

In-Class Assignment 4

Searching

Wednesday, January 23, 2008

Boolean Searching

The following is a list of “webpages”:

- | | |
|-------------------|------------------------------|
| 1. rock paper | 7. rock paper scissors |
| 2. rock scissors | 8. rock paper bomb |
| 3. rock bomb | 9. rock scissors bomb |
| 4. paper scissors | 10. paper scissors bomb |
| 5. paper bomb | 11. rock paper scissors bomb |
| 6. scissors bomb | |

Which of the webpage numbers do the following queries match?

e.g. “rock AND bomb” matches 3, 8, 9, and 11.

- rock OR paper OR scissors
- (rock AND paper) OR (scissors AND bomb)
- (rock OR paper) AND (scissors OR bomb)
- rock AND (paper OR (scissors AND bomb))
- ((rock AND paper) OR scissors) AND bomb

Regular Expressions

1. What do the following regular expressions match?

(a) `/m(a|e|i|o|u)d/`

(b) `/m(a|e|i|o|u)d?/`

(c) `/m(a|e|i|o|u)?d/`

(d) `/qwer*ty/`

(e) `/qw(er)*ty/`

(f) `/qw(er)+ty/`

(g) `/qwer.ty/`

(h) `/qwer.*ty/`

(i) `/qwe(r.)*ty/`

2. Write regular expressions which match the following lists of strings:

(a) cabin, kabin

(b) cabin, kabin, cabins, kabins

(c) cabin, kabin, cababin, kababin, cabababin, kabababin, ...

(d) a, as, asd, asf, asdf, ad, adf, af

3. Go to <http://www.lexmasterclass.com/exercises/regex/index.html> and do Exercises 1-4. Note that you can try out the regular expressions you type in by clicking on the Submit button. (Warning: these are fairly difficult!)

If you are unsure about what to do, take a look at Exercise 1 first. Some possible answers are:

```
.*p(o|i|a|t| ).*  
.*p[^e]t.*  
.*p[^er]+t.*
```

If you type any of these in and click submit, you can see how the left column is completely red and the right column is completely black. You can type in any regular expression and see which parts of the strings match. If you find these too difficult, just try out different regular expressions to see what you are able to match without worrying about one column or the other.

Exercise 2:

Exercise 3:

Exercise 4: