Advanced Computer Applications

Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**What is Boolean Searching?**

 When you search for something with a search engine and search for a single term, you are likely to get thousands of hits, most of which are useless. Adding more related terms may help improve the hits at the top of the list, but it will also bring in thousands more hits, too. Boolean expressions (named after the English mathematician George Boole) give you a way to control what you do and don’t find in your hit lists. Boolean logic consists of three logical operators:

* OR
* AND
* NOT

Interested in searching for edible shellfish? Consider these two expressions:

**clams AND oysters**

**clams OR oysters**

 If you search for these two expressions, the first will find you documents that include **both** terms. If you used the second, you will get all the documents that include either clams or oysters, or both. The second will produce many more hits than the first. The AND relationship limits your hits. The OR relationship expands your number of hits. With most search engines, you can string together more than two terms. For example, you can search for:

**clams AND oysters AND scallops**

 This expression limits the search further than the original search. Any documents found in the first search that did not include scallops, will be gone from the list. Expressions with more than two values are easy to understand if they use all ANDS or all ORS, but what happens if they are mixed? Consider this one that you might use if you are looking for a seafood restaurant:

**Clams OR oysters AND restaurant**

 This expression is ambiguous unless you know exactly how the search engine will react to it. If you read it left to right, it will first locate documents with clams or oysters and then limit it only to those that also include restaurant. That is probably what you had in mind in conducting this search. But some search engines will read right to left. In that case it will find all documents that include both oysters and restaurant, but then add in all documents that also include the word clam, whether or not restaurant is part of it. A very different result! The usual way to handle this is to put parentheses around the terms to be evaluated together. This is the correct way to look for seafood restaurants:

**(clams OR oysters) AND restaurant**

Expressions can be more complex. This expression will lead you to shellfish restaurants in
Boston or Cambridge:

**(clams OR oysters) AND restaurant AND (Boston OR Cambridge)**

 The above expression will find shellfish restaurants in Cambridge, England, too. To limit this, we can add a NOT expression.

**(clams OR oysters) AND restaurant AND (Boston OR Cambridge) AND NOT England**

*\*\*\*The NOT term has been changed in the last few years and does not work in some search engines. In google, you will have to do an advanced search and type in the words that you want to exclude in the box that asks for them.*

For all practical purposes, NOTs are always used with multiple terms and combined with ANDs. It would be pointless to search for any of these:

**NOT books**

**Television OR NOT books**

 Both of the above expressions would return every document that didn’t have the word “book” in it. The second would add to the list all documents that had the word television in, too, even if the word “books” was present. There would be millions of hits, in either case, and it would be useless. NOT can be very useful in limiting out unwanted hits. The two expressions that follow will give you very different results:

**dungeons AND dragons**

**dungeons AND NOT dragons**

Three of the major search engines let you use Boolean expressions just as described here—Alta Vista’s Advanced Mode (<http://www.altavista.com> ), Excite (<http://www.excite.com> ), and HotBot (<http://www.hotbot.com/> ). Others let you use variants of Boolean logic. One variant you will sometimes see are symbolic equivalents of AND, OR, and NOT, respectively the symbols & (ampersand), (pipe or vertical line), and ! (exclamation point). There are rarely, if ever, required in place of the words. Another common variation is to use the symbol + (plus) to enforce the appearance of a word, rather than using AND between required words. The symbol – (minus or dash) is used as a NOT in this system.
There is no exact equivalent of OR in this simpler variation. Boolean logic searching is based on Boolean Algebra. If you are interested in learning more about the mathematical basis of Boolean Algebra, visit this site:

<http://mathworld.wolfram.com/BooleanAlgebra.html>

**How Do Search Engines Work?**

 Search Engines for the general web (like all those listed above) do not really search the World Wide Web directly. Each one searches a database of the full text of web pages selected from the billions of web pages out there residing on servers. When you search the web using a search engine, you are always searching a somewhat stale copy of the real web page. When you click on links provided in a search engine’s search results, you retrieve from the server the current version of the page.

If a web page is never linked to in any other page, search engine spiders cannot find it. The only way a brand new page – one that no other page has ever linked to – can get into a search engine is for its URL to be sent by some human to the search engine companies as a request that the new page be included. All search engine companies offer ways to do this.

After spiders find pages, they pass them onto another computer program for “indexing.” This program identifies the text, links, and other content in the page and stores it in the search engine database’s files so that the database can be searched by keyword and whatever more advanced approaches are offered, and the page will be found if your search matches its contents.

Some types of pages and links are excluded from most search engines by policy. Others are excluded because search engine spiders cannot access them. Pages that are excluded are referred to as the “Invisible Web” – what you don’t see in search engine results. The Invisible Web is estimated to be two to three or more times bigger than the visible web.