# Homework 6 Search engine indexing and PageRank

1. Chris owns a small online bicycle store selling bike parts and accessories.   
   Customers often find the store via a web search using a search engine.  
   1. Explain the purpose of web crawler software used by search engines. [2]

To get all the information and linked pages to the pages already on the search engine’s index. For example, in google, the web crawler goes on google’s index, and visits the sites already on the index, it then fetches all the sites and pages that are linked to the site it is visiting, and adds it to the index including information about them, like what the sites are about.

* 1. In the construction of the website, Chris included metatags in each web page.
     1. Give **two** examples of metatags that Chris may have applied to the   
        home page of his website. [1]

Bike parts

Bike accessories

* + 1. Explain how metatags are used in the search process. [2]

**The website will be on an index of a search engine, and the details of the website include it’s metatags. When people search for something that is related or part of the metatags in his website, depending on his website’s relevancy, it will be ranked and displayed to the searcher**

1. Describe what is meant by search engine indexing, and how this affects the user   
   of a search engine [4]

Search engine indexing is using databases of websites to find the webpage the user is looking for. The database contains information on the websites in the database, including metatags and descriptions that tell the search engine what each webpage is about. This proves more useful results to the user, saving the time from having to manually find through many many web pages for the type of webpage they are looking for. Web crawlers are used to add to the index, they visit websites already in the index and the websites linked to these already present websites, and add them to the index, so that they can be referenced if the user needs to see them.

1. The PageRank algorithm is used by Google to decide how relevant a web page is. The original algorithm is:

**PR(A) = (1-d) + d (PR(Ti)/C(Ti) + … + PR(Tn)/C(Tn))**

* **PR(A)** is the PageRank of page A
* **PR(Ti)** is the PageRank of pages Ti which link to page A
* **d** is the damping factor
* **C(Ti)** is the number of outbound links on page Ti
  1. The algorithm uses a damping factor *d*. Explain the purpose of *d*. [2]

To prevent page rank of inbound site divided by number of outbound sites of that inbound site from having too much influence in determining the page rank. It also means that the average user will stop their session or enter a new address after 6 click through links.

* 1. Suggest three factors that affect the rank of a web page. [3]

Number of inbound sites/pages

The number of outbound sites the inbound sites have

The metatags that the web page has

Total 14 Marks