

Fundamentos de la Búsqueda

Uso eficiente de sistemas de información para profesionales de la Comunicación



Lluís Codina
Univ. Pompeu Fabra
Facultad de Documentación
Junio 2014

Contenido de esta presentación

2

- La Información como recurso
- La saturación informativa
- Búsquedas avanzadas
 - Booleanas
 - Parametrizadas
 - Combinadas
- Propuesta de ejercicios

La información como recurso

3

- Un recurso
 - Facilita alcanzar objetivos
- En **general**, la información permite
 - Predecir el futuro
 - ✦ Gestionar el entorno
 - ✦ Reducir las consecuencias de futuros hechos adversos
 - ✦ Incrementar las ventajas de futuros hechos favorables
 - Optimizar todos los demás recursos
 - ✦ A más información menos recursos (tiempo, dinero, etc.)
 - ✦ Conseguir mejores objetivos con menos recursos
- En **comunicación**, la información añade
 - Calidad
 - Credibilidad
 - Exclusividad

Entornos

4

- **Intensivos en información**
 - **Académicos e Investigadores**
 - **Estudiantes universitarios**
 - Doctorando y estudiantes de tercer ciclo
 - Profesores de universidad
 - Sectores I+D+i
 - etc.
 - **Profesionales**
 - **Comunicadores**
 - **Periodistas**
 - Economistas
 - Politólogos
 - etc.
- **Sistemas de información relevantes**
 - **Motores de búsqueda especializados**
 - Bases de datos para académicos
 - **Bases de datos para profesionales**

Sistemas relevantes para comunicadores y comunicólogos

5

- **Buscadores**
 - Google > búsqueda avanzada
 - Google Scholar
- **Bases de datos**
 - Prensa
 - ✦ MyNews
 - ✦ Factiva
 - Multimedia
 - ✦ AGE Fotostock
 - ✦ Getty Images
 - ✦ Corbis

El Problema

6

- Nuestro ecosistema informativo
 - Cada vez es más rico
 - ✦ Más información
 - ✦ Más fuentes
 - ✦ Más formatos
 - ✦ Más tecnologías y canales
- Pero el día sigue teniendo 24 horas
 - No podemos consumir toda la información que nos afecta profesionalmente
- Cada vez se requieren mejores competencias informacionales
 - “Somos” la información a la que nos exponemos

Consecuencias

7

- Es letal
 - Mucha información (ruido)
 - Perder información (silencio)
- Necesitamos
 - **Solamente** información relevante (precisión), pero
 - **Toda** la información relevante (exhaustividad)
- El balance necesario
 - La precisión reduce la exhaustividad
 - Hemos de elegir entre:
 - ✦ Aumentar precisión (y perder exhaustividad)
 - ✦ Aumentar la exhaustividad (y perder precisión)

Las soluciones

8

- Necesitamos
 - Conocer las fuentes
 - ✦ Fuentes críticas de cada sector de actividad o rama del conocimiento. En nuestro caso: fuentes para la comunicación social
 - Saber buscar
 - ✦ Conocer los principios unificadores
 - ✦ Utilizar las búsquedas avanzadas

Tipos de búsqueda

9

- Simple
 - Término único
 - Término compuesto
 - Frase en lenguaje natural
- Avanzada
 - Parametrizada
 - Booleana
 - Combinada
- Formas de expresión
 - Sintaxis directa
 - Formulario (asistida)

Búsqueda simple

10

- Término único
 - Ejemplos: **web**
 - Problema: **ruido**
 - Opción: no usar nunca términos únicos
- Término compuesto: conceptos que necesitan dos o más términos para ser expresados
 - Ejemplos: **John Ford, juegos olímpicos, economía política**
 - Problema: **ruido** en versión falsas coordinaciones. Ejemplo: el sitios web de coches de la marca Ford de un tal John Smith.
 - Opción: usar comillas, p.e. “john ford”, ”juegos olimpicos”, etc.
- Frase en lenguaje natural
 - ✦ Ejemplo: **la web móvil y el futuro de los medios de comunicación**
 - ✦ Útil para encontrar documentos MUY específicos, si es que existen...
 - ✦ Problema general: **silencio**, falsas coordinaciones
 - ✦ Opción: convertir la frase natural en una ecuación booleana
- En general
 - Uso de alfabeto pobre, p.e: “juegos olimpicos” (no “Juegos Olímpicos”), “economia politica”, etc.
 - La lengua de la pregunta determina la lengua del documento

Búsquedas avanzadas

11

- Tipología de búsquedas avanzadas
 - Parametrizadas
 - Booleanas
 - Combinadas

Parametrizadas

12

- Restricción de la búsqueda por campos o parámetros
- Los más habituales (en la web)
 - Título
 - Tema
 - Fecha de publicación
- Otros parámetros (en bases de datos)
 - Autor
 - Lengua
 - Fuente
 - Categoría, etc.
- Ejemplo
 - Título: “instituto noos”
 - Entre: 2009 - 2012

Ejercicio - 1

13

- Haga dos o tres prácticas de búsqueda parametrizada en Google mediante el formulario de búsqueda avanzada, combinando un tema a su elección y, en cada caso: (a) un formato de documento, (b) un campo determinado, (c) un tipo de licencia o una fecha de publicación.
- Documente la práctica con capturas de pantalla comentadas.

Booleanas

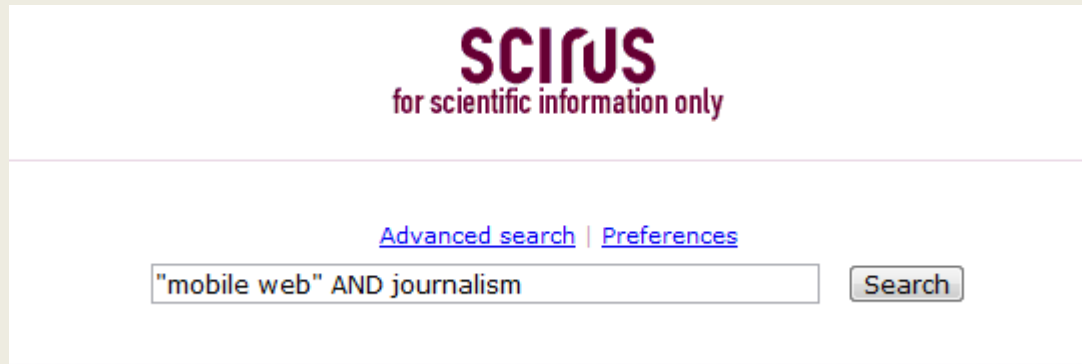
14

- Basada en álgebra de Boole
- Operadores principales
 - AND
 - OR
 - NOT
- Forma simple
 - **t1 OP t2**
 - Dónde:
 - ✦ **t1** es una palabra clave
 - ✦ **OP** es uno de los operadores principales AND, OR, NOT
 - Ejemplo: "**transparency index**" **AND** **spain**

Ejemplo

15

- Búsqueda booleana simple (caja de búsqueda de Scirus)



SCIRUS
for scientific information only

[Advanced search](#) | [Preferences](#)

Booleanas compuestas

16

- Forma compuesta. Variaciones:
- $t1 \text{ OP } (t2 \text{ OP } t3)$
- $(t1 \text{ OP } t2) \text{ OP } (t3 \text{ OP } t4)$
- Ejemplos
 - Situación: necesitamos encontrar información sobre comunicación científica y ética o deontología. Posible solución:

`"comunicacion científica" AND (etica OR deontologia)`
 - 2. Situación: necesitamos encontrar información sobre la web 2.0, las redes sociales, el periodismo, la comunicación corporativa, etc. , todo a la vez. Posible solución:

`("web 2.0" OR "redes sociales") AND (periodismo OR "comunicación corporativa")`

Ejemplo con Factiva

17

- Búsqueda booleana compuesta

Búsqueda de
texto libre

Modo de búsqueda

Ejemplos

educación AND (españa OR "unión europea")

Concept Explorer [Beta]

Incluir los medios de comunicación añadidos recientemente

Ejercicio 2

18

- Pruebe a realizar la búsqueda del ejemplo 1 (diapositiva 16) en Google con el formulario de búsqueda avanzada.
- Documente la práctica con capturas de pantalla del formulario de búsqueda y de la primera página de resultados de cada búsqueda.

Booleanas - 3

19

- **Uso**
 - **AND** > reduce resultados y aumenta precisión
 - **OR** > aumenta resultados y reduce precisión
 - **NOT** > reduce resultados y aumenta precisión
- **Balance:**
 - **AND** > aumenta precisión pero elimina información
 - **OR** > aumenta resultados pero incrementa el ruido
 - **NOT** > reduce el ruido pero elimina información

Problemas - 1

20

- **AND >**

- Falsamente intuitivo
- No corresponde necesariamente con el uso del “y” gramatical
- Ejemplo: “necesito información sobre la obras de teatro **y** novelas que han sido adaptadas para cine **y** televisión”
- Traducción booleana **incorrecta**:
 - ✦ **(teatro AND novela) AND (cine AND televisión)**
 - ✦ Problema: el usuario perderá información muy relevante
- Traducción booleana **correcta**:
 - ✦ **(teatro OR novela) AND (cine OR televisión)**
 - ✦ Ventaja: el usuario consigue una gran precisión sin pérdida de información

Problemas - 2

21

- **OR**
 - Anti intuitivo
 - ✦ término 1 o bien término 2, o ambos
 - ✦ cuantos más términos, más ruido, no menos
 - En la lengua natural, la partícula “O” puede tener dos significados:
 - ✦ Incluyente
 - ✦ Excluyente
 - Cómo operador booleano siempre es **incluyente**

Problemas 3

22

- **NOT**
 - Anti intuitivo
 - Puede destruir información: elimina por igual información relevante y no relevante
- Ejemplo: “necesito información sobre publicidad de refrescos, pero siempre que sean bebidas para adultos”
- Con la ecuación: (**refrescos AND publicidad**)
NOT jóvenes
- Perdemos este documento: “Resultados de una investigación sobre publicidad de marcas de refrescos para jóvenes y adultos”

Cuadro resumen

23

Operador	Resultados	Precisión	Ruido
AND	-	+	-
OR	+	-	+
NOT	-	+	-

Búsquedas Combinadas: toda la potencia en marcha

24

- Son las más habituales en sistemas profesionales y científicos
- Combinan ecuaciones de los dos tipos
 - Parametrizadas
 - Booleanas
- Ejemplo
 - (Título artículo= “web móvil” AND publicidad) AND (Años= 2010 -2014)
 - Interpretación: buscamos artículos sobre la web móvil y la publicidad, pero siempre que esas palabras clave aparezcan en el título, pero solamente si han publicado en los últimos años.

Búsquedas avanzadas: balance

25

- El medio más potente para expresar nuestras necesidades de información sin ambigüedad, en especial las búsquedas combinadas (booleanas + parametrizadas)
- Especialmente útil en bases de datos profesionales (p.e. MyNews o Factiva) y científicas.
- Utilidad algo limitada en Google:
 - ✦ Implementación incompleta en el formulario de búsqueda avanzada
 - ✦ Resultados pueden ser inconsistentes en la sintaxis directa

Ejemplo

26

- Formulario de búsqueda combinada en CSIC

The screenshot shows a search interface with three columns: 'Buscar:', 'en los campos:', and 'opciones:'. The 'Buscar:' column contains three rows of search terms: 'web móvil', 'periodismo', and 'radio'. The 'en los campos:' column contains three rows of 'Descriptores'. The 'opciones:' column contains three rows of 'todas las palabras'. A red vertical line is on the left of the search terms, and a red arrow points to the 'en los campos:' header. Below the search terms is a note: '(Nota: Campos básicos significa buscar en título, resumen, descriptores, identificadores y topónimos)'. Below the note is a 'Clasificación:' dropdown. Below that are 'Año publicación:' (después de 2010) and 'Siglos:' (igual a). Below those are 'Periodo Histórico:' (inicio: igual a, fin: igual a). At the bottom right are 'Buscar' and 'Limpiar' buttons.

Buscar:	en los campos:	opciones:
<input type="text" value="web móvil"/>	Descriptores	todas las palabras
<input type="text" value="periodismo"/>	Descriptores	todas las palabras
<input type="text" value="radio"/>	Descriptores	todas las palabras

(Nota: Campos básicos significa buscar en título, resumen, descriptores, identificadores y topónimos).

Clasificación:

Año publicación: después de Siglos: igual a

Periodo Histórico: inicio: igual a fin: igual a

Ejercicio - 3

27

- Dada la siguiente necesidad de información: “becas y ayudas para producciones de cine y vídeo”, ¿cuál de las siguientes ecuaciones es mejor?
 - a) (becas AND ayudas) AND (cine AND vídeo)
 - b) (becas OR ayudas) AND (cine OR vídeo)
- Copie el enunciado de esta pregunta y la respuesta elegida en su informe.

Ejercicio - 4

28

- Prepare una búsqueda combinada de un tema de su interés que sea:
 - Parametrizada, por ejemplo, mediante un tema limitado al título
 - Booleana cruzando dos conceptos
- Primero escriba la ecuación booleana, después vea si es posible representarla en el formulario de búsqueda avanzada de Google. Si le ha sido posible, haga las capturas correspondientes.
- En caso contrario, indique el motivo por el cual el formulario no le ha sido útil. En este caso, intente hacer la búsqueda mediante sintaxis libre. Valore los resultados. Haga capturas de pantalla.

Ejercicio 5

29

- Repita la búsqueda del Ejercicio 4 (o piense en un tema de búsqueda nuevo, si lo prefiere) pero use en esta ocasión la versión académica de Google > Google Scholar (o Google Académico).
- Compare el tipo de resultados de Google “normal” con el tipo de resultados de Google Scholar mediante capturas de pantalla comentadas.

Ejercicios finales de comprobación (EFC) - 1

30

- Dada la siguiente pregunta: “necesito información sobre mascotas”, ¿cuál es la mejor ecuación?:
 - a) `perros AND gatos`
 - b) `perros OR gatos`

Ejercicios finales de comprobación - 2

31

- ¿Podría proponer un enunciado plausible en lenguaje natural de esta ecuación booleana?:
 - (legislación OR directivas) AND (niños OR adolescentes) AND (prensa OR noticias)

Ejercicio finales de comprobación - 3

32

- ¿Podría proponer una formalización posible como ecuación booleana del siguiente problema de información?:
 - “el futuro de la comunicación y del periodismo en la era de la web y de los dispositivos móviles como tablets y smartphones”