How Google Works in 2017: Are Search Engines Really Dumb and Should Educators Even Care?

Paul Barron
pbbarron@gmail.com
paul@duckduckgo.com

All Right Reserved. This presentation may be copied and distributed for nonprofit educational purposes only. Revised 7-1-2017
How do you define “Research?”

For middle and high school teachers the definition of “research” has changed.

That change is reflected in how students approach the research task.

“When asked how their students “do research,” the first response in every student and teacher focus group was ‘Google’.”

“Teachers say, for students today, ‘Research = Google-ing’.”

“How Teens Do Research in the Digital World”
http://pewinternet.org/Reports/2012/Student-Research
Do you agree?

- For the …

... “to Google” is a lifestyle, a behavior, a habit pattern.
If educators hope ...

- To change students’ excessive use of Google, educators must learn how Google works, in order ...
- ... to influence students to integrate Google use with other reliable sources of information.
Learning Objective

- Increase our understanding of how Google works by dispelling some search engine myths.
Learning Objectives: Dispel ...

- Search engine myths: Google ...
  - accepts pay for placement,
  - understands the search query,
  - treats all sites and top level domains equally when determining results,
More Google Myths

- Google …
  - ranks the results based on the **popularity** of the webpage with searchers, and
  - returns the **best answer** to a query in the **first result**.

- Another myth: Students read webpages.

Longwood University 2017 SLI: Curiosity Ignites Literacy
Learning Objectives

- Understand Google’s ranking factors that you - the searcher - can exploit to find relevant results

- Items **on** the webpage
  - Words and their location
    - Title of the webpage
    - URL (webpage address)

- Items **off** the webpage
  - **Links** to a webpage
Myth: Google Accepts “Pay for Ranking”

“At Google we take our commitment to delivering useful and impartial search results very seriously.”

“We don’t ever accept payment to add a site to our index, update it more often, or improve its ranking.”

Matt Cutts
Head of Google’s Web Spam Team
http://goo.gl/S40MJJ
Google accepts payment for advertising.

Teacher School Supplies - ReallyGoodStuff.com
Ad results.reallygoodstuff.com
The Most Trusted Name Since 1992. Designed By Teachers For Teachers!
Bulletin Boards & Post... Classroom Wish Lists
Why Buy From Us? Organizers & Storage

Discount School Supply® - DiscountSchoolSupply.com
Ad www.discountschoolsupply.com/
Shop The Most Extreme Bargains Ever. Savings Over 70% Off!
110% Low Price Guarantee · Fast Delivery · Top Brands · High Quality Pro...
Motor Skills Activities · Teacher's Resources · Extreme Savings · 25% Off ...

Images for "teaching school supplies"

Google’s Ad Revenue Hits $19 Billion
www.adweek.com
July 2016

Longwood University 2017 SLI: Curiosity Ignites Literacy
Why learn how Google works?

“We expect a lot search engines. We ask them vague questions about topics that we are unfamiliar and anticipate a concise organized response.”

“You would have better success if you laid your head on the keyboard and coaxed the computer to read your mind.”

Understanding Search Engines: Mathematical Modeling and Text Retrieval
Michael W. Berry and Murray Browne
From Google’s Senior VP of Search

- “Google interacts with users on purely transactional terms.”

- “We type a **half-formed** thought into the query bar and wade through **blue links** until we find what we are after.”

- “Or we assume if it’s not indexed by Google, it doesn’t exist.”

Is that what your students believe?

The technology giant faces the biggest shift since its founding

[http://time.com/google-now/](http://time.com/google-now/)

Longwood University 2017 SLI: Curiosity Ignites Literacy
We all know that our students ...
The reality today?

“Libraries once seemed like the best answer to the question, Where do I find …?”

“Today the search engine [Google] rules.”

“No Brief Candle: Preconceiving Research Libraries for the 21st Century;” Part II
http://www.clir.org/pubs/reports/pub142/pub142.pdf
Should he have seen the librarian first!

"MY TERM PAPER HAS HYPERTEXT LINKS, FULL-SCREEN VIDEO CLIPS, ANIMATED FONTS, AND AWESOME 3D SPECIAL EFFECTS. . . NOW ALL I NEED IS A TOPIC."
Research Sources for Middle & High School Students

1. Google or other online search engine (94%)
2. Wikipedia or other online encyclopedia (75%)
3. YouTube or other social media sites (52%)
4. Their peers (42%)
8. Online databases (EBSCO, JSTOR, or Grolier (17%)
9. Research librarian at school (16%)

“How Teens Do Research in the Digital World”
http://pewinternet.org/Reports/2012/Student-Research
Myth: Students are Google experts.

“Students overuse Google and misuse scholarly databases. Indeed, they’re not very good at using Google for these purposes.”

“Google’s own research scientists lament that students are unable to take advantage of the resources that are available to those who know how to find them.”

Report of the Commission on the Future of the UC Berkeley Library
http://goo.gl/iKER2f
Unfortunate Facts at UC Berkeley

“At the undergraduate level, what is anecdotally apparent to most faculty and librarians:

- Students lack skills needed to use digital resources for research.
- They are adept at finding information for personal purposes; but …
  - … those skills often aren’t sufficient to accomplish their academic work effectively.”

“Report of the Commission on the Future of the UC Berkeley Library;”
http://goo.gl/iKER2f
UC Berkeley 2017 World Ranking #10

<table>
<thead>
<tr>
<th>2016-17 rank</th>
<th>Institution</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>University of Oxford</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>2</td>
<td>California Institute of Technology</td>
<td>United States</td>
</tr>
<tr>
<td>3</td>
<td>Stanford University</td>
<td>United States</td>
</tr>
<tr>
<td>4</td>
<td>University of Cambridge</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>5</td>
<td>Massachusetts Institute of Technology</td>
<td>United States</td>
</tr>
<tr>
<td>6</td>
<td>Harvard University</td>
<td>United States</td>
</tr>
<tr>
<td>7</td>
<td>Princeton University</td>
<td>United States</td>
</tr>
<tr>
<td>8</td>
<td>Imperial College London</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>9</td>
<td>ETH Zurich – Swiss Federal Institute of Technology Zurich</td>
<td>Switzerland</td>
</tr>
<tr>
<td>=10</td>
<td>University of California, Berkeley</td>
<td>United States</td>
</tr>
<tr>
<td>=10</td>
<td>University of Chicago</td>
<td>United States</td>
</tr>
</tbody>
</table>

https://goo.gl/ppTrPp
Myth: Google Understands Your Query

- “Search engines have no understanding of words or language.”
- “(They) don't recognize user intent and can't distinguish goal-oriented search from browsing search.”
In 2013 Google announced that...

“Google will focus on trying to understand the meanings of phrases and concepts as opposed to matching keywords in a search query to the same words on Web pages.”

“Google Alters Search to Handle More Complex Queries”
New York Times; September 26, 2013
goo.gl/iuEtH8
If Google does not understand …

… the words in my query …

… how does Google know what to look for in the database?
“We analyze what the words in your search query mean.”

How?

“We … try to decipher what strings of words we should look up in the index.”

“to decipher” = to match exact words and synonyms

https://www.google.com/search/howsearchworks/algorithms/
standard deviation
standard deviation calculator
standard deviation formula
standard normal table
standard error
standard error formula
standard form
standard error calculator
standard
standard deduction 2016
Google Matching #2

standards of learning
standards of learning va
standards of nursing practice
standards of care
standards of mathematical practice
standards of practice
standards of excellence
standards of review
standards of proof
standards of conduct
Google Matching #3

standards of learning
standards of learning va
standards of learning test
standards of learning practice tests
standards of learning math
standards of learning test scores
standards of learning science
standards of learning english
standards of learning virginia social studies
standards of learning kindergarten virginia

Google Search  I'm Feeling Lucky
If Google doesn’t understand my query …

… and only matches words, what does Google consider when selecting and ranking the results in response to my query?
Top 2 Ranking Factors of 200

- **Google Senior Search Quality Strategist**

Andrey Lipattsev confirms that:

- webpage content, and

- links to a webpage are the most important ranking factors.

The State of Links: Yesterday's Ranking Factor?
https://moz.com/blog/state-of-links?
And webpage content is ...
Words on the Webpage

- How often do the search terms appear on the webpage?
- Do the search terms appear in the title or the URL?
- Are synonyms for the search terms on the webpage?

Facts about Google and Competition

http://www.google.com/press/competition/howgooglesearchworks.html
An Update to our Search Algorithms (8/10/12)
What Google Considers Off the Webpage

**Links**

- **PageRank** – A measure of the number and the quality of links to a webpage.

- Important webpages receive more links from other webpages.

Facts about Google and Competition
www.google.com/press/competition/howgooglesearchworks.html
Why is knowing the ranking ...  

...factors useful to me?

- **Question Time!**
  - What is the only control that you have to influence the results that Google returns?
And the answer is!

Words in the query.

How would you evaluate your students’ query development capabilities?
Myth: Students craft detailed queries.

“Students perceive themselves as skilled searchers of Google and other search tools.

“[Educators] know that these perceptions aren’t true.”

“Undergraduate students rated their information literacy skills highly, but their search queries and behaviors did not support this. They were not sophisticated users of Google at all, let alone library resources.”

“What Do Librarians Do, Exactly?”
The Informed Librarian Online; http://goo.gl/gTIYFD
“Searchers turn so quickly to Google that they don't think about what they're searching for.”

“It's surprising, we see people trying to find out something about a topic, but never mention the topic.”
Queries by Middle School Students

“A predominate difficulty students experience while performing Web-based research is constructing effective search strings.”

“[M]iddle school students demonstrate unsophisticated skills when constructing search strings, using mainly broad terms and phrases.”

“Internet Searching by K-12 Students: A Research-based Process Model”
http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1b/a8/26.pdf
Queries by High School Students

“[H]igh school students struggle with conceptualizing the topic for their query, sometimes omitting required concepts.”

“Internet Searching by K-12 Students: A Research-based Process Model”
http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1b/a8/26.pdf
College students are better; right?

“[S]earch engines performed poorly because of college students’ inability to construct appropriate search statements.”

Nowicki, Stacy
Student vs. Search Engine: Undergraduates Rank Results for Relevance portal: Libraries and the Academy - Volume 3, Number 3, July 2003
Why isn’t Google’s #1 ranking factor …

... the fact that if the site is popular with us, the site isn’t the first result in Google?
Why not the searchers’ preferences?

"We believe the approach which relies heavily on an individual's tastes and preferences [to rank results] just doesn't produce the quality and relevant ranking that our algorithms do."

Amit Singhal; Google Fellow
“This is tough stuff;” 25 February 2010
http://googlepolicyeurope.blogspot.com/2010/02/this-stuff-is-tough.html
But wait, in 2015 ...

- “User experience is a ranking signal for mobile results, not for desktop results.”

  Google: User Experience As A Ranking Signal Is Currently Only For Mobile Results, Not Desktop Results
  https://goo.gl/r6wmqR

- Why?
  - Do you use your smartphone for academic research?
  - What are mobile phone users searching for with Google?
Why not use popularity for desktop searches?

“We have all been trained to trust Google and click on the first result.”

“How Google Measures Search Quality,” with Peter Norvig, Director of Research at Google Inc.
Datawocky
http://tinyurl.com/6mpt4u

“College students trust Google; they click on the number one abstract most of the time, even when the abstracts are less relevant.”

In Google We Trust: Users’ Decisions on Rank, Position, and Relevance; Laura Granka
Journal of Computer-Mediated Communication
Trusting Google too Much?

“For informational queries … if a result on page 4, provides better information than the results on the first three pages, users will not know this result exists!”

“Therefore, usage behavior does not provide the best feedback on the rankings.”

“But we are the best results!”

“How Google Measures Search Quality”
Datawocky
http://tinyurl.com/6mpt4u
Google’s Laura Granka, Ph.D.

Laura Granka, PhD
User Experience Lead

currently I lead and manage a large team of user experience researchers and software engineers at Google, spanning across Search, Geo, and Identity.

previously I joined Google in 2004, as a researcher on Search. Since then, I have worked in multiple areas of the company, including three years managing the user experience research team across Google Social.

education I completed my PhD at Stanford University, and both my M.S. and B.S. at Cornell University. See my resume for complete work and research history.
How do students decide what to believe?

- Question! What factor loomed largest?

  - “A site's placement in the search results.”
  - “Students ignore a site’s sponsoring organization and an article's author, blindly trusting the search engine to put the most reliable results first.”

Trust Online: Young Adults’ Evaluation of Web Content
http://webuse.org/pdf/HargittaiEtAlTrustOnlineIJoC10.pdf

Longwood University 2017 SLI: Curiosity Ignites Literacy 45
More Blind Trust!

“Kids won’t click to the second page of Google results because they expect Google to place what they need at the top [of the search results].”

Study Ties College Success to Students’ Exposure to a High School Librarian
http://goo.gl/9JBPrZ
Let's test that expectation.

Google returns all sites with the words, 

*martin* and *luther* and *king* and *school* and *flyers*
Google’s 1st Result (7-1-2017)

Download flyers to pass out at your school. - Martin Luther King Jr. - A ...

www.martinlutherking.org/flyer.html

Bring the dream to your school! This is the time of year many students are asked to write reports about Martin Luther King and the civil rights movement. Print out ...

Images for martin luther king school flyers

More images for martin luther king school flyers  Report images
Martin Luther King, Jr.
A True Historical Examination

That night King retired to his room at the Willard Hotel. There FBI bugs reportedly picked up 14 hours of party chatter, the clinking of glasses and the sounds of illicit sex - including King's cries of "I'm f-ing for God" and "I'm not a Negro tonight!"

Newsweek Magazine
January 19, 1996
Page 62

Martin Luther King.org Homepage
Martin Luther King.org is hosted by ...

… “a community of racial realists and idealists.”

Welcome to Stormfront.

We are a community of racial realists and Idealists. We are White Nationalists who support true diversity and a homeland for all peoples. Thousands of organizations promote the interests, values and heritage of non-White minorities. We promote ours.

We are the voice of the new, embattled White minority!

Tell the truth and fear no one!
The student wants to know ... 

Why was that site returned as the 1st result among the 857,000 results!?! 

I thought Google always returned the best results.
Checking for .edu **Links** to the Webpage

Remember the importance of PageRank which measures the number and quality of *links* to a webpage.

- **Link** Check – Returns results that are linked to a site; for example, .edu sites that are linked to Martin Luther King.org.
QUESTION
By reviewing the webpage description can you determine the purpose of the .edu sites’ linking to Martin Luther King.org?
The Value of Quality Links

“With PageRank, five or six high-quality links from websites would be valued much more highly than twice as many links from less reputable or established sites.”

Librarian Central
How does Google collect and rank results?
http://www.google.com/librariancenter/articles/0512_01.html
“Google places a heavy bias on informational resources; .edu and .gov sites tend to rank higher than others.”

“Google is the best at determining true link quality and places a lot of weight on domain trust levels.”

“Can You Please Them All?”
http://www.bruceclay.com/blog/archives/2006/08/can_you_please.html
When reputable webpage author(s) repeatedly link to a webpage, or when highly regarded universities, governments, or organizations, link to a webpage, the rank of the linked-to webpage increases, regardless of whether the page is relevant.
“Tailor-made for skimming, scanning and using the “find” function rather than reading slowly and thoughtfully.”

In 2017 users read at most 28% of the words on the page.
Myth: Students read webpages?

- **1997** - “They don't. Users scan webpage.”

- **2006** - Users scan webpage
  - Users focus on words at the top or left side of the page, while barely glancing at words that appeared elsewhere.

How Little Do Users Read?
http://www.nngroup.com/articles/how-little-do-users-read/
From 2005 to 2014

The average user scanned more results in 2014 vs. 2005, but spent less time looking at each result before clicking a result.

THE EVOLUTION OF GOOGLE SEARCH RESULTS PAGES & THEIR EFFECTS ON USER BEHAVIOUR
www.mediative.com
And in 2017

What explains the change in the scan pattern?
Consider this ...

“The computer screen is … literally a small thing [that] may display just over 300 words. If this world becomes our reality, we actually are relying on less information, not the more that is available.”

“The Google-ization of Knowledge”
Natasja Larson, Laura Servage, and Jim Parsons; Faculty of Education; University of Alberta
http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/28/03/99.pdf
Do you agree that...

“There are **consequences to** our students and our educational system **if we allow Google to** define the parameters of effective research.”

The University of Google: Education in the (Post) Information Age
Tara Brabazon
“In universities the Google Generation do the dumbest things you can possibly imagine.”

“Scholarly searching is not an intuitive skill; students cannot learn by imitating peers.”

“That is where librarians [and classroom teachers] come in; teach them what is possible.”
Google Helpouts

- Search Help Center
  - https://support.google.com/websearch#topic=3081620
  - https://goo.gl/Vot32N

- Advanced Search Options
  - https://www.google.com/advanced_search
  - https://goo.gl/vGcSrY

- Refine Web Searches
  - https://support.google.com/websearch/answer/2466433
  - https://goo.gl/CDc1P2
Videos

- Watch Google's Top Ranking Engineer, Paul Haahr, Talk On How Google Works
  - https://goo.gl/bjm8i1

- How Search Works
  - https://www.google.com/search/howsearchworks/

- How Does Google Search Work?
  - https://www.youtube.com/watch?v=KyCYoGusqs

- How Search Works
  - https://www.youtube.com/watch?v=BNHR6IQJGZs