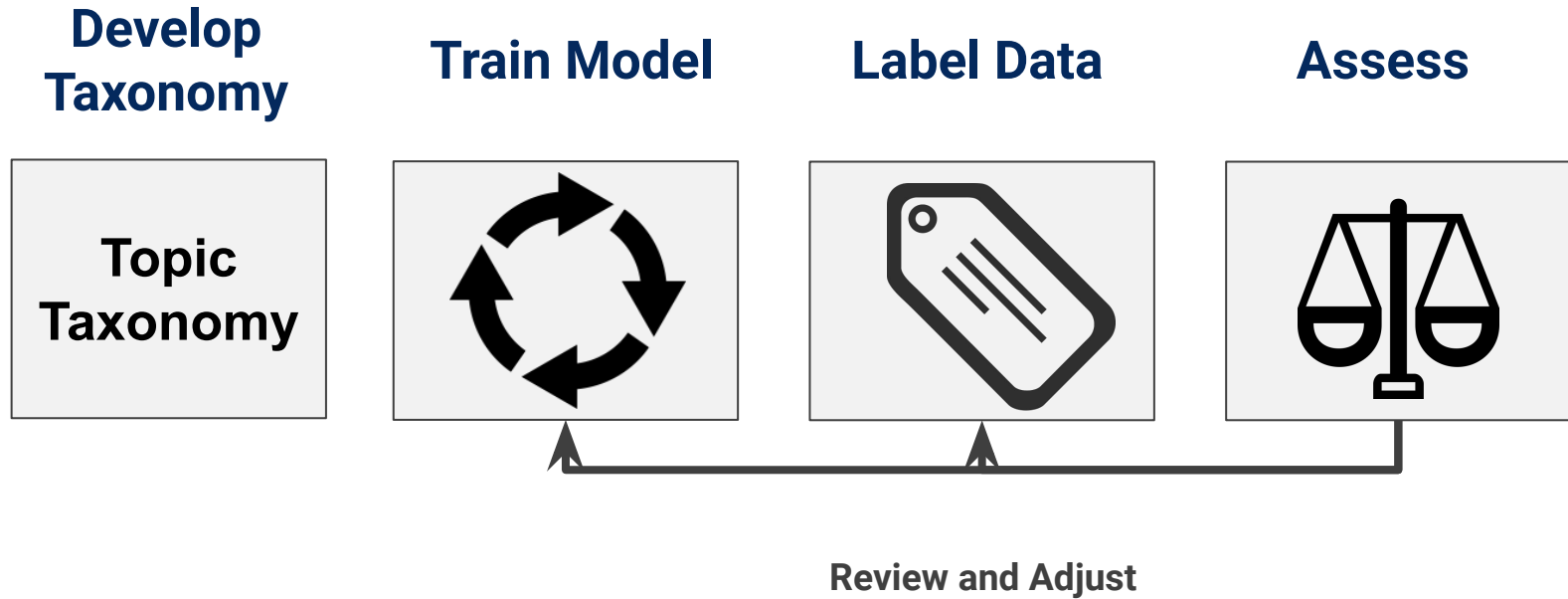
Automatic classification increases consistency

- **Too many categories.** Manual classification is error prone, because consistently deciding the right classification among dozens of choices is very hard for people
- **People are inconsistent.** Even the same people disagree with themselves—35% of document coders disagree with themselves when given the same task a few days later



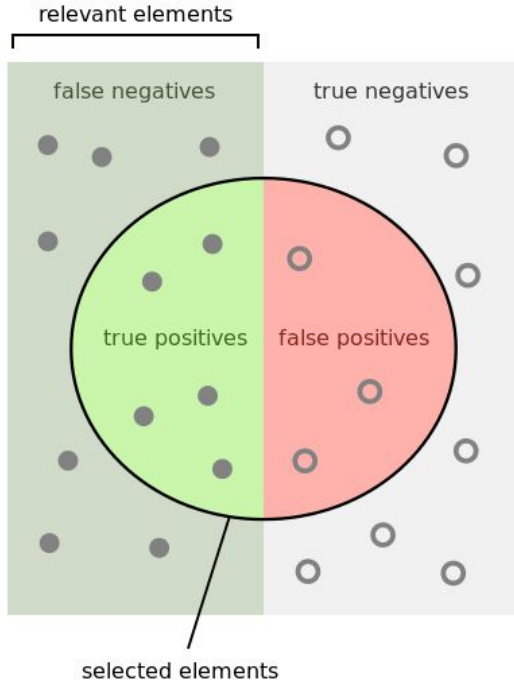


How does auto-tagging work?





Measure accuracy with precision and recall



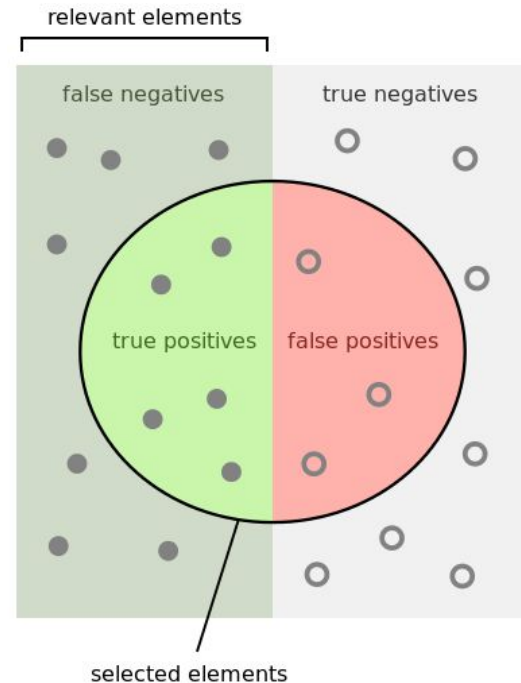
- **Precision:** How many of the selected elements are true positives?
- **Recall:** How many of the relevant elements are selected?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$
$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$



What's good? F-Measure tells you.

- F-Measure trades off precision and recall
- Higher F-Measure are better
- You can weight precision and recall differently
- When in doubt, weight them the same





Topics also can be used programmatically

- APIs can add topics to your CMS
- APIs can embed recommendations into your page template

Related Products

**Specifications for Band Saw
Blades (Metal Cutting)**

Standards

Hobs

Standards

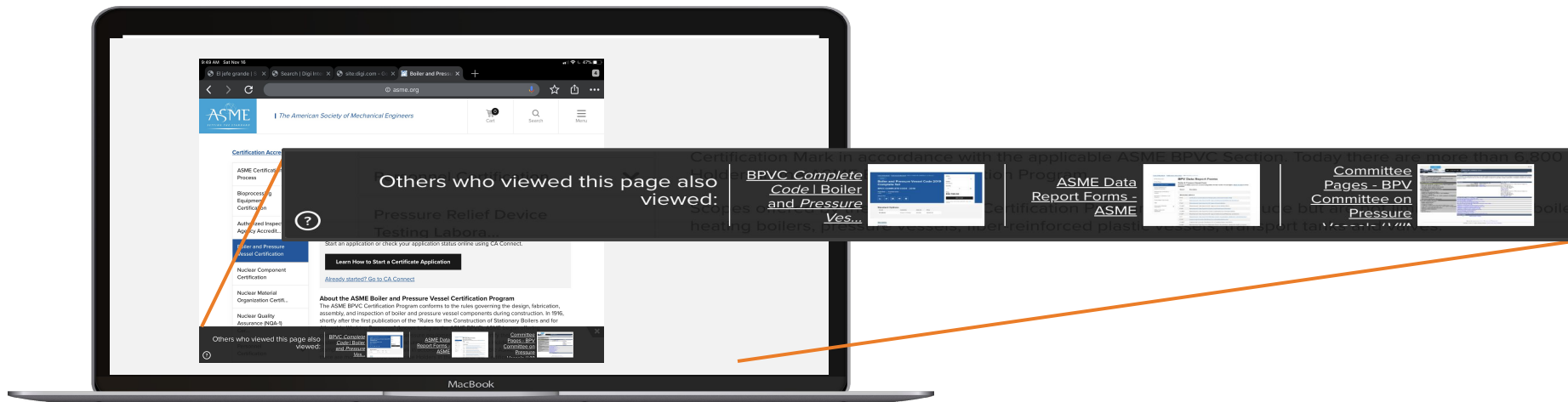
**Tapered and Reduced Cross
Section Retaining Rings (Inch
Series)**

Standards



Topics fuel content recommendation

Present the next best action so each user succeeds



Presented as a modal, but can be embedded on the page by your CMS. using an API



Questions?

Mike Moran (Mike@SoloSegment.com)
