

# Fundamentos de la Búsqueda en la Web Para Periodistas y Comunicadores



Lluís Codina  
Grupo DigiDoc – Departamento de  
Comunicación  
Universitat Pompeu Fabra

# Contenido de esta presentación

2

- La Información como recurso
- Entornos intensivos en información
- La saturación informativa
- Búsquedas avanzadas
  - Booleanas
  - Parametrizadas
  - Combinadas
- El monocultivo informacional
- 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

# La Búsqueda de Información en el Periodismos

4

- En el **Periodismo** en particular, y en la **Comunicación** en general, la búsqueda y el uso de sistemas de información añade:
  - Calidad
  - Credibilidad
  - Exclusividad

# Entornos

5

- 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

6

- **Buscadores**
  - **Generalistas**
    - ✦ **Google**
    - ✦ **Bing**
    - ✦ **DuckDuckGo**
  - **Especializados:**
    - ✦ Sistemas de respuestas: **Wolfram Alpha**
    - ✦ Buscadores académicos: **Google Scholar**
- **Bases de datos**
  - **Prensa**
    - ✦ MyNews
    - ✦ Factiva
  - **Multimedia**
    - ✦ AGE Fotostock
    - ✦ Getty Images
    - ✦ Corbis

# El Problema

7

- 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

8

- 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)



# Soluciones – 1: Abandonar el Monocultivo

9

- No se trata de cambiar de buscador
  - Si el usamos funciona bien, ¿porqué cambiarlo?
- No se trata de usar una vez un buscador y otra vez, otro
  - No solucionamos nada si una vez usamos el buscador A, y otras el buscador B
- Se trata de **añadir** un n-ésimo buscador a nuestra cartera de herramientas profesionales
  - El buscador A en búsquedas puntuales
  - El buscador A + B en búsquedas informacionales
  - Y, posiblemente, los buscadores A + B + C en determinados proyectos

# Soluciones 2 - Conocer las fuentes

10

- 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

11

- **Simple**
  - Término único
  - Término compuesto
  - Frase en lenguaje natural
- **Avanzada**
  - Parametrizada
  - Booleana
  - Combinada
- **Formas de expresión**
  - Pseudo código: independiente del sistema
  - Sintaxis directa: dependiente del sistema
  - Formulario: interfaz del sistema

# Búsqueda simple - 1

12

- Término único
  - Ejemplos: **television**
  - Problema: excesivo **ruido**
  - Opción: no usar nunca términos únicos
- Término compuesto: conceptos que necesitan dos o más términos para ser expresados
  - Ejemplos: **John Ford, ciencia ficción, 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”, ”ciencia ficción”, etc.

# Búsqueda simple - 2

13

- Frase en lenguaje natural
  - Ejemplo: **principales tendencias del periodismo en 2015**
  - Útil para encontrar documentos MUY específicos y seguramente MUY relevantes (si es que existen...)
  - **Cada vez más útil** en Google:
    - Capacidad de interpretar las intenciones del usuario y la semántica de la frase completa
  - Problema general: **silencio**, falsas coordinaciones. Pérdida de información. Falta de precisión de los resultados.
  - Opción: convertir la frase natural en una ecuación booleana.

# Recomendaciones generales

14

- Uso de comillas en términos compuestos
  - “juegos olimpicos” mejor que Juegos Olímpicos
- Uso de alfabeto pobre:
  - “juegos olimpicos” mejor que **J**uegos **O**límpicos
  - “ciencia ficcion” mejor que: ciencia **f**icción
- La lengua de la pregunta (pre)determina la lengua del documento

# Búsquedas avanzadas

15

- Tipología de búsquedas avanzadas
  - Parametrizadas
  - Booleanas
  - Combinadas

# Parametrizadas

16

- Restricción de la búsqueda por campos o parámetros
- Los campos o parámetros más habituales (en la web)
  - Título
  - Cuerpo de la página
  - Fecha de publicación
- Otros parámetros (en bases de datos)
  - Autor
  - Lengua
  - Fuente
  - Categoría, etc.
- Ejemplo
  - Título: festival cannes
  - Entre: 2012 - 2014



# Ejercicio - 1

17

- Haga **dos 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) una zona del documento (p.e. título).
- Documente la práctica con capturas de pantalla comentadas del formulario de búsqueda y de la página de resultados correspondiente.

# Booleanas

18

- Basada en álgebra de Boole
- Operadores principales
  - AND
  - OR
  - NOT
- Forma simple
  - $t1 \text{ OP } t2$
  - Dónde:
    - ✦  $t1$  es una palabra clave (p.e. barcelona, festivales, cine, televisión, etc.)
    - ✦  $\text{OP}$  es uno de los operadores principales AND, OR, NOT
  - Ejemplos en pseudo código:
    - ✦ barcelona AND festivales
    - ✦ Cine OR television
  - Ejemplo (siguiendo el formulario de búsqueda avanzada de Google)
    - ✦ **AND > Todas las palabras:** barcelona festivales
    - ✦ **OR > Cualquiera de las palabras:** cine television

# Booleanas compuestas

19

- Forma compuesta. Variaciones:
- **Primer caso:**  $t1 \text{ OP } (t2 \text{ OP } t3)$
- **Segundo caso:**  $(t1 \text{ OP } t2) \text{ OP } (t3 \text{ OP } t4)$
- Ejemplos
  - **Primer caso:** necesitamos encontrar información sobre “ciencia ficción” y cine o televisión. Posible solución en pseudo código:

`"ciencia ficción" AND (cine OR televisión)`

- ✦ Tiene solución en el formulario de Google . ¿Cuál sería? Haga pruebas.

- **Segundo caso:** necesitamos encontrar información sobre la web 2.0, las redes sociales, el periodismo, la comunicación corporativa, etc. , todo a la vez. Posible solución en pseudo código:

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

- ✦ ¿Tiene solución en el formulario de Google?
- ✦ Si es que NO, en este caso necesitamos **sintaxis directa**

# Ejercicio 2

20

- Pruebe a realizar las búsquedas siguientes expresadas con pseudo código:  
Barcelona AND (cine OR television)  
barcelona AND festivales AND (cine OR television OR audiovisual)  
con el **formulario** de búsqueda avanzada de Google
- Documente la práctica con capturas de pantalla del formulario de búsqueda y de la primera página de resultados de cada búsqueda.

# Repaso Booleanas - 3

21

- **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

22

- **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

23

- **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

24

- **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”



# Ejercicio - 3

25

- 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.

# Cuadro resumen

26

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

# Búsquedas Combinadas: la mejor amiga de los Periodistas

27

- Son las más habituales en sistemas profesionales y académicos
- Combinan ecuaciones de los dos tipos
  - Parametrizadas
  - Booleanas
- Ejemplo
  - (Título artículo= “web móvil” AND publicidad) AND (Años= 2010 -2015)
  - Interpretación: buscamos artículos sobre la web móvil y la publicidad, pero siempre que esas palabras clave aparezcan en el título y solamente si han publicado en los últimos años.

# Búsquedas avanzadas: balance

28

- 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: **Factiva**, **Getty Images**, **Corbis**, etc. y académicas: **Google Scholar**.
- Sin embargo, utilidad **limitada en Google “normal”**:
  - ✦ Implementación incompleta en el formulario de búsqueda avanzada
  - ✦ Resultados pueden ser inconsistentes en la sintaxis directa
- En el caso de Google “normal”: cada vez más eficaz (para compensar) en búsquedas con lenguaje natural.

# Ejemplo

29

- Formulario de búsqueda combinada en CSIC

The screenshot shows a search interface with the following components:

- Buscar:** Three rows of search criteria. The first row has a dropdown menu (pointed to by a red vertical arrow) and the text "web móvil". The second row has a dropdown menu with "Y" and the text "periodismo". The third row has a dropdown menu with "O" and the text "radio".
- en los campos:** Three dropdown menus, all set to "Descriptores" (pointed to by a red arrow).
- opciones:** Three dropdown menus, all set to "todas las palabras".
- Clasificación:** A dropdown menu.
- Año publicación:** A dropdown menu set to "después de" followed by a text box containing "2010" (pointed to by a red arrow).
- Siglos:** A dropdown menu set to "igual a" followed by a text box.
- Periodo Histórico:** Two dropdown menus set to "igual a" followed by text boxes for "inicio" and "fin".
- Buttons:** "Buscar" and "Limpiar" buttons at the bottom right.

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

# Ejercicio - 4

30

- Prepare una búsqueda combinada de un tema de su interés que sea:
  - **Parametrizada**, por ejemplo, mediante un tema limitado al **título**
  - **Booleana**, en base a utilizar dos conceptos combinados mediante OR, AND, NOT (no hace falta todos los operadores, uno o dos de ellos).
- Valore los resultados. Haga capturas de pantalla.

# Evitar el monocultivo

31

- Usar “otro” buscador (además de Google)
- Buscadores generalistas:
  - Bing
  - DuckDuckGo
  - Ask
- Buscadores especializados
  - Scholar > Información académica
  - Mendeley > Información académica + Red Social
  - Wolfram Alpha > Sistema de respuestas

# Ejercicio

32

- Ir a: <http://tools.seochat.com/>
  - Hacer una breve revisión en diagonal de las herramientas disponibles. Hacer una captura de pantalla comentada
- Ir a: <http://tools.seochat.com/tools/google-vs-yahoo-search-results/>
  - Hacer una comparación Google v Bing para un tema de su interés (puede ser un tema de actualidad o un tema académico vinculado con el Grado en Periodismo)
  - Hacer una captura de pantalla comentada



# Ejercicio

33

- Realizar la misma búsqueda (a elección del estudiante) en estos tres buscadores:
  - Google
  - Bing
  - DuckDuckGo
- Hacer una breve comparativa en formato libre (p.e. mediante una tabla y/o con capturas de pantalla) de la relevancia de los cinco primeros resultados de cada buscador.
- Redactar unas breves conclusiones.

# Google Scholar

34

- La versión Google para entornos académicos
- Únicamente resultados publicados en servidores muy fiables (universidades, fundaciones científicas, centros de investigación, etc.)
- Cada vez más utilidades:
  - Análisis de citas
  - Biblioteca personalizada
  - Exportación de referencias

# Ejercicio 5

35

- Repita alguna de las búsquedas anteriores (o piense en un tema de búsqueda nuevo, si lo prefiere) pero use en esta ocasión la versión académica de Google > Google Scholar.
- Compare las **páginas de resultados**. Observe las opciones de navegación laterales y las opciones situadas debajo de cada resultado. Intente comprender su función. Indíquelas en el pie de la captura.

# Sistema de respuestas

36

- No proporcionan listas de enlaces, sino respuestas directes
- Útiles para resolver necesidades de información concretas
- No resuelven búsquedas informacionales ni transaccionales

# Wolfram Alpha

37

- Ir a: Wolfram Alpha > <http://www.wolframalpha.com/>
- Vaya primero a los ejemplos de uso:  
<http://www.wolframalpha.com/examples/>
  - Haga pruebas libremente, ponga a prueba, p.e., algunos de los ejemplos de búsqueda sugeridos para entender mejor este buscador
  - Haga dos capturas de pantalla comentadas de dos elementos significativos de Wolfram Alpha

# Para evitar el monocultivo – 1: Tipología de necesidades de información

38

- **Necesidades de Información Concretas - NIC:**
  - Navegacionales, transaccionales, informacionales simples
    - ✦ Una sola respuesta o un dato concreto soluciona la necesidad
    - ✦ Ejemplos:
      - ¿Cuántos presidentes han tenido en los Estados Unidos de América?
      - ¿Cuáles son las cinco ciudades más pobladas del mundo?
      - Unesco (en el sentido, “quiero ir al sitio de la Unesco”)
- **Necesidades de Información Orientadas a un Problema –NIOP**
  - Informacionales complejas, académicas
    - ✦ No hay una respuesta unívoca, necesitamos razonamiento (discurso) y reflexión; más información suele generar aún más necesidades de información. No hay una solución predeterminada.
    - ✦ Ejemplos:
      - Realidad aumentada y periodismo
      - Modelos de negocio en cibermedios
      - Principales tendencias de comunicación en los próximos años

# Para evitar el monocultivo – 2: Buscadores candidatos

39

- **Sea:**
  - **A:** el buscador habitual, probablemente, Google
  - **B:** un segundo buscador alternativo, probablemente, Bing
  - **C:** un tercer buscador alternativo, posiblemente, DuckDuckGo o Ask o Yahoo
  - **D:** una pareja de buscadores académicos, probablemente, Google Scholar + Mendeley o Wordlcat

# Para evitar el monocultivo – 3: Estrategia

40

- Ejemplo de aplicación de la regla de **añadir** (no sustituir) buscadores al buscador habitual
  - Supongamos esta distribución de necesidades de información:
    - ✦ 7 de cada 10 serán tipo **NIC**
    - ✦ 2 de cada 10 serán tipo **NIOP**-Profesional
    - ✦ 1 de cada 10 será tipo **NIOP**-Académica
  - La posible estrategia **multicultivo** puede ser:
    - ✦ 70 % de las búsquedas, tipo **NIC**: buscador A
    - ✦ 20 % de las búsquedas, tipo **NIOP**-profesionales: buscador A + B | C
    - ✦ 10 % de las búsquedas, tipo **NIOP**-académicas: D



# Más información y sitio del autor

41

- <http://www.lluiscodina.com/category/diagramas/>
- <http://www.lluiscodina.com/>