

VI SEMINARIO-IIE-IMP-ININ
SOBRE ESPECIALIDADES TECNOLOGICAS

Julio 15 de 1992.

5556

EL ALGEBRA BOOLEANA EN LA RECUPERACION DE INFORMACION

PROFR. RAFAEL OVIEDO GALINDO.

I.M.P.

RESUMEN

Muy lejos están las complejas bases de datos actuales de las mucho más sencillas de hace treinta años, cuando sus índices se reducían a uno invertido, y los operadores lógicos para manejarlo se limitaban al AND y al OR.

Las bases de datos de hoy día, con sus índices básicos y adicionales, con su lenguaje controlado y operadores que aparentemente van más allá de la lógica de Bool, imponen la necesidad de hacer una revisión de la teoría básica y fundamental que está detrás de todo este desarrollo.

En este trabajo pretendo demostrar que los operadores lógicos AND, OR y NOT siguen siendo tan poderosos en el manejo de las bases de datos como lo fueron en un principio, y que aunque enmascarados en aparentes complejidades, siguen apuntalando hoy día, la brillante evolución de la recuperación automatizada y en línea.

1. INTRODUCCION.

1.1 Objetivo de las bases de datos bibliográficas:

- Almacenar información.
- Recuperar información.

1.2 Medios para alcanzar el objetivo:

- Indices inversos.
- Los Indices Duales Coordinados Inversos incorporan la lógica booleana.

1.3 Ejemplo:

ALGEBRA					
003	009	<u>014</u>	034	055	070
<u>071</u>	102	117	<u>304</u>	406	500

- Fig. 1 -

BOOL					
007	<u>014</u>	015	024	<u>071</u>	099
101	<u>304</u>				

- Fig. 2 -

ALGEBRA AND BOOL

3 documentos

ALGEBRA OR BOOL

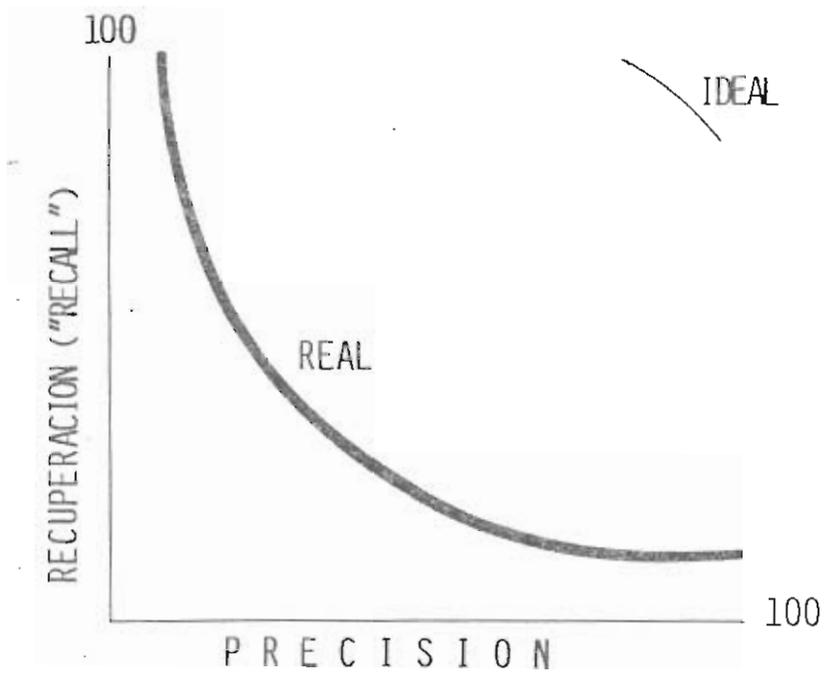
20 documentos

ALGEBRA NOT BOOL

9 documentos

2. RECUPERACION VS PRECISION.

Los resultados demostraron que la lógica booleana es MUY PODEROSA pero IMPERFECTA para recuperar la información.



- Fig. 3 -

OR VS AND

RECUPERACION

PRECISION

ALTA

BAJA



OR

AND



BAJA

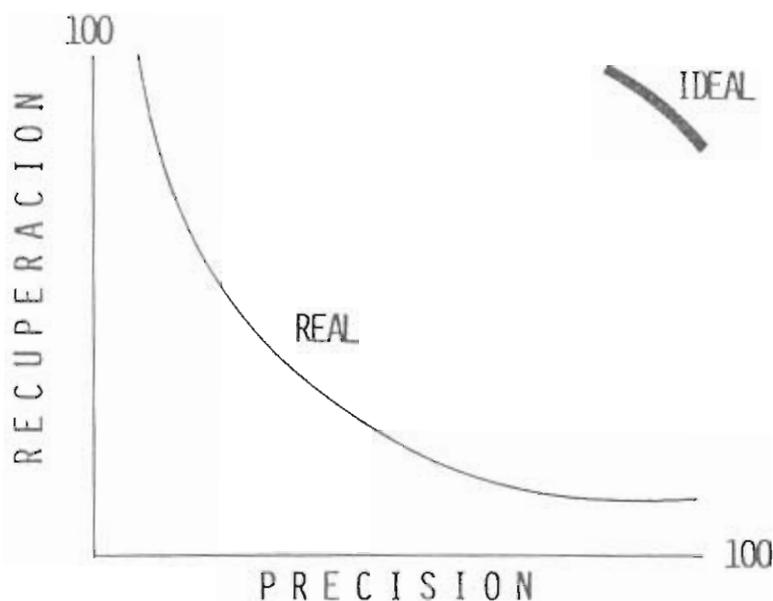
ALTA

- Fig. 4 -

COROLARIO: "Se ha argumentado que la lógica booleana idealmente no es adecuada para ser la base de la combinación de términos en la recuperación de información" (7.2).

3. MAS ALLA DE LA ALGEBRA BOOLEANA.

3.1 Desarrollos que pretenden acercar al ideal:



- Fig. 5 -

3.1.1 Jerarquización NATURAL del lenguaje.

Comprensión o extensión de los términos.

Ej. Herramienta - martillo

Las palabras de paro ("stop words") modifican la extensión.

3.1.2 Jerarquización TECNICA del lenguaje

Tesauros	Términos amplios
y	Términos reducidos
Descriptores:	Términos relacionados
	Use (por)

Identificadores

Encabezados de Sección

Códigos

3.1.3 Indices

básicos

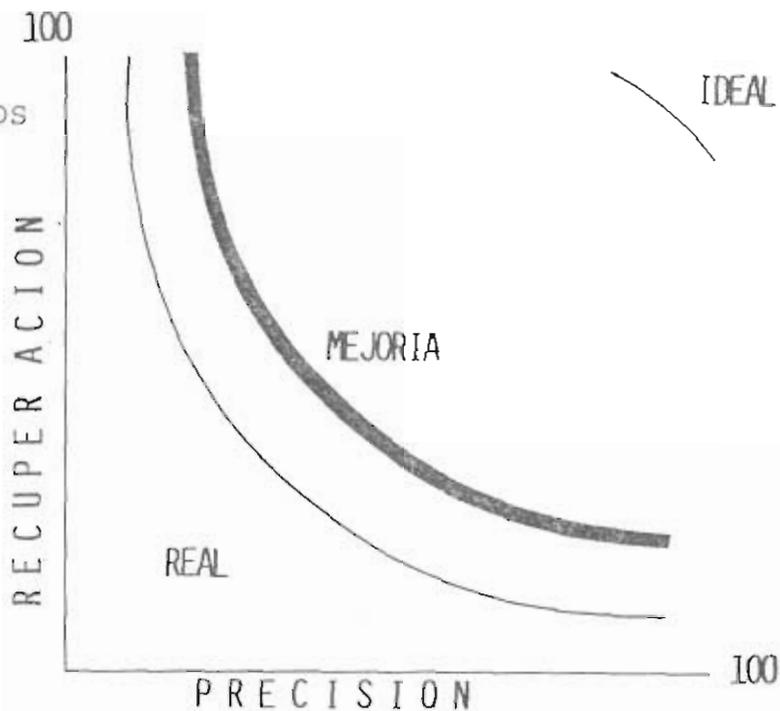
adicionales

3.1.4 Truncación

3.1.5 Proximidad

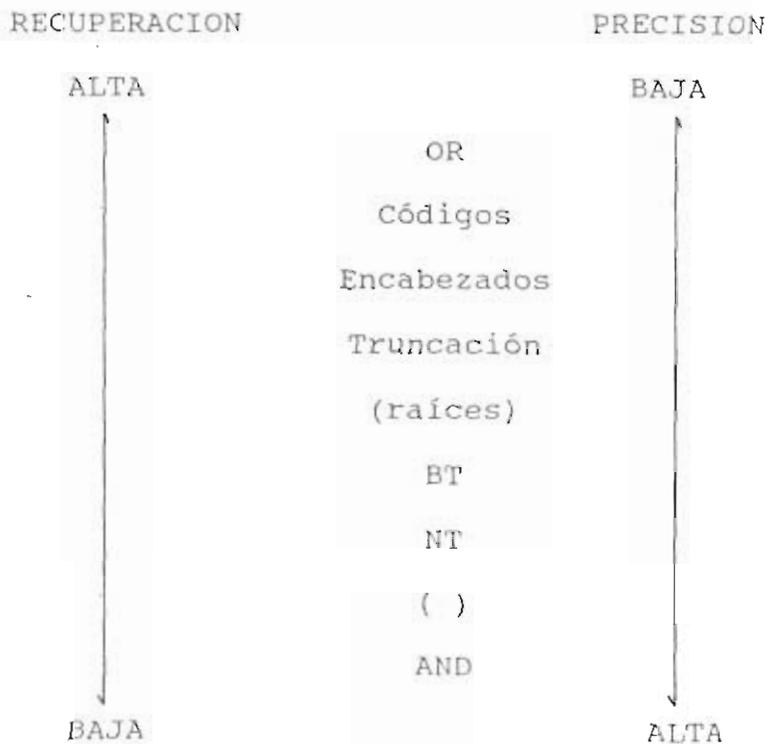
3.1.6 Búsqueda simultánea a través de archivos

3.2 RESULTADOS



- Fig. 6 -

3.2.1 EXHAUSTIVIDAD VS PRECISION



- Fig. 7 -

EN LENGUAJE LOGICO:

Códigos

Encabezados

} muchos OR

Truncación = álgebra OR algía OR algodón OR alborrismo...

BT = NT OR NT OR NT...

() = álgebra AND RESTRINGIDO booleana

NT = BT NOT (NT1 OR NT2...)

4. PODER DE LA LOGICA BOOLENA.

4.1 Funciona - no se ha encontrado mejor solución.

4.2 Sin complicaciones excesivas (sistemas expertos o inteligencia artificial) es lo que más se acerca al reconocimiento del lenguaje natural.

4.3 Indispensable para búsquedas a través de índices y campos.

4.4 Extremadamente útil y poderosa para búsquedas simultáneas a través de archivos.

5. EJEMPLOS

5.1 Pregunta: ADHESIVIDAD Y APLICACIONES DE RESINAS DENTALES.

Interés particular en resinas acrílicas o mixtas, autocurables o curables por luz ultravioleta, o laser; resinas acrílicas y mixtas para puentes y coronas; resinas acrílicas o mixtas protectoras de pulpa y dentina.

Conceptos:	A- resinas dentales	I- fotocurable
	B- adhesividad	J- puente
	C- acrílicas	K- coronas
	D- mixtas	L- resinas protectoras de pulpa
	E- autocurables	M- resinas protectoras de dentina
	F- luz ultravioleta	N- protección
	G- uv	O- pulpa
	H- laser	P- dentina

Estrategia

$A \cdot B \cdot (((C+D) \cdot (E+(I \cdot (F+G+H))))) + L+M+(N \cdot (O+P)))$

5.2 Pregunta: MODELOS MATEMATICOS PARA CATALIZADORES USADOS EN CONVERTIDORES CATALITICOS.

Estrategia (búsqueda simultánea a través de archivos):

```
(mathematical model../DE OR modeling, OR (matemactical  
( ) model...)) AND ((catalyst../ DE OR catalyst.. OR  
SH = catalyst-) AND (catalytic muffler/DE OR  
(catalytic ( ) muffler) OR (catalytic ( ) converter..)  
OR muffler.. OR SH = catalytic conversions))
```

5.3 Pregunta: Se solicita que la bibliografía corresponda a documentos escritos en inglés, francés y español.

Estrategia: Conj. n/english OR (conj. n AND (la = english OR la= french OR la = spanish)).

6. CONCLUSIONES

6.1 Aunque los operadores booleanos no son lo ideal para recuperar información, siguen siendo actualmente poderosos y aún no superados.

6.2 Son muy útiles para búsquedas a través de índices, así como para no quedar a merced de los indizadores.

6.3 Son muy efectivos para búsquedas simultáneas a través de archivos.

6.4 Son muy útiles para elaborar estrategias (inclusive "anidadas") que pueden ser enviadas con un solo "enter" para superar problemas de ruido.

7. BIBLIOGRAFIA.

7.1- Bates Marcia J., HOW TO USE CONTROLLED VOCABULARY MORE EFFECTIVELY IN ON LINE SEARCHING, Online, november 1988, pp. 45-56.

7.2- Bates Marcia J., THE FALLACY OF THE PERFECT THIRTY-ITEM ON LINE SEARCH.

7.3- Hearter Stephen P., HEURISTICS FOR ONLINE INFORMATION RETRIEVAL: A TYPOLOGY AND PRELIMINARY LISTING, Online Review, Vol. 9, No. 5, pp. 407-424, 1985.

7.4- Hildret Charles R., BEYON BOOLEAN: DESIGN THE NEXT GENERATION OF ONLINE CATALOGS, Library Trends, Spring 1987, pp. 647-667.