

# Aggregation and Search: Baskets for Berrypicking

NKOS 2020 Consolidated Workshops

## Agenda

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

## How

Footnote chasing

Citation Indexing

Journal Run

Area Scanning

Author Index

SUBJECT INDEX



9, 127-32  
America, United State  
A/iman, A. B., 254  
Aibworth,



## Who



## Where



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

## How



Enter Queries below		Truncate	Listings
Title	<input type="text"/>	<input checked="" type="checkbox"/>	
Author	<input type="text"/>	<input checked="" type="checkbox"/>	List of authors
Organisation	<input type="text"/>	<input checked="" type="checkbox"/>	List of organisations
Text search	<input type="text"/>	<input checked="" type="checkbox"/>	
Subject	<input type="text"/>	<input checked="" type="checkbox"/>	List of subjects
Year	<input type="text"/>	<input checked="" type="checkbox"/>	
Show	50 hits on 1 page		
Search in	Books		
Submit form		Clear form	Help
			Expert search form



## Who



## Where



**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.

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

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

Login

Rent from Deepdyve

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

### Ranking Algorithm

Trusted host domain  
Link popularity  
External links to page  
Meta keywords  
Visitor time on site  
Mobile-friendly  
Speed  
SSL certificate  
Schema.org markup  
Keywords in URL  
Keywords in H1

## 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

Content Types	Health Topics	Industries	Substances
FAQs	Children's Health	Agriculture	Allergens
Forms & Applications	Food Safety	Automobile Repair	Biological Contaminants
News & Announcements	Health Advisories	Chemical	Carcinogens
Policies & Procedures	Health Effects	Construction	Chemicals
Publications	Health Risks	Dry Cleaning	Explosives
Presentations	Occupational Health	Electronics & Computer	Liquid Waste
Regulated Product Information	Pesticide Effects	Energy	Microorganisms
Reports	Seniors' Health	Extractive	Ozone
Tools & Databases	Sun Protection	Food Processing	Pesticides
Transcripts & Statements	Toxicity	Leather Tanning & Finishing	Radioactive Waste

***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

The screenshot shows the EPA website's search interface. At the top, the EPA logo and navigation menu are visible. The search bar contains the word "Allergen" and a "Search" button. Below the search bar, there are several filter panels: "Special collection" with options like Professional content (1934) and Regulatory content (1511); "File type" with PDF (2400) and Web pages (978); "Resource type" with Reports and Assessments (159) and Overviews and Factsheets (138); and "Topic" with a dropdown menu. The search results show "Results 1 through 10 of 3,978 for Allergen". The first result is titled "THE ALLERGENIC POTENTIAL OF INDOOR AIR FUNGAL CONTAMINANTS" and includes a list of authors and a "Show more" link. The second result is "Determining if a Cleaning Product is a Pesticide Under FIFRA" and the third is "Asthma Triggers: Gain Control".

**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.

**NASA NATIONAL AERONAUTICS AND SPACE ADMINISTRATION**

Search

**219958 items**

by Organization	by Subject	by Missions and Projects	by Date
<a href="#">NASA Affiliated Institutions</a> 1378	<a href="#">Aeronautics</a> 26532	<a href="#">Aerospace Technology</a> 60	<a href="#">1972</a> 8392
<a href="#">NASA Centers</a> 76545	<a href="#">Astronautics</a> 31758	<a href="#">Biological and Physical Research</a> 68	<a href="#">1973</a> 8512
<a href="#">NASA Contractors</a> 10108	<a href="#">Chemistry and Materials</a> 17086	<a href="#">Data</a> 140	<a href="#">1974</a> 7828
<a href="#">NASA Enterprises</a> 815	<a href="#">Engineering</a> 39631	<a href="#">Earth Sciences</a> 1497	<a href="#">1975</a> 7704
<a href="#">NASA Headquarters</a> 4042	<a href="#">Geosciences</a> 30770	<a href="#">Human Exploration and Development...</a> 10680	<a href="#">1992</a> 8131
<a href="#">Other NASA Partners</a> 999	<a href="#">Mathematical and Computer Sciences</a> 13286	<a href="#">Planetary Missions</a> 4819	<a href="#">1993</a> 8519
	<a href="#">Space Sciences</a> 22685	<a href="#">Space Sciences</a> 9467	<a href="#">1994</a> 7712
	<a href="#">4 more</a>		<a href="#">74 more</a>

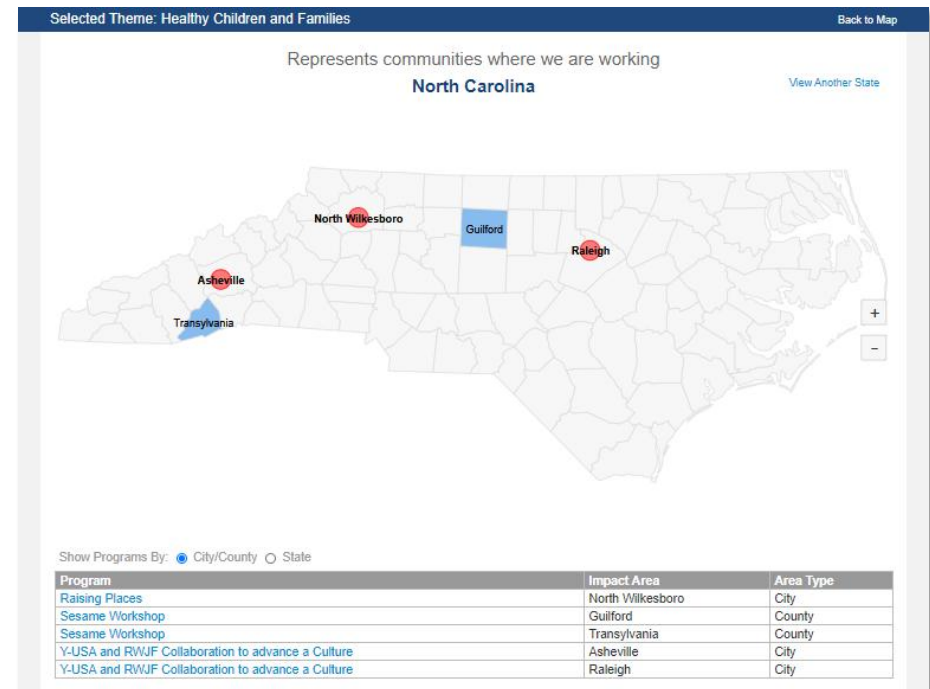
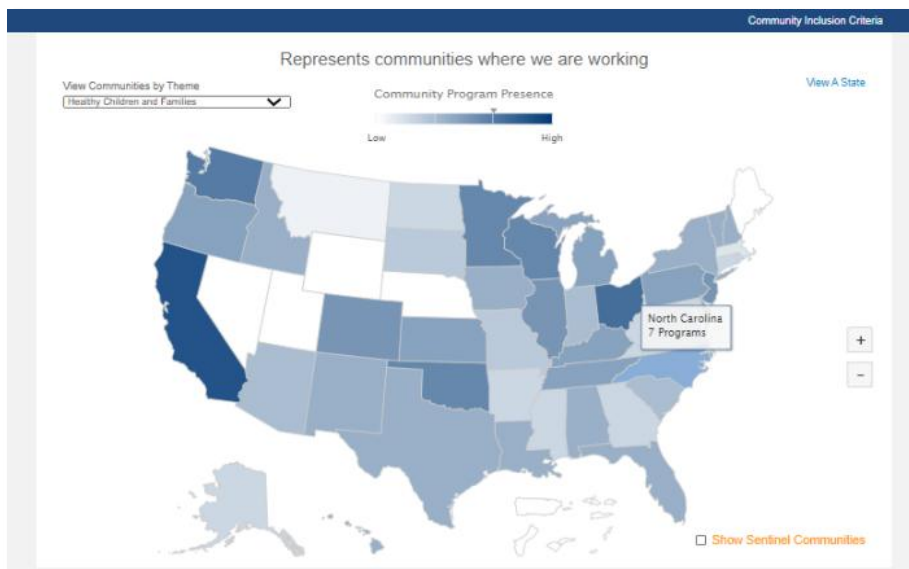
  

by Competencies	by Information Type	by Collection
<a href="#">Business</a> 386	<a href="#">Catalogs and Databases</a> 32	<a href="#">Lessons Learned</a> 1370
<a href="#">Engineering</a> 393	<a href="#">Designs and Specifications</a> 62	<a href="#">NTRS</a> 213900
<a href="#">Mission</a> 555	<a href="#">Plans and Agendas</a> 158	<a href="#">SIRTF</a> 4054
<a href="#">Scientific</a> 410	<a href="#">Results and Analyses</a> 260	<a href="#">Webb</a> 634
<a href="#">Technical</a> 218	<a href="#">Reviews and Lessons Learned</a> 1819	
	<a href="#">Status Reports</a> 119	
	<a href="#">Technical Reports</a> 229	
	<a href="#">6 more</a>	

**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



Back

### Awards by Healthy Communities

	Fund ID	Title	Organization	Amount	Start Date	End Date
1	76239	Ensuring the Strong Families Fund's success in...	Coalition for Reproductive Health	1,166,912	01/15/2019	12/14/2028
2	76339	Advancing the practice of pooled community h...	Georgia State University Research Foundation	925,000	04/01/2019	03/31/2021
3	76345	Producing a white paper examining the role of...	Brookings Institution	50,000	03/15/2019	08/14/2019
4	76359	Advancing the Build Healthy Places Network's L...	Build Healthy Places	2,700,000	04/15/2019	04/14/2022
5	76360	Exploring the role of impact capital in creati...	Commonwealth Foundation	281,250	03/01/2019	02/29/2020
6	76392	Supporting the Convergence Partnership's 201...	Trust Foundation	636,563	04/15/2019	04/14/2020
7	76405	Changing mindsets of business leaders throug...	Ernst & Young	40,000	04/15/2019	07/14/2019
8	76408	Supporting the Asset Funders Network's 2019 g...	Rockefeller Foundation	25,000	04/15/2019	06/30/2019
9	76410	Engaging small and midsize cities to partici...	National League of Cities Institute	2,499,795	05/15/2019	11/14/2021
10	76463	Advancing health equity through mixed-incom...	Case Western Reserve University, Buck Commu...	600,000	05/01/2019	04/30/2021
11	76464	Supporting workshops for quiltline profession...	North America Quiltline Foundation	5,000	06/01/2019	10/31/2019
12	76507	Finalizing planning for the Improving Health by...	Coalition for Reproductive Health	150,000	07/01/2019	10/31/2019
13	76629	Promoting health equity in the tobacco contro...	Tandem LLC	504,650	08/15/2019	10/14/2020
14	76657	Strengthening and expanding the Purpose Built...	Purpose Built Communities Foundation, Inc.	2,244,190	09/15/2019	09/14/2022
15	76664	Completing and disseminating a resource man...	National Housing Law Project	20,000	08/01/2019	10/31/2019
16	76688	Monitoring the rollout of IQOS in Atlanta to pr...	Georgia State University Research Foundation	358,678	08/15/2019	08/14/2020
17	76761	Supporting systematic learning and coordinati...	Aspen Institute	2,000,000	09/15/2019	09/14/2021
18	76821	Informing rural-development investments and...	Urban Institute	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.