Search Engines: Page Rank and Retrieval

Unit 3. Information and the Internet

Revision Date: Jan 12, 2020
Duration: 2 50-minute sessions

Lesson Summary

Summary

This lesson has two main objectives.

The first focuses on search engine algorithms and the impact search engines have on our lives. Search engine page rank algorithms rely on many factors to predict what someone is looking for. The business advantage of appearing on the front page of a Google search is tremendous. However, as more information is tracked about our interests and preferences in order to customize the results of our searches, we have to ask whether or not the loss of privacy is worth the results.

The second objective is to introduce students to creating a visual artifact (knowledge required for performance tasks). Students will research a page ranking subtopic, prepare a one minute speech, and (if possible) create a video to accompany the speech.

Outcomes

- A presentation guides the discussion of how search engines work, what page rank is, and how results differ for a variety of reasons.
- Students will understand what metadata is, its relationship to data, and its uses.
- Students then do a quick research project to gather information on assigned, related topics
- Students create a 1-minute speech and presentation on their chosen research topic. If the classroom has the equipment, they will create a video artifact on their topic to share with the class either in the classroom or as homework posted online

Overview

Session One

1. Getting Started (5 min) - Journaling on online search methodology
2. Activity (45 min) - Students discuss page rank and begin research for presentations

Session Two

1. Getting Started (5 min) - Journaling on search engine data mining and introduce presentation activity
2. Activity (35 min) - Create Presentations/Videos on previously researched topics.
3. Presentation of videos or talks (10 min)
4. Optional Homework: Have students watch the rest of the videos made and write about what they learned.

Session 3 New for 2020-2021

1. Getting Started (5 min) - Think-pair-share
2. Activity (35 min) - Metadata exploration
3. Wrap Up (5 min) - Metadata and Privacy

Learning Objectives

CSP Objectives

- *EU DAT-2 - Programs can be used to process data, which allows users to discover information and create new knowledge.*

- LO DAT-2.A - Describe what information can be extracted from data.
- LO DAT-2.B - Describe what information can be extracted from metadata.
- LO DAT-2.C: - Identify the challenges associated with processing data.
- LO DAT-2.D - Extract information from data using a program.
- LO DAT-2.E - Explain how programs can be used to gain insight and knowledge from data.

- *EU AAP-2 - The way statements are sequenced and combined in a program determines the computed result. Programs incorporate iteration and selection constructs to represent repetition and make decisions to handle varied input values.*
    - LO AAP-2.A - Express an algorithm that uses sequencing without using a programming language.

- *EU IOC-1 - While computing innovations are typically designed to achieve a specific purpose, they may have unintended consequences.*
    - LO IOC-1.B - Explain how a computing innovation can have an impact beyond its intended purpose.
    - LO IOC-1.F - Explain how the use of computing can raise legal and ethical concerns.

- *EU IOC-2 - The use of computing innovations may involve risks to your personal safety and identity.*
    - LO IOC-2.A - Describe the risks to privacy from collecting and storing personal data on a computer system.

Key Concepts

Students will understand that the page rank algorithm depends on many factors, has changed over time, and has a large impact on the traffic that a site gets.

Students will give examples of how their activity online is tracked and how the knowledge of them is used to taylor the results and the possible repercussions.

Students will create an artifact using screen capture of themselves discussing and analyzing an aspect of searching.

Essential Questions

- How can computing extend traditional forms of human expression and experience?
- How can computation be employed to help people process data and information to gain insight and knowledge?
- What is the Internet, how is it built, and how does it function?

How can computing extend traditional forms of human expression and enhance people's ability to find information and solutions?

Teacher Resources

Student computer usage for this lesson is: **required**

In the Lesson Resources folder:

- "Search Engine Background" NOTE: This document explains the content of each slide in the presentation WITH answers to the questions in the presentation.
- "PageRank" - slide presentation
- Handouts for students (can be placed on the student's drives or printed out on paper):
    - "PageRank Student Handout" (optional notes to go along with the PowerPoint - gives students a place to answer questions posed in the presentation)
    - "1 minute talk directions.odt" (to help students organize their video)
    - "Sample 1 minute script on keyword matching"
- "Sample 1 minute video artifact on keywords.swf" (1:15)

Online Videos:

- "How Search Works by Matt Cutts" https://www.youtube.com/watch?v=BNHR6IQJGZs (https://www.youtube.com/watch?v=BNHR6IQJGZs) (review of spiders, how they follow links and fetch the pages to index. followed by page rank and spam avoidance) (3:14)

Sites Used in this Lesson:

- Screencast-O-Matic http://screencast-o-matic.com/ (http://screencast-o-matic.com/) Used by students to create an video artifact of their screen to accompany their speech.

- Jing http://www.techsmith.com/jing.html (http://www.techsmith.com/jing.html) Free download for screen capture to create artifact to accompany speech.
- PageRank Checker http://checkpagerank.net/ (http://checkpagerank.net/) Students use this page to check the page rank and other information about web sites.

Lesson Plan

Session 1

Getting Started (5 min)

Students should take a few minutes to journal on the following question:

Which are you more likely to do if you don't see an answer to a search request on the first page: click forward to page 2 of the results or ask the question differently? Why?

(Encourage students to discover that it is very valuable to business to appear at the top of the search engine rankings and that often thousands or millions of results are returned in a single search.)

Activity (45 min)

Part 1 (25 min) - How Search Engines Work

(Use the PageRank presentation in the lesson folder to guide the discussion.)

Note: Guidelines for the teacher are in the "Teacher Notes on PageRank Presentation" document. This document also contains an answer key. (*Students can record their notes in the "PageRank Student Handout".*)

1. Watch the three-minute video on Google search closely to pick up details. Pause, take notes, and discuss as needed.
2. Emphasize that results are found in the database, and then information filtering systems, which take large data sets and eliminate data that is not of interest, help find the most relevant information and identify patterns in the information to better organize it.
3. Allow students to generate ideas on why one webpage might have a higher PageRank than the other. [slide 3]
4. Look at the HTML code of the webpage in the PowerPoint to discover the frequency of keywords including synonyms, and occurrence in titles and metatags. [slide 4] (*Student handout also has a printout of the HTML code for students to get a closer look.*)
5. Read *Fast Fact* the Protect PII on Social Media (https://www.doncio.navy.mil/ContentView.aspx?id=4383) published by the US Navy.

   What concerns does the Navy have about the information posted online?

   Do these same concerns apply to people not in the military?

6. Discuss possible reasons why two different people can get different results doing the same search.
7. Read the first three sections *Internet is forever, except when it is not (https://askleo.com/the-internet-is-forever-except-when-its-not/)* from Ask Leo.com.

   How hard is it to permanently delete information once it is posted on the internet?

   What advice would you give to a young person about posting their information online?

Part 2 (20 min) Preliminary Research for creation of video artifacts/PowerPoints

1. Assign topics for research to student groups. (There are additional topics in the "Search Engine Background" information document if desired). Here are several suggestions:
   - What are additional factors in page rank?
   - What do people do to achieve SEO? (search engine optimization)
   - What is Google bombing? How does it work?
   - How much storage is needed to store Google's index? How many server farms are needed to store it all? What is the design philosophy of server farms?
   - What's the environmental impact of server farms? How do they try to stay green?
   - How does advertising affect search engines? Is it necessary? What is "pay per click" and "click fraud"?

- How is Google getting good at finding things like pictures, videos and other kinds of information beyond just words?
- How do directories work? Show some examples, such as https://dir.yahoo.com/ (https://dir.yahoo.com/) .

2. **If you have video recording equipment:** Demonstrate how to create a high-quality video artifact (the kind students might choose to create for their Performance Task).
   - Student handout: 1 minute talk directions -- go over this handout with them.
   - Show the Sample 1 minute video artifact on keyword matching (in the Lesson Resources folder).
3. With the remaining time, have students begin their research on their chosen topic.

Session 2

Getting Started (5 minutes)

Journal (3 min)

Why could it be beneficial for a search engine to keep track of what people are searching for? In what ways do computers enhance our ability to solve problems? Discuss.

(Possible answer to lead students toward: Topics sporadically become popular, and knowing what results people like can make it easy to suggest sites to others looking for similar things. History data can also enable a search engine to suggest a search phrase when a single word or only a few letters are typed in. The better a search engine knows what you are looking for, the better it can filter results to include results relevant to your query.)

Introduction (2 min)

Explain that students will be creating a presentation on the topic they researched in the last session. This presentation should be scripted, and make use of a PowerPoint and sources from the internet. They will have 30 minutes to make this presentation. (Slide 8 is made for video creation, but works well for general presentations too.)

For classes with enough video recording equipment for all groups:

Explain that students will create their own video explanations of how one feature of search engines works. Go over the "1 minute talk directions.odt" together to help students organize their video. The creative design process to develop such a short, focused product requires good teamwork, organization, and creativity. Plan out what the key message is, what visuals will add the most value, and then craft the wording to fit within the 1 minute time frame.

Activity (35 minutes)

Students should split into their groups and begin work. Allow only 10 minutes for additional research as needed. They will take the remaining 25 minutes to:

a. Make a PowerPoint, gather search results to use as examples, create a rough script, and practice their presentation.
b. **If they are making a video:** Write a script, either create a PowerPoint to voiceover **or** choose some search results to analyze, and practice. The last 5-10 minutes should be used to record a 1 minute video clip of their presentation allowing for multiple retakes)

Presentation of videos or talks (10 minutes)

Show as many videos/ group presentations as you can share with the class. If there are videos, assign the remainder to be watched as homework and have students bring in notes on the key points learned from each video.

Session 3 New for 2020-2021

Getting Started

Warmup:

1 What can we conclude from the following two statements?

- Computers store web pages.
- Everything stored in a computer is stored in bits.

2 How do you think the computer can tell which bits are for the title of a page and should be large and which bits represent plain text on the page and should be smaller?

Introduction

Say: According to the Dublin Core Metadata Initiative Metadata Basics page (), metadata is "data about data" -- specifically, descriptive metadata -- is structured data about anything that can be named, such as Web pages, books, journal articles, images, songs, products, processes, people (and their activities), research data, concepts, and services. Creative Commons Attribution 3.0 Unported License (http://creativecommons.org/licenses/by/3.0/)

In this session, we will investigate what metadata is, its relationship to data, and its uses.

Activity

Part 1

Watch the first 40 seconds of this video to complete these two sentences.

Metadata are pieces of information that make studies _____ & _____.

Metadata helps you make_____ of the _____.

https://youtu.be/-4_MFhi4GpU (https://youtu.be/-4_MFhi4GpU)

Watch the first 90 seconds of this video. After watching the Meta... What? Metadata! (https://youtu.be/3sLKVYYOM40) video, define metadata in your own words.

Say: Let's answer two questions about the metadata for this image.

- What is it used for?
- What are the consequences of changing it?

.

Colin Hines www.ColinHinesPhotography.com [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0/ (http://creativecommons.org/licenses/by-sa/3.0/))]

Now view its metadata.

Think-pair-share: Discuss with your elbow partner

- What is it used for?
- What are the consequences of changing it?

Part 2

Say: If metadata tells us what data means - what information it contains - then what tells us what metadata means.?

Consider the XML for these messages.

```
<messages>
  <note id="23">
    <to>Marcus</to>
    <from>Lin</from>
    <heading>Reminder</heading>
    <body>Don't forget our appointment!</body>
  </note>
  <note id="24">
    <to>Lin</to>
    <from>Marcus</from>
    <heading>Re: Reminder</heading>
    <body>What meeting</body>
  </note>
```

</messages>

Ask: How does the metadata make using the data itself more effective?

The metadata tells us what the data (in black) means. To find out what XML, means we need to consult a standard. Examine the standards recommended by the W3C for XML (https://www.w3.org/TR/xml/).

Say: On a web page the data is what people need to see. The metadata is what computer systems use to present the data in useful ways. One large computer system is the internet. Every message/file transmitted on the internet is first broken into small fragments and each fragment packed in packets. These packets contain a lot of metadata that make delivery of the packets by the routers (computers) of the internet both possible and reliable.

Part 3

Visit the IP packet structure web page (https://www.freesoft.org/CIE/Course/Section3/7.htm)and examine the metadata that is part of every packet transmitted on the internet and answer these questions.

- What two addresses are part of the metadata in every packet?
- What part of the metadata is used to put the fragments back in the order they were before the message/file was broken into fragments?

Wrap up:

Watch this video from International Privacy.org (https://youtu.be/xP_e56DsymA) then discuss these two questions with your elbow partner.

- What concern should non-technically trained people have about metadata?
- What would happen if metadata was banned successfully?

Options for Differentiated Instruction

For a shorter class, don't have students take notes, just discuss the slides.

For the Explore performance task, each student should be able to create their own artifact. You could have the students work on the presentations individually in this lesson, as a practice for the Explore task, if your class is fairly competent with the technology. For students with less experience (or to save time during presentations), it could be beneficial to have students create these artifacts in pairs, with some pairs repeating topics for comparison.

Evidence of Learning

Formative Assessment

Students share best definitions of page rank related terms

Students analyze web pages for reasons for differences in page rank

---

Summative Assessment

Students create a one-minute video clip on a topic related to the operation of search engines.

 (http://www.umbc.edu/)   (http://www.umd.edu/)   (http://www.nsf.gov/)