Aggregation and Search: Baskets for Berrypicking

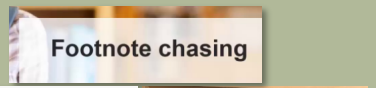
NKOS 2020 Consolidated Workshops

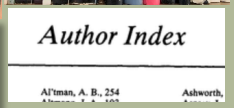
Copyright 2020 Taxonomy Strategies LLC. All rights reserved.

Agenda

- Assumptions and context of Marcia Bates' "berrypicking" vision
- Methods for exploring large amounts of information

How





Where

Who

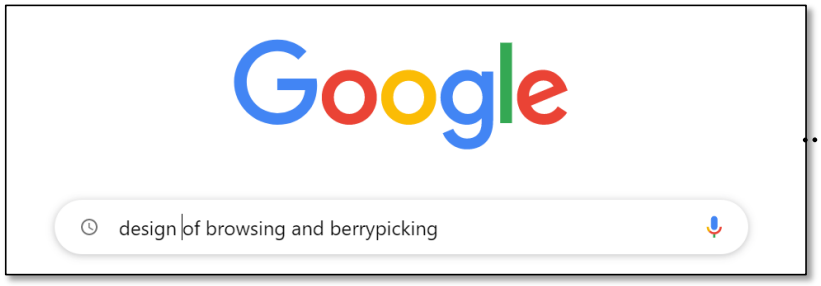
Users switch from one searching mode to another in a physical library.

How

Where

Who

Users can switch among or combine searching modes in a digital library.



About 521,000 results (0.40 seconds)

www.emerald.com › ... › Volume 13 Issue 5

The design of browsing and berrypicking techniques for the ...

May 1, 1989 - First, a new model of searching in online and other information systems, called '**berrypicking**', is discussed. This model, it is argued, is much ...

by MJ Bates - 1989 - Cited by 2370 - Related articles

You've visited this page 2 times. Last visit: 8/24/20

What Marcia Bates did not foresee in 1989 …

Please note you might not have access to this content

You may be able to access this content by login via Shibboleth, Open Athens or with your Emerald account. **Login**

To rent this content from Deepdyve, please click the button. **Rent from Deepdyve**

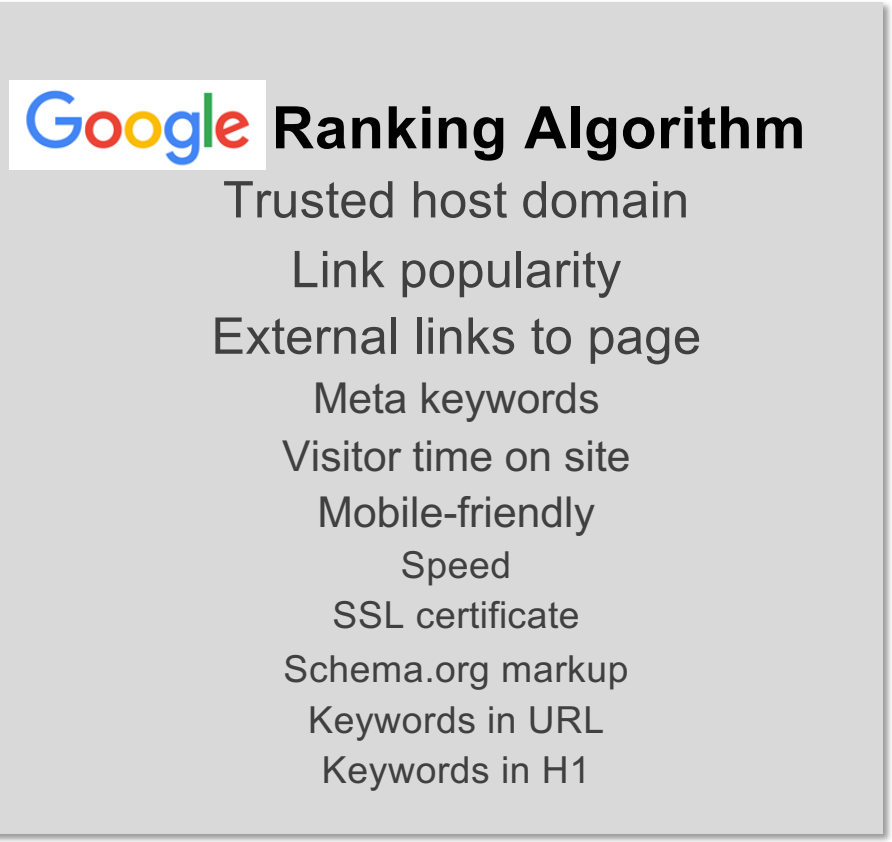
If you think you should have access to this content, click the button to contact our support team. **Contact us**

Methods for exploring large amounts of information

WWW/Enterprise Search Interfaces	Berrypicking Search Interfaces
Natural language processing and analytics	Citation searching
Find an expert	Footnote chasing
Guided navigation	Subject searches
Search results as collections	Area scanning
Visualize collections	
Knowledge graphs	

Natural language processing and especially analytics are Citation Searching on steroids

- NLP is deployed on a massive scale.
  - Identify and index meaningful entities beyond simple term frequency and document length.
- Websites and content are instrumented with usage analytics.
  - Usage analytics rank and promote “popular” information items, similar to citation searching
    - Hyperlinks
    - Visit frequency
    - Other factors

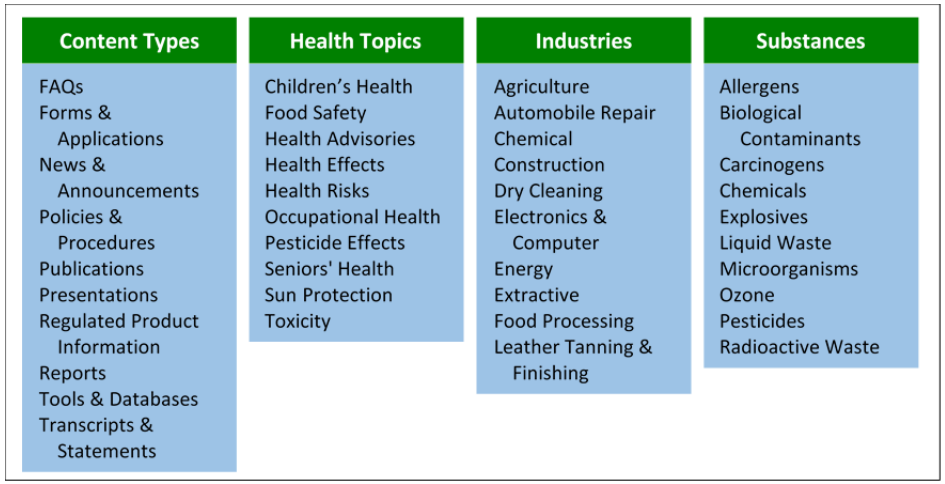


How do you find an expert? … by footnote chasing

- Assumptions
  - Full-text search
  - Comprehensive collection
- Plan B: Ask an expert
    - Email a colleague
    - Expertise directories
    - LinkedIn, Research Gate, etc.
    - Facebook
    - Chatbots
  - It's like footnote chasing

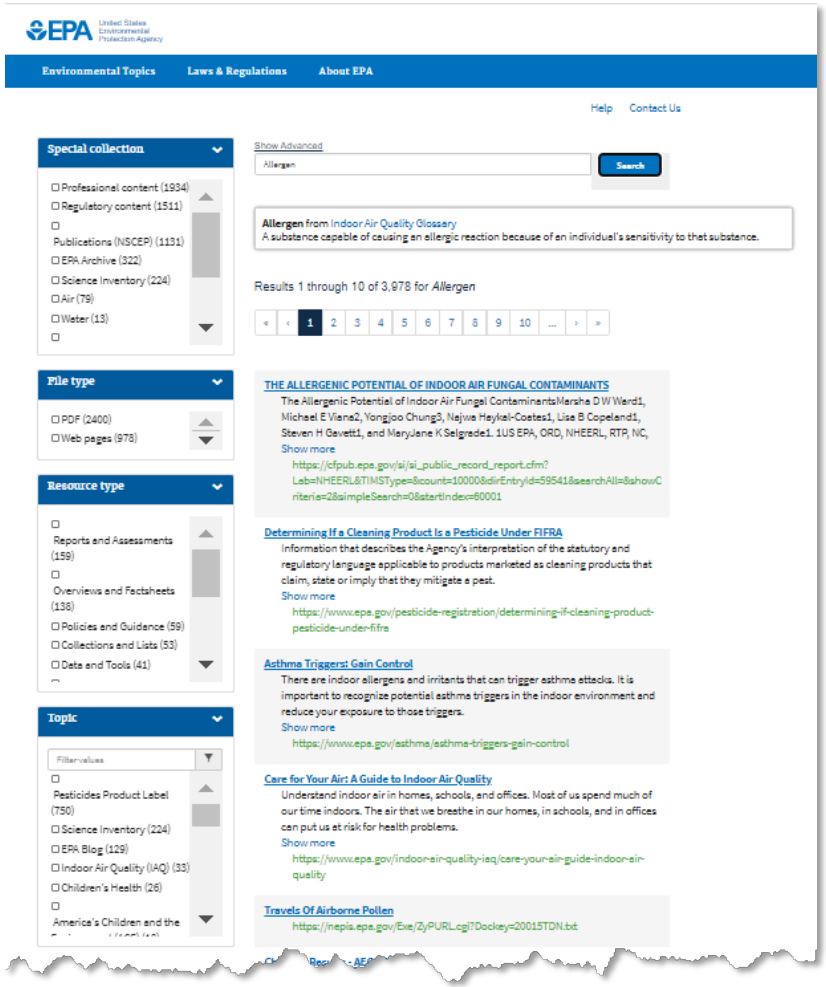
Guided navigation is the new Subject Searching paradigm

- Take advantage of ubiquitous search as an entry point for browsing
- Break the paradigm that the relevant result must be near the top of the results
- Guided navigation is a model for refining a very large text search collection in a few clicks



Busch's Golden Rule: Four metadata-controlled vocabularies of 10 values each have the same discriminatory power as one taxonomy of 10,000 values.

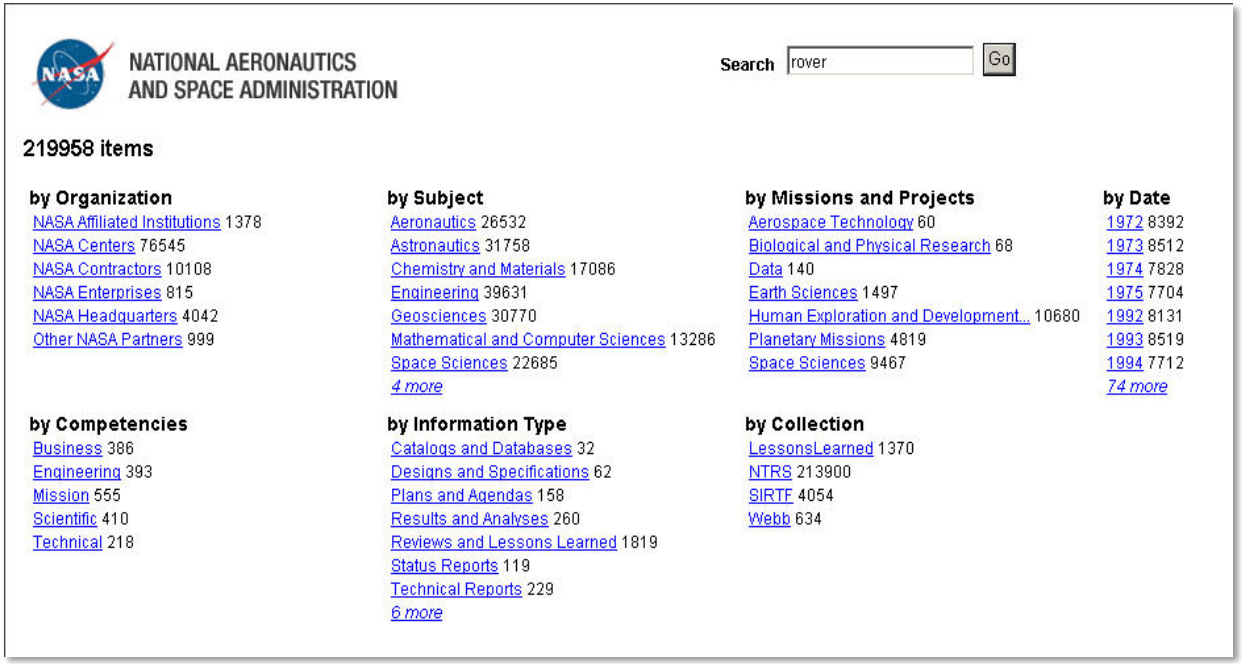
Guided navigation on a content website



Guided navigation applied on a content site epa.gov.

Search results as collections: A type of area scanning

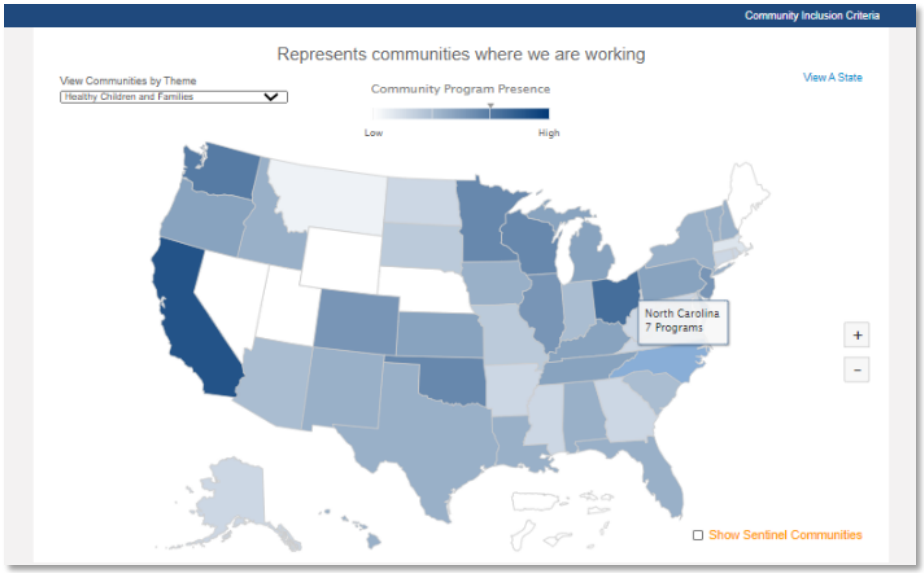
- Every search should be thought of as a collection of results, instead of presenting text search results as a list of references.
- Provide the user with an overview of the available information, and invite them to refine or start with a new search.

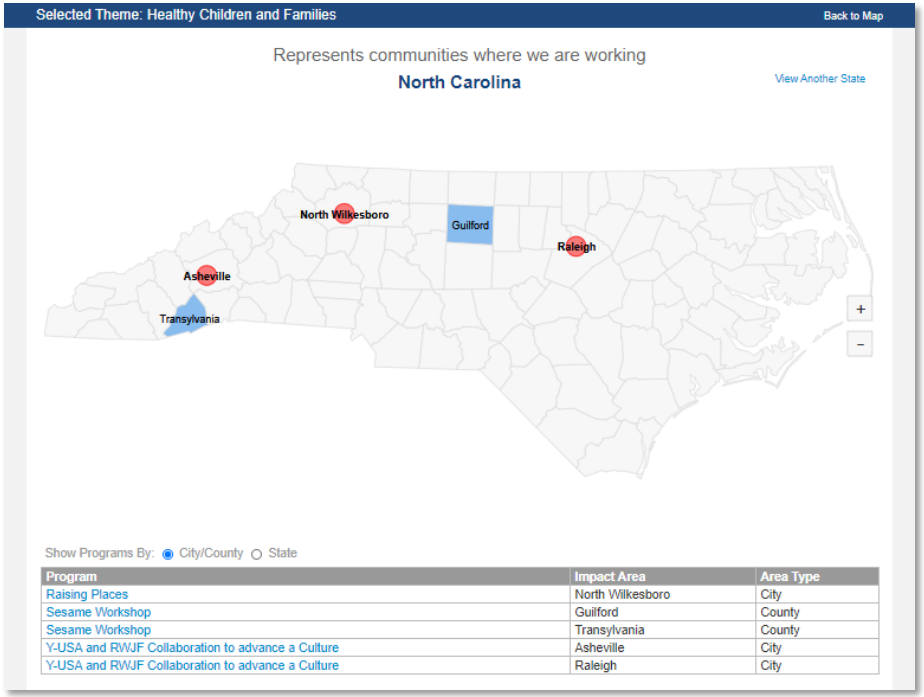


Collection of more than 200,000 search results for [Mars] rover in the top occurring categories of the NASA Taxonomy, a faceted KOS.

Visualizing collections

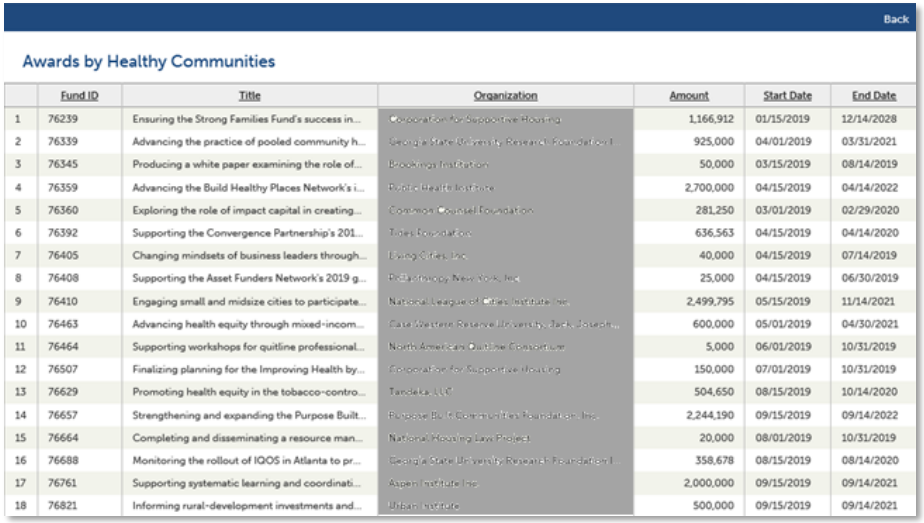
- Visualize collections of search results with maps and charts instead of lists of references.





A map visualization of search results that displays themes (topics) for U.S. states, and a drill-down to a state with county/city items.

Visualizing collections with charts and drill-downs

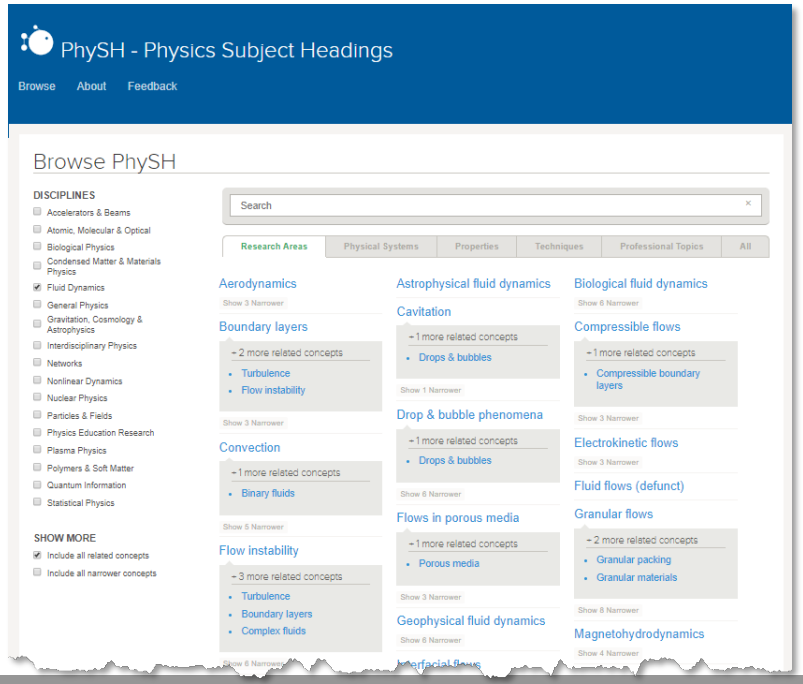


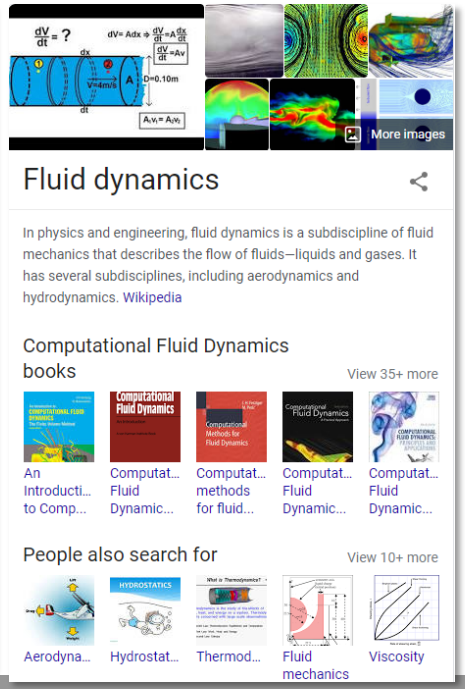
	Fund ID	Title	Organization	Amount	Start Date	End Date
1	76239	Ensuring the Strong Families Fund's success in...	[redacted]	1,166,912	01/15/2019	12/14/2028
2	76339	Advancing the practice of pooled community h...	[redacted]	925,000	04/01/2019	03/31/2021
3	76345	Producing a white paper examining the role of...	[redacted]	50,000	03/15/2019	08/14/2019
4	76359	Advancing the Build Healthy Places Network's i...	[redacted]	2,700,000	04/15/2019	04/14/2022
5	76360	Exploring the role of impact capital in creating...	[redacted]	281,250	03/01/2019	02/29/2020
6	76392	Supporting the Convergence Partnership's 201...	[redacted]	636,563	04/15/2019	04/14/2020
7	76405	Changing mindsets of business leaders through...	[redacted]	40,000	04/15/2019	07/14/2019
8	76408	Supporting the Asset Funders Network's 2019 g...	[redacted]	25,000	04/15/2019	06/30/2019
9	76410	Engaging small and midsize cities to participate...	[redacted]	2,499,795	05/15/2019	11/14/2021
10	76463	Advancing health equity through mixed-incom...	[redacted]	600,000	05/01/2019	04/30/2021
11	76464	Supporting workshops for quitline professional...	[redacted]	5,000	06/01/2019	10/31/2019
12	76507	Finalizing planning for the Improving Health by...	[redacted]	150,000	07/01/2019	10/31/2019
13	76629	Promoting health equity in the tobacco-contro...	[redacted]	504,650	08/15/2019	10/14/2020
14	76657	Strengthening and expanding the Purpose Built...	[redacted]	2,244,190	09/15/2019	09/14/2022
15	76664	Completing and disseminating a resource man...	[redacted]	20,000	08/01/2019	10/31/2019
16	76688	Monitoring the rollout of IQOS in Atlanta to pr...	[redacted]	358,678	08/15/2019	08/14/2020
17	76761	Supporting systematic learning and coordinati...	[redacted]	2,000,000	09/15/2019	09/14/2021
18	76821	Informing rural-development investments and...	[redacted]	500,000	09/15/2019	09/14/2021

A chart visualization that shows total and KPI amounts awarded by lines of business and in summary for the whole enterprise.

Knowledge graphs

- Representations of an organization's knowledge assets, content, and data—people, places, documents, multimedia, data, etc.—and how these things are related to each other.
- Typically, this is an ontology that defines classes for the things, properties for the things, and relationships between the things.





An ontology for the physics domain with the knowledge graph for the same concept designed to be presented on the search results page.

KOS are the baskets for gathering “berries”

- The purpose of KOS is not to find items or answers, but to group or aggregate content into collections for review or further refinement.
- Consider the search results user experience when designing KOS .

Resources

- M. Bates. “The design of browsing and berrypicking techniques for the online search interface.” 13(5) *Online Review* 407-424, and in: M. Bates. *Information Searching Theory and Practice: Selected Works*. Vol. 2. Berkeley: Ketchikan Press, 2016. pp. 257-278.
- V. Bush. “As we may think.” *The Atlantic* (July 1945). https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/. Last checked: 6/8/2020.
- S. Papa. “The faceted navigation and search revolution.” *KM World* (March 23, 2006) https://www.kmworld.com/Articles/White-Paper/Article/The-Faceted-Navigation-and-Search-Revolution-15378.aspx. Last checked: 6/8/2020.
- *NASA Taxonomy*. Last updated: 05/08/2012. https://vocabularyserver.com/nasa/. Last checked: 6/9/2020.
- *PhySH – Physics Subject Headings*. American Physical Society. https://physh.aps.org/. Last checked: 6/11/2020.

Summary

The goal of search is to reliably find what you are looking for, to be able to type in a highly variable query and return the most relevant result or the right answer every time. These days, effective search relies to a large extent on natural language processing and analytics. The purpose of KOS is not to find items or answers, but to group or aggregate content into collections for review or further refinement. This can be pre-search to build a collection to search on rather than the whole universe, or it can be post-search to characterize the search result set, or refine the results. It's important to consider the kind of search result user experience when the KOS is designed. The aggregation scenario means a broad and shallow scheme with discrete categories is needed. The focus needs to be on designing the baskets for gathering "berries" rather than the berries themselves that users will be picking. This paper lays out some use cases for this aggregation scenario and presents some examples.