

Assignment 2

Linguistics 445/515
Python & Searching

Due Monday, September 29, 2008

1. In this question, you're going to critically evaluate different kinds of search engines and sources of information. Furthermore, you will be evaluating your own ability to search, through the use of accurate search terms and boolean expressions.
 - (a) Identify a (research) topic that you need more information about. In one paragraph, give a clear description of what you want to find and identify highly specific pieces of information you want. For example, wanting more information on the *yeti* is too vague: wanting to know its shoe size, its location(s), and organizations that serve to protect it are much better topics. So, be clear and be specific.
 - (b) Evaluate the capabilities of the following resources, in terms of accuracy, reliability of information, helpfulness, ease of use, speed, amount of distractors (e.g., paid links), and any other criteria that you deem relevant.
 - i. <http://www.google.com>
 - ii. <http://www.ask.com>
 - iii. <http://en.wikipedia.org/wiki/WP:RD>
 - iv. <http://www.iucat.iu.edu/>
 - (c) Of course, how good each resource is depends a bit on you. What queries did you find to be most successful in your search? Why? Which were the least successful? Why? Did your successful queries vary by resource?
2. Below is a list of the authors of 7 hypothetical books. So, for example, *knight* and *davis* wrote book 4, but *crean* did not.

- | | |
|-----------------|-----------------------|
| 1) knight | 5) knight crean |
| 2) davis | 6) davis crean |
| 3) crean | 7) knight davis crean |
| 4) knight davis | |

For each of the following Boolean expressions, write down which author lists match the expression. For example, the Boolean expression *davis* matches 2, 4, 6, and 7. Additionally, for (d) and (e), draw a Venn diagram representing the query.

- (a) knight
- (b) knight AND crean
- (c) knight OR davis OR crean
- (d) knight AND (davis OR crean)
- (e) (knight AND davis) OR crean

3. Your name is John Smith, and you have a bit of a problem: no one can find your homepage.
 - (a) Knowing what you know now about how search engines work and how they rank results, how would you make your website more findable?
 - (b) If you were to add structure to the internet, what structure would you propose to make your page more findable? That is, should there be some sort of categorization to webpages and what should that categorization be?
4. When we talked about how weblinking is used to rank webpages, we mentioned how one page's popularity helps determine another's. But there's a problem: where do we start? How can we know another page's popularity if it, too, is based on other page's popularity.

Assume that we initially assign every page a score of 1.

- (a) If every page has the same score, namely 1, what does that mean for our page ranks? In other words: how is a page's score now calculated?
 - (b) Let's say we start with a score of 1, but we want to figure out different scores for different pages. How do we go about doing that? What is the next step?
5.
 - Make sure you have access to a computer that has Python.
 - Write a short program in Python which takes a string variable, calculates its length, and then prints out one letter per line (hint: string slices). In this case, your variable will contain the string "Hello world", but we might want to change that in the future, so be sure to use a `while` loop.

Your output should look like this:

```
H
e
l
l
o

w
o
r
l
d
.
```