Information Seeking Behavior

INFO 518 · October 15, 2018 · Chris Holstrom

Information Seeking Activity

- What was Seattle's nickname before Emerald City?

- What did the old nickname mean?

Talk about:
- What process did you follow and what tools did you use?
- How did you feel during the process?
- How was this an artificial process? What do you "usually" do?
- What do you want to do next?

What is information seeking behavior?

Information Seeking Behavior

The way people search for and use information.

The process or activity of attempting to obtain information.

These are just Wikipedia definitions.

Information Seeking Behavior

"The user's constructive activity of finding meaning from information in order to extend his or her state of knowledge on a particular problem or topic."

Kuhlthau

How did we get to Information Seeking Behavior?

Cataloging and Classification: **1880's - 1900's**

Information Retrieval: **1940's - present**

Information Seeking Behavior: **1980's - 1990's**

User Experience: **1990's- present**

These are very rough dates.

End dates also don't mean a complete end to work in the area. For example, cataloging and classification work is ongoing, but the dates indicated are when lots of foundational work happened.

Information seeking and information seeking behavior are still going, but they have been somewhat subsumed by human computer interaction / user experience / usability.

Cataloging and Classification

3rd Century BCE: Library of Alexandria

1880: Cutter Expansive Classification

1885: Dewey Decimal Classification

1904: Otlet and Universal Decimal Classification

- Focus on retrieval, mostly by librarians.
- System focus rather than human focus.
- Controlled vocabularies.
- Best for known-item search.

Libraries got bigger, and eventually we had to organize them so we could find stuff.

We don't have the power of computers yet, but we have the power of librarians.

They put everything in its "right place" and then help people find information.

Early in this era, the act information seeking was about asking the librarian. Later, as access to the tools grew, it became "learning the system" and figuring out how to retrieve information under the organization system. Researchers weren't considering the information seeking journey that researchers went on or paying particular attention to how people searched or what terms they used.

Note the call numbers on the book spines. These are NOT about information seeking. They are about cataloging -- what have we got and how do we uniquely identify it in the system. Very few humans can actually read these things.
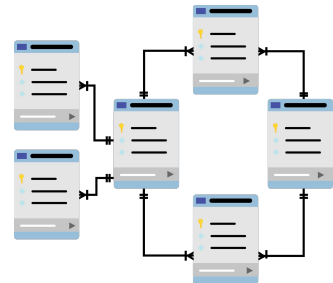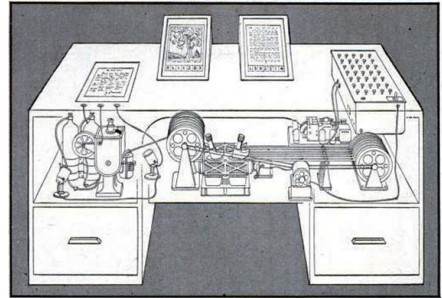
Note that this history is VERY Western-centric.

Information Retrieval

1940's-1950's: Computers arrive to save us all!

Or, rather, to make "the system" faster, smarter, and more complicated. **Focus is still not on the person.**

Over the decades IR systems have handled larger corpora and employed more sophisticated algorithms.

Modern search engines are advanced IR systems.

The picture is of Vannevar Bush's memex, which was an early vision of the Internet, with ready access to all kinds of information.

The first mention of a computer searching for information is 1948 by a guy with an almost-awesome last name: Holmstrom.

Information Retrieval: Precision and Recall

1950's - 1960's: Cranfield Tests by Cyril Cleverdon, UK

Precision: Fraction of the retrieved documents that are relevant to the query.

Recall: Percentage of all of the relevant documents that the query retrieved.

Cleverdon was a librarian and computer scientist

All about measuring the performance of the system, not about the person's experience, feelings, thought process, and learning while they search for information.

Example: Precision and Recall

Initial Search Results
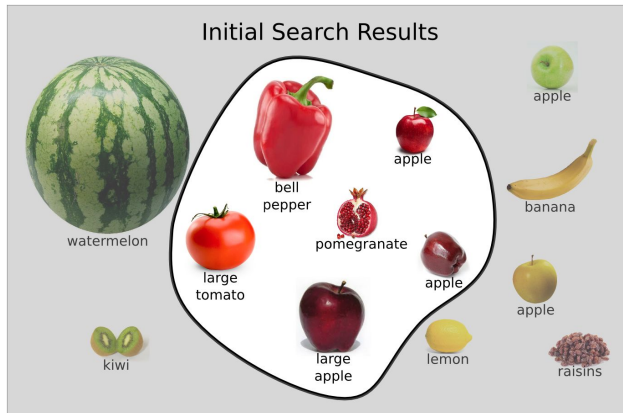
Searching for apples with:

```
(medium OR large) AND red
AND fruit
```

Precision = 3/6 = 50%

Recall = 3/5 = 60%

https://opensourceconnections.com/blog/2016/03/30/search-precision-and-recall-by-example/

Example: Precision and Recall

Initial Search Results
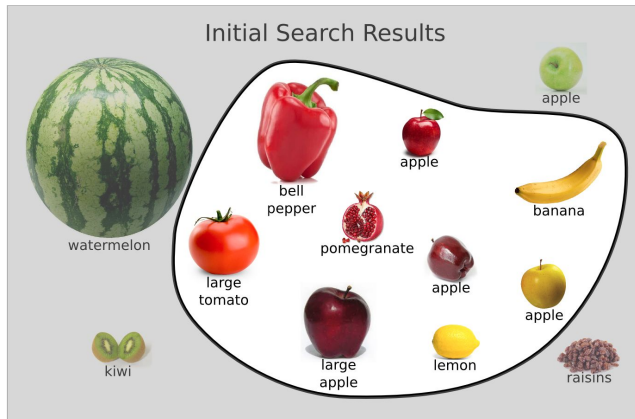
Searching for apples with:

```
(medium OR large) AND (red
OR yellow) AND fruit
```

Precision = 4/9 = 44% ⬇

Recall = 4/5 = 80% ⬆

https://opensourceconnections.com/blog/2016/03/30/search-precision-and-recall-by-example/

Example: Precision and Recall

"Tight" Search Results

apple, bell pepper, watermelon, large tomato, apple, banana, pomegranate, apple, apple, kiwi, large apple, lemon, raisins
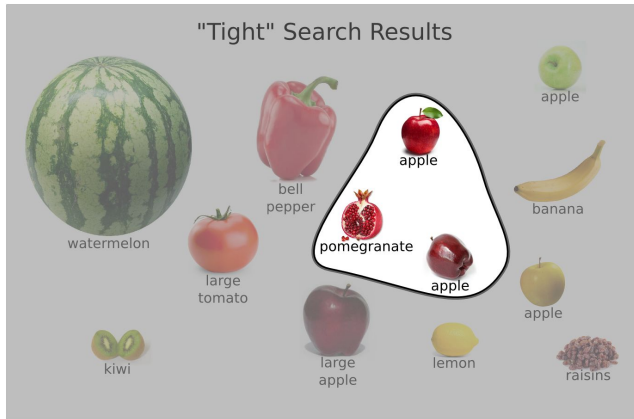
Searching for apples with:

`Medium AND red AND fruit`

Precision = 2/3 = 66% ⬆

Recall = 2/5 = 40% ⬇

https://opensourceconnections.com/blog/2016/03/30/search-precision-and-recall-by-example/

While a person could adjust their search, this is really all about how the system responds to different search terms.

You improve precision and recall by improving the system, not by looking at people's information seeking behaviors.

Advanced Queries: "Learn the System"

Here's an example of optimizing for precision and recall, but information seeking behavior:

```
DTYPE("CLINICAL TRIAL*" OR "CONTROLLED CLINICAL TRIAL" OR "MULTICENTER
STUDY" OR "RANDOMIZED CONTROLLED TRIAL" OR "EQUIVALENCE TRIAL") OR
MESH.EXACT.EXPLODE("CLINICAL TRIALS AS TOPIC")
```

MEDLINE Search Fields

Information Seeking

Differs from Information Retrieval because it **focuses on the person, not the system**.

1950's: Early studies about researchers.

1981: Wilson's model

1980's - 1990's: Most studies, including the Kuhlthau
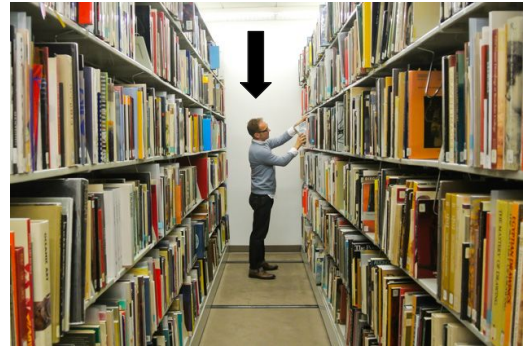
Information Seeking: Wilson

1981: Thomas D. Wilson

"an attempt ... to map the processes involved in what was known at the time as 'user need research.'"
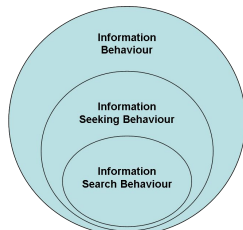
Wilson's Nested Model of Conceptual Areas (Wilson 1999)
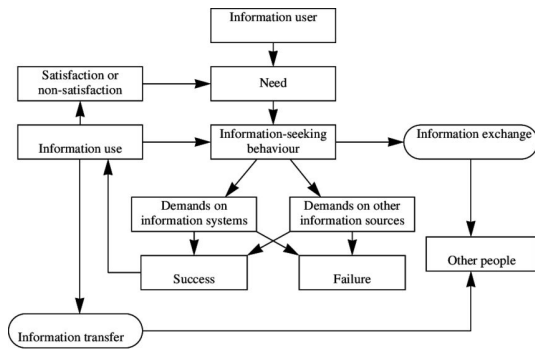
Circles: More and more specific behaviors. Note that Kuhlthau writes about Information Search Behavior, which I'll admit to using somewhat synonymously with Information Seeking Behavior. Information Seeking is broader -- more before and after search and more places to look for information.

The picture on the right is Wilson's first model. He updated it twice in the 1990's. The details are not that important to our discussion today. What's important is that it starts with the user, it describes an iterative process, and that it involves a user, an information system, and information sources.

Kuhlthau's Information Search Process

Published in 1991

Summarizes five experiments to understand how **people** (mostly students) search for information

Presents a six stage model for the search process.

Calvin: "By breaking down ISP into a six-stage process with 3 lenses from which to view the process, I feel that Kuhlthau has constructed an effective model, that is still extremely relevant and whose scope has outlived earlier and more simplistic models."

TABLE 2. Information search process (ISP).

Stages in ISP	Feelings Common to Each Stage	Thoughts Common to Each Stage	Actions Common to Each Stage	Appropriate Task According to Kuhlthau Model
1. Initiation	Uncertainty	General/ Vague	Seeking Background Information	Recognize
2. Selection	Optimism			Identify
3. Exploration	Confusion/ Frustration/ Doubt		Seeking Relevant Information	Investigate
4. Formulation	Clarity	Narrowed/ Clearer		Formulate
5. Collection	Sense of Direction/ Confidence	Increased Interest	Seeking Relevant or Focused Information	Gather
6. Presentation	Relief/ Satisfaction or Disappointment	Clearer or Focused		Complete

Summary: Kuhlthau ran a series of 5 experiments, primarily on students, to understand how people searched for information. She was specifically studying the people's behaviors, not the systems that they used to find information.

She identified a pattern of behavior and developed a six stage model for the information search process, as shown in the figure.

Kuhlthau: Three Realms of Activity

A model representing the user's sense-making process of information seeking ought to incorporate three realms of activity:

- physical, actual actions taken;
- affective, feelings experienced;
- and cognitive, thoughts concerning both process and content.

A person moves from the initial state of information need to the goal state of resolution by a series of choices made through a complex interplay within these three realms (MacMullin & Taylor, 1984).

Note that these realms are all about the person and NOT the system.

What are they doing? What are they feeling? What are they thinking?

We want to understand the answers to these questions at each stage of information seeking.

Kuhlthau: Six Stages of the Info Search Process

1. **Initiation:** Just getting started, don't really know what to do. 😵

Kuhlthau: Six Stages of the Info Search Process

1. **Initiation:** Just getting started, don't really know what to do. 🤔
2. **Selection:** Figure out the general topic, feeling better. 😄

Kuhlthau: Six Stages of the Info Search Process

1. **Initiation:** Just getting started, don't really know what to do. 🤔
2. **Selection:** Figure out the general topic, feeling better. 😄
3. **Exploration:** Investigate the general topic. Feel confused and frustrated as you try to understand what to search for and map what you find to your understanding of the world. 🤯

Kuhlthau: Six Stages of the Info Search Process

1. **Initiation:** Just getting started, don't really know what to do. 🤔
2. **Selection:** Figure out the general topic, feeling better. 😄
3. **Exploration:** Investigate the general topic. Feel confused and frustrated as you try to understand what to search for and map what you find to your understanding of the world. 🤯
4. **Formulation:** Turning point. Search becomes more focused and personalized. 😁

Kuhlthau: Six Stages of the Info Search Process

1. **Initiation:** Just getting started, don't really know what to do. 🤭
2. **Selection:** Figure out the general topic, feeling better. 😄
3. **Exploration:** Investigate the general topic. Feel confused and frustrated as you try to understand what to search for and map what you find to your understanding of the world. 🤯
4. **Formulation:** Turning point. Search becomes more focused and personalized. 😁
5. **Collection:** Know how to search effectively and select the best resources. 😄

Kuhlthau: Six Stages of the Info Search Process

1. **Initiation:** Just getting started, don't really know what to do. 🤔
2. **Selection:** Figure out the general topic, feeling better. 😄
3. **Exploration:** Investigate the general topic. Feel confused and frustrated as you try to understand what to search for and map what you find to your understanding of the world. 🤯
4. **Formulation:** Turning point. Search becomes more focused and personalized. 😁
5. **Collection:** Know how to search effectively and select the best resources. 😄
6. **Presentation:** Relief and satisfaction as you finish the task and write a paper or give a talk. 😁

Kuhlthau: Does It Hold True For You?

Lisa: "I think some of this model still hold true to how I begin a search for information on topics I am unfamiliar with. It may take me awhile to determine what information is relevant to my needs. However, my initial idea of what I think is relevant may change over time and I may move back to an earlier stage in the model for a brief moment or jump to a later stage of the model."

Related to Information Seeking Behavior

Principle of Least Effort or Satisficing

Sense-Making

User-Centered Indexing

Affordances

Principles of Least Effort and "Satisficing"

Zipf (1949), Simon (1956)

When people search for information, they typically:
- don't search until they find all of the information or even the best information
- don't employ the most sophisticated search method

Instead, they:
- search until they have acceptable information
- employ the easiest or most accessible search strategy

Am I the only one around here that tries to do things with the least effort possible and expects a good result?!

someecards
user card

What does this tell us about exhaustive classification systems and sophisticated information retrieval systems?

Sense-Making

Brenda Dervin, 1980's-1990's

When people seek information, they are trying to explain a discontinuity, something that doesn't make sense.

As they seek, they make sense of the discontinuity and acquire new knowledge. As they learn and build a mental model, they iterate and focus their information seeking behavior.

Similar to Belkin's Anomalous States of Knowledge
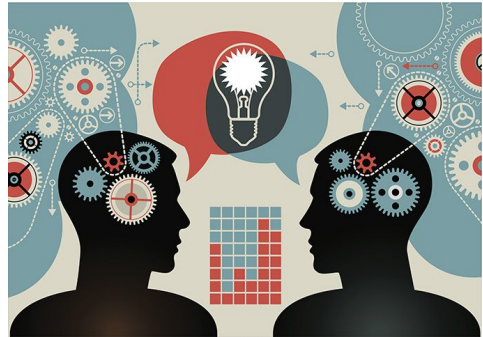
Belin is cited in the Kuhlthau reading

What People Ask Librarians

Can you help me find this book?

It is a green book ...
... about this size, and ...
... I think the author's name was Doug.
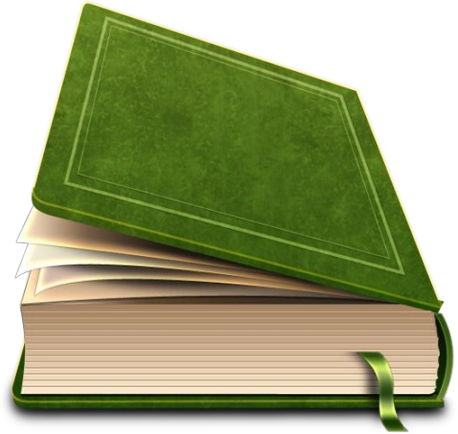
At first, librarians had to know the system.

But research in information seeking behavior helped make systems more user-friendly and more capable of supporting this type of search.

You probably still need a librarian for this particular question, but you can get closer because you can search on more fields and you because search systems help you learn more information as you go -- sense-making.

User-Centered Indexing

Do you index to describe the resources or to anticipate user queries?
Most indexing is document-centric, although some systems adapt to discovered mismatches between controlled vocabulary and regular speech.

Do you index globally or per-user?
Global is the norm because it is document-centric, but automated indexing could support personalized indexes.

A big part of what you can do with user-centered indexing is map user terms to the controlled vocabulary.

For example, suppose that people keep searching for "kitty" and getting no results. They are getting no results because your system has a controlled vocabulary and uses only "cat." If you think about the user, as you do when you consider information seeking baehavior and

Information Seeking Behavior in Gen Y Students

Weiland, 2004

Summary: Ugh, kids these days with their TVs and their Internet and their different learning styles and their lack of cognitive development and their laziness. They don't know how to do proper research.

Sandeep: "I found it condescending that Weiller casually states, 'Most college faculty and librarians are painfully aware of how often students seem to be incapable of thinking critically about …' without citing evidence."

Calvin: "I found the introduction to Weiler's paper to be a little alarmist."

Information Seeking Behavior in Gen Y Students

Weiland, 2004

Sur___ ____g_ __ ___ days with their TVs and their Internet ____ ___ent learning styl__ __ __ __k of cognitive development and the__ ____ess. They don't know how to __ ___ ___arch.

Information Seeking Behavior in Gen Y Students

Weiler, 2004

Okay, okay, that's not really what it is about.

I wanted to acknowledge that some people might have some issues with this paper before we summarize and discuss it.

So, why did I choose this paper?

Information Seeking Behavior in Gen Y Students

Key points:

Motivation, critical thinking, and learning theories are key to understanding information seeking behaviors.

Changes in technology change how we search, what we have access to, and how we process information.

The Internet has emerged as the primary place to seek for information.

Specific to how we search, consider Boolean operators: AND, OR, XOR, etc. These are very important if you have to adapt to the system. But many systems have adapted to how people think and speak and search.

Information Seeking Behavior in Gen Y Students

Drew: "The main thing that stood out to me when comparing the early information seeking behavior models by Kuhlthau and others to the Weiler study was the emphasis on fleshing out things like motivation and the personal skill set of the individual users. Kuhlthau described an almost inflexible information-seeking approach that necessitated the library and interacting with neatly categorized systems. Weiler brought attention to the style of learning and the skills that people develop over time or may be more naturally attuned to.

 "In my mind, this is reflective of the technology present in modern day processes for information seeking when compared to the classical approach. The students in Kuhlthau's study were nearly entirely limited to the library and had to interact with that system in a very straight-forward way. Modern advances in computing platforms have made it easier for content in all formats (video, text, interactive media, audio, etc.) to be uploaded, shared, distributed, etc."

- Dealing with more and more information, plus different types of information

Gen Y Info-Seeking: Can I Trust the Source?

Weiler: "... issues of time and levels of difficulty in obtaining information are usually of more concern to students than issues of accuracy."

Milly: "With current world events and events in politics, as well as increased information access for Gen Y, I think that the perception that many individuals in this age group now have is that the veracity of information is the most important--not how easy it is to obtain."

Is more difficult-to-find information more accurate? Think about the early internet and how anyone could post anything. We've built up some mechanisms for reputation on the Internet, but it isn't perfect.

Note that Weiler noted that only 73% of professors accepted unqualified use of the Internet in research because they feared students would find inaccurate, low-quality information. That number is almost certainly 100% now.

Talk about the rise of misinformation and how this shapes our satisficing search behaviors.

Lisa, Erin, and Drew also discussed this topic in their responses.

Weiler: Verbalizing Search Queries

Erin: "The difficulty that people have in verbalizing complex information needs rings true with the application (or attempted applications) of voice interfaces for conducting searches (like Ok Google, Cortana, and Alexa). Information seeking behaviors are easily achieved through verbalization for simple (mostly fact-based) queries, but more complex queries are much harder to verbalize."

Affordances: Facets, Menus, Refined Search

Affordances: Natural Language Search and Smart Algorithms

Not just natural language and a smart algorithm behind it. You also have the shortcut of displaying the answer directly on the results page.
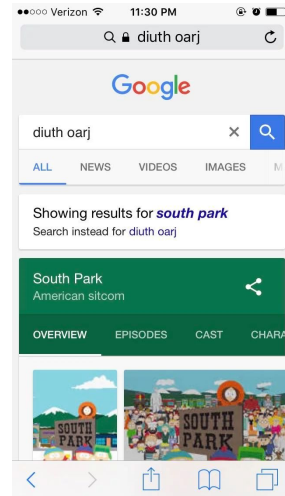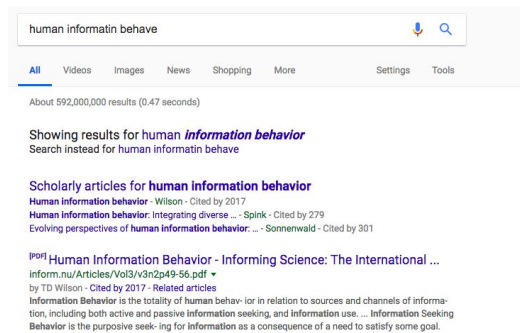
Affordances: Autocomplete

Affordances: Typo Correction

The South Park one was reported by someone who was amazed to get the right result when searching while drunk.

Ideation Activity

Describe your "dream information machine."

- What process would you or the machine follow?
- What affordances would it support?
- What types of algorithms would it support?
- Is it an information retrieval system or an information seeking tool?
- What would be different if it was specifically for an academic researcher?

"Dream information machine" is taken from the Gen Y paper.

Close activity by talking about how academics search differently. The network of citations is somewhat unique to this genre and very important for understanding related works. Some researchers know their journals and stick to just those -- or they wait to see when specific authors have published new papers or books.

Information Seeking Behavior

Information seeking behavior is about **people** and the **processes** (physical, mental, and emotional) that they employ to find information. The people, behaviors, processes, systems, and information involved are continuously changing and adapting, but we'll always want information to make sense of the world.

The best tools balance information retrieval technology and information seeking sensibilities.