

## PowerPoint Presentation Script

### Slide 1:

- What are algorithms? Put very simply, algorithms are a set of step by step instructions. In the realm of computer science, algorithms are a specific set of instructions and operations that a computer program carries out to accomplish various tasks. For example, a user of a computer program could be asked to enter in 3 random numbers. The program could then follow a set of specific steps in order arrange the numbers from smallest to largest.
  - Examples of said steps:
    - Ask the user to enter 3 numbers.
    - Look at the numbers in the order that the user entered them.
      - Is the first number larger than the second number?
    - If so, swap the 2 numbers.
    - Otherwise, do nothing.
      - Is the second number larger than the third number?
    - If so, swap the second and third number
    - Otherwise do nothing.
      - Repeat steps 2-4 until correct order is achieved
      - Return new sequence to user

### Slide 2:

- Why are algorithms important?
- Scientists and engineers have to make use of their time wisely. Well-built algorithms help optimize time by finding and programming the quickest path to get something done. For example, Instagram uses an algorithm which determines which posts to show you at the top of your feed, based on your browsing habits. This way, you spend less time looking for posts that you might be interested in. Algorithms can be used to sort a large set of information based on a set of structural rules, such as step by step instructions. For example, usually when you search for something on Google, there are many results, even pages and pages of results. However, more often than not, you never have to venture off that first page of results, because the most important and relevant items are on that first page. This is because of a specific algorithm that Google uses in order to sort through and figure which websites are the most important, based on relevance and website rank. We will go into further depth of this specific algorithm later on.

### Slide 3:

- First let's look at a real-life example of an algorithm for a task that doesn't require a computer at all: making a grilled cheese sandwich. When you make a grilled cheese, you generally follow specific steps to reach the desired outcome. First you need bread. Then you need to butter the bread. Cheese is then added. The final step is to cook the grilled cheese.

### Slide 4:

- Only upon completing all of these steps are you able to enjoy the final result. This is exactly how an algorithm works: it follows steps.

Name:

Date:

Class:

**Slide 5:**

- Google PageRank is a more complex example of an algorithm. This is the algorithm that Google uses to determine the *importance* of a website, based on how many websites are linked to that website.
- The algorithm assigns each webpage a number, called a *PageRank* (or *PR*), on a scale from 1 to 10. Higher numbers correspond to higher importance, which means the higher that webpage will appear in search results.

**Slide 6:**

- The PR is one of the main factors that Google uses to determine what order to show you the search results when you search for a website. Another important factor is *relevance*. For example, if I were to search for “soccer” notice how there are no pictures of bananas in the search results. This is because bananas are not relevant to my search query, but soccer balls, soccer players, and soccer fields are relevant.
- In general, Google shows you the most relevant and highest PR websites first. Relevance is actually computed by a separate algorithm used by Google that aids to organize data. The higher a website’s PR is, keeping the relevance of the search in mind, the higher it will be in the search results (if it is relevant to your query).

**Slide 6 (Higher Levels ppt) & Slide 7 (Lower Levels ppt):**

- How does Google determine which websites are the most important, and deserve the highest PR? Interestingly, this is based on looking at how all of the websites link to each other. As you know, when you are visiting a website, you will often see links to other websites. A website reviewing bicycle helmets might include links to Amazon, where you can buy those helmets. An article about a soccer team might include links to ESPN bios for each player. And so on.
- In general, important websites should be those that have *lots of other websites linking to them*. However, PageRank does not merely involve *counting* how many links are coming into each website. Instead, PageRank says that important websites should be those that have lots of other *important* websites linking to them.

**Slide 7 (Higher Levels ppt):**

- Consider the example shown. There we see two websites (one in red and one in blue), each with two other websites linking to them. However, the red website has two very important government websites linking to it, whereas the blue website has two less important online shopping websites linking to it.

**Slide 8 (Higher Levels ppt):**

- So, in this case, Google’s algorithm should assign a higher PR to the red website. Importance of a website is determined by credibility of that website. Credibility here means how likely it is that the information you are gaining from that source is true or not. For example, if I googled “what kind of fabric are astronaut suits made of?”, Google would give NASA a higher page rank because they would be a more credible source for that information than a clothing store would be.

Name:

Date:

Class:

**Slide 9 (Higher Levels ppt):**

- [Examples of different PRs.] Notice how Facebook and YouTube have perfect PageRanks, while the website for the Sweetbriar only has a PR of 1. This is because websites like Facebook and YouTube have many other important websites linking to them, while the Sweetbriar website is from a small organization, so significantly fewer websites link to its website. Warning: Note that a high PR does *not* automatically mean that a website has good information or is necessarily more credible. All of this is simply determined by which websites link to which other ones.

**Slide 10 (Higher Levels ppt) & Slide 8 (Lower Levels ppt):**

- Here we see an example of how PR can be determined simply by looking at which websites link to which other ones; these links define how they are connected in a network. In this example, pretend that each circle represents a website, and the arrows represent links from one website to another. Scientists and engineers often use pictures like this to understand *networks* of interacting machines, websites, etc. So, in this example, both red websites link to the blue one, one red website links to the green one, and the blue website also links to the green one. In this example, the red websites should have the lowest PR, since no other websites link to them. This is indeed the case, and so the red circles are drawn to have the smallest size (the size of each circle represents its PR). The blue website is the next largest (next highest PR), since it has links from two red websites. And the green website is the largest (highest PR) of them all, since it has links from one red and one blue website.

**Slide 11 (Higher Levels ppt):**

- This can quickly become very complicated! Here we see a more complex example of a network of websites. Here, the yellow circle has 6 links coming into it, whereas the light blue circle only has one link coming into it; so why does the light blue circle have a larger PR than the yellow circle? It is because the one link to the light blue website is from the red website, which has the highest PR among all of them. This is more important than all of the links to the yellow website from other websites with low PR. Another thing to note is that the number of links coming *out* of a website does not affect its PR.
- This leads to an interesting problem: to determine a website's PR, you need to know the PR of the websites that link to it. But to determine the PR of those sites, you need to know the PR of all the websites that link to those, and so on creating a vast hierarchy of interconnected links. This seems like a never-ending process.
- To compute PageRank, Google actually uses a very clever computer program that is based on mathematical concepts from a field called "linear algebra". Linear algebra is a topic that is used widely throughout science and engineering and is taught in college.
- This lesson continues with an associated activity called "Acting Like an Algorithm". In that activity, we're not going to talk about linear algebra or the actual steps inside Google's algorithm. Instead, it turns out there is a very fun trick we can use to come up with the same answer as Google's algorithm. This trick involves simple concepts of probability and randomness, combined with the iterative process of a ball game. By playing a game, we can determine which websites in a network have the highest PR.