



RÁPIDA Y PERTINENTE BÚSQUEDA POR INTERNET MEDIANTE OPERADORES BOOLEANOS

Bayardo Villegas V.

*Departamento de Matemáticas, Facultad de Ciencias, Pontificia Universidad Javeriana, Bogotá,
Colombia*

e-mail: bayardo.villegas@javeriana.edu.co

RESUMEN

Internet es una excelente y moderna herramienta para buscar información acerca de un problema particular, es como tener acceso simultáneo a muchas bibliotecas del mundo. En cualquier investigación generalmente se consigue una gran cantidad de información y mucha de ella no está relacionada con el problema de interés; el uso adecuado de los operadores booleanos produce mejores resultados en la búsqueda de información. En este escrito, la teoría básica de proposiciones y conjuntos y sus aplicaciones a la búsqueda de información en internet se discute y se ilustra con varios ejemplos.

Palabras clave: búsqueda, conjunto, operador *booleano*, proposición, proposición funcional.

ABSTRACT

Internet is a great and modern tool to look for information about a particular problem, it is like to have simultaneous access to several libraries in the world. In any search one usually gets a huge of information and some of it is not related to the problem of interest; the proper use of boolean operators leads to better results of the search. In this paper, the basic theory of sets and propositions and its applications to the search in internet is discussed and illustrated with several examples.

Key words: Search, set, boolean operator, proposition, functional proposition.

INTRODUCCIÓN

La búsqueda de información por internet tiene el problema de que el volumen de tal información es a veces muy grande con mucha de ella no pertinente. Una manera de resolver esto es pedir al “buscador” que consiga información que contenga simultáneamente todas las palabras clave del tema de interés, lo que, por otro lado, restringe demasiado la búsqueda. Esto último es solucionable colocando en el *query* todos los “sinónimos” (en sentido propio o en el de análogos, derivados de una pala-

bra primitiva truncada, en la línea del tema, nombres de autores relacionados con el tema, etc.) de las palabras clave: una manera para hacer que los escritos obtenidos en una tal búsqueda sean los que incluyan al tiempo todas las palabras clave y los sinónimos de éstas es mediante “operadores *booleanos*”.

MARCO TEÓRICO

En esta sección se anota la mínima teoría matemática para trabajar con “operadores *booleanos*”.

Proposiciones

Una *proposición* es una frase de la cual puede decirse, inequívocamente, que es una sola de las siguientes cosas: verdadera, falsa (Lipschutz, 1970). Por ej., las frases “la Tierra gira alrededor del Sol”, “*E. coli* es un procariota”, “Roma es la capital de Francia” y “ $2 > 4$ ”, son proposiciones, las dos primeras verdaderas y las últimas falsas; “¿qué día es hoy?” no es proposición.

Ahora unas operaciones en proposiciones (Pinzón, 1973): dadas las proposiciones p, q ,

- La *disyunción* de p, q , con denotación “ $p \vee q$ ” leída “ p o q ”, es falsa sólo si ambas proposiciones son falsas.
- La *conjunción* de p, q , con denotación “ $p \wedge q$ ” leída “ p y q ”, es verdadera sólo si ambas proposiciones son verdaderas.
- La *negación* de p , con denotación “ $\neg p$ ” leída “no p ”, es verdadera sólo si p es falsa y viceversa.

Para ilustrar estas operaciones sean p, q, r y s denotaciones, respectivamente, de las proposiciones del anterior ejemplo: se tiene que $p \wedge q, p \vee q, p \vee r$ y $\neg r$ son verdaderas, $p \wedge r, r \vee s$, y $\neg p$ son falsas.

Los anteriores signos “ \wedge ”, “ \vee ” y “ \neg ” se denominan “operadores lógicos booleanos” (frecuentemente se omite “lógicos”).

Por otra parte, una *función proposicional* $p(x)$ es una frase que contiene una variable x que al sustituirla por objetos determinados tal frase se convierte en varias proposiciones correspondientes (Kneebone, 1963); si para cierto objeto \mathbf{O} la proposición resultante es verdadera, se dice que “ \mathbf{O} satisface $p(x)$ ”. Por ej., a la función proposicional “ x es un planeta que gira alrededor del Sol”,

el objeto (*la*) *tierra* la satisface.

Conjuntos

Si X es un conjunto y $p(x)$ una función proposicional satisfecha por cada x elemento de X , entonces se escribe $X = \{x \mid p(x)\}$ y es de resaltar que “ $x \in X$ equivale a x satisface $p(x)$ ”; por otra parte, el número de elementos de X se denota $\#(X)$.

Bien, ahora unas definiciones de algunas operaciones entre conjuntos (Kleiman y Kleiman, 1975): dados los conjuntos X y Y ,

- La *unión* de X y Y , con denotación “ $X \cup Y$ ” leída “ X unión Y ”, es el conjunto de los elementos que están en A o en B :

$$X \cup Y = \{x \mid x \in X \vee x \in Y\}$$

La *intersección* de X y Y , con denotación “ $X \cap Y$ ” leída “ X intersección Y ”, es el conjunto de los elementos que pertenecen simultáneamente a X y a Y :

$$X \cap Y = \{x \mid x \in X \wedge x \in Y\}$$

- La *diferencia* X menos Y , denotada “ $X - Y$ ”, es

$$X - Y = \{x \mid x \in X \wedge x \notin Y\}$$

Se tienen las siguientes propiedades (Chung 1978): $X \cap Y \subset X \cup Y, X - Y \subset X \cup Y, X \cap Y = Y \cap X, (X \cap Y) \cap Z = X \cap (Y \cap Z), X \cup Y = Y \cup X, (X \cup Y) \cup Z = X \cup (Y \cup Z), (X \cup Y) \cap Z \neq X \cup (Y \cap Z), (X \cap Y) \cup Z \neq X \cap (Y \cup Z), X - Y \neq Y - X, (X - Y) - Z \neq X - (Y - Z), (X \cup Y) - Z \neq X \cup (Y - Z)$.

BÚSQUEDA BOOLEANA

Los operadores *booleanos* “ \wedge ”, “ \vee ” y “ \neg ” en muchos buscadores se remplazan, respectivamente, por “AND”, “OR” y “NOT” (en algunos buscadores en minúsculas).



Se ilustrará la “búsqueda *booleana*” (es decir, búsqueda usando operadores *booleanos*) por internet usando un buscador, como, por ej., <http://www.ncbi.nlm.nih.gov/pubmed>, mediante los siguientes ejemplos:

Ejemplo 1. Supóngase que se está interesado en buscar información sobre el tema enzimas “topoisomerasas” (se usarán palabras en inglés porque en este idioma hay mayor número de artículos publicados) y entonces habrá que hallar el conjunto

$$A = \{x / x \text{ es un escrito que contiene la palabra topoisomerase}\};$$

si se le da al computador el *query* topoisomerase, el buscador averiguará todos los x que satisfagan la proposición funcional $p(x)$ “ x es un escrito que contiene la palabra topoisomerase”, es decir, ¡por la equivalencia antes resaltada!, todos los x que pertenecen a A , o sea A , resultando que $\#(A) = 7558$ (a fecha 04 – 06 – 03).

Ejemplo 2. Sea $B = \{x / x \text{ es un escrito que contiene la palabra gyrase}\}$; algunos autores usan gyrase por topoisomerase y entonces la búsqueda anterior, no restringida arbitrariamente, es hallar el conjunto

$$\{x / x \text{ es un escrito que tiene una u otra de las palabras topoisomerase o gyrase}\} = A \cup B;$$

query: topoisomerase OR gyrase, $\#(A \cup B) = 8826$ (obsérvese en este ejemplo que las palabras clave ligadas por “OR” no necesariamente han de ser sinónimas y entonces podría también buscarse información, al tiempo, sobre temas diferentes).

Ejemplo 3. Sea $C = \{x / x \text{ un escrito que tiene las palabras sucesivas conserved sequence\}$: puede ser que sólo se esté interesado en escritos que aludan a secuencias conservadas de los enzimas anteriores y entonces el conjunto a buscar es

$$\{x \mid x \text{ es un escrito que contiene una u otra de las palabras topoisomerase o gyrase y simultáneamente las palabras sucesivas conserved sequence\} = (A \cup B) \cap C;$$

query: (topoisomerase OR gyrase) AND “conserved sequence”, $\#((A \cup B) \cap C) = 70$.

Obsérvese en este ejemplo que en el *query* se ha escrito “conserved sequence”: palabras entre comillas el buscador las procesa como frase; si se escribe conserved sequence, así sin comillas, el buscador asume conserved AND sequence y hallará escritos en los cuales estén ambas palabras así sea muy separadas y entonces la búsqueda será más amplia, tanto que el número de escritos encontrados bajo el *query* (topoisomerase OR gyrase) AND conserved sequence es 111; también nótese el uso adecuado de paréntesis en el *query* (recuérdese que $(X \cup Y) \cap Z \neq X \cup (Y \cap Z)$ pues en búsqueda *booleana* es muy importante el empleo correcto de las propiedades, antes anotadas, de las operaciones entre conjuntos).

Ejemplo 4. Sea $D = \{x \mid x \text{ es un escrito que contiene alguna palabra derivada de bacteri*}\}$: supóngase ahora que la anterior búsqueda se quiere más específica en el sentido de que excluya artículos que aludan a bacterias y entonces el conjunto a buscar es

$$(A \cup B) \cap C - D;$$

query: (topoisomerase OR gyrase) AND “conserved sequence” NOT bacteri*, $\#((A \cup B) \cap C - D) = 54$ (¡54! artículos frente a los ¡8826! resultantes en el ejemplo 2).

BONDADES DE LA BÚSQUEDA BOOLEANA

Pueden anotarse las siguientes:

- Diferentes autores pueden referirse al mismo tema bajo palabras clave

sinónimas: el operador “OR” resuelve este hecho y así la búsqueda no resulta restringida arbitrariamente.

- Puede buscarse información al tiempo sobre temas diferentes.
- Supónganse un investigador que necesita información sobre un tema que no es el de su especialidad, o un estudiante también con una tal necesidad: ya que mucho de las especialidades es sus jergas propias, puede suceder que en el *query* se introduzcan palabras mal escritas o aún que nada tengan que ver con tal tema y sin embargo, no habrá problema si esas palabras van ligadas por “OR” a otra(s) pertinente(s) bien escrita(s). Por otra parte, una búsqueda, por desconocimiento de sinónimos de palabras clave, inicialmente puede resultar restringida: la búsqueda booleana permite la incorporación de correcciones y sinónimos aprendidos en sucesivas búsquedas; un especialista obviaría tal incorporación y entonces la búsqueda *booleana* le agilizaría hallar escritos sobre temas de su interés.
- Permite al no especialista acceder a información especializada.
- El operador OR hace la búsqueda lo más amplia posible, mientras que el NOT, el AND y las comillas filtran la anterior a la pertinente (recuérdese que $X \cap Y$ y $X-Y$ son subconjuntos de $X \cup Y$ y que palabras entre comillas son pro-

cesadas como frase); es más, el uso adecuado e iterado de tales filtros y de paréntesis permite que la búsqueda sea lo “ampliamente específica” que se quiera: en este sentido una mejor búsqueda que la del ejemplo 4 sería con el *query* (topoisomerase OR gyrase) AND (“conserved sequence” OR “conserved region” OR “conserved motif” OR “conserved domain”) NOT bacteri*, para el cual el cardinal del conjunto de escritos encontrados es 66 (a fecha 04 - 06 - 03).

LITERATURA CITADA

- CHUNG K.L. *ELEMENTARY PROBABILITY THEORY WITH STOCHASTIC PROCESSES*. SPRINGER-VERLAG NEW YORK INC., NEW YORK, USA. 1978
- KLEIMAN A., KLEIMAN E. *CONJUNTOS. APLICACIONES MATEMÁTICAS A LA ADMINISTRACIÓN*. EDITORIAL LIMUSA, S. A. MÉXICO, MÉXICO D. F., MÉXICO. 1975.
- KNEEBONE G.T. *MATHEMATICAL LOGIC AND THE FOUNDATIONS OF MATHEMATICS*. D. VAN NOSTRAND COMPANY LIMITED, NEW YORK, USA. 1963.
- LIPSCHUTZ S. *TEORÍA DE CONJUNTOS Y TEMAS AFINES*. MCGRAW-HILL DE MÉXICO S. A., MÉXICO D. F., MÉXICO. 1970.
- PINZÓN A. 1973. *CONJUNTOS Y ESTRUCTURAS*. HARLA, S. A. MÉXICO, MÉXICO D. F., MÉXICO.

[HTTP://WWW.NCBI.NLM.NIH.GOV/PUBMED](http://www.ncbi.nlm.nih.gov/pubmed)

Recibido: 26-05-2003
Aceptado: 28-08-2003