Índice

Problemas: demasiados resultados

Debemos ser más específicos

- Añadir palabras clave para dar más información

  `ropa vieja` vs. `receta ropa vieja`

- Emplear palabras más específicas

  `coche de segunda mano` vs. `Peugeot 207 de segunda mano`

- Retringir la búsqueda mediante AND
- Excluir mediante NOT palabras poco útiles o que no se correspondan con la acepción deseada

  `jaguar -coche`

- Buscar FRASES EXACTAS
- Primar las palabras más relevantes:
  - Poniéndolas de primeras
  - Repetiéndolas
- Reintroducir mayúsculas y/o tildes

Problemas (cont.): pocos o ningún resultado

Debemos ser más generales

- Eliminar palabras clave dejando las más relevantes
- Ampliar la búsqueda mediante OR
- Emplear sinónimos o variantes
- Comprobar ortografía (comprobar sugerencias)
- Eliminar mayúsculas y tildes
- Buscar en inglés o viceversa (localización)
- Emplear otro buscador

Problemas (cont.): demasiado lento

- **Búsqueda demasiado compleja**
  - Eliminar palabras "demasiado" comunes (*stopwords*): artículos, preposiciones, etc.
  - Evitar consultas demasiado grandes
  - Emplear otro buscador
- **Sobrecarga en la red o en el buscador**
  - Buscar en otro momento
  - Usar buscador más cercano (geográficamente)
  - Desactivar descarga de gráficos
  - Emplear otro buscador

Consejos y sugerencias

- Ni todo está en Internet, ni todo es correcto
- Piensa en cómo puede estar escrito en la página lo que buscas
- No limitarse a un único buscador
  - Alternativa: metabuscadores
- Utilizar los operadores selectivamente y con cuidado
- Evitar palabras demasiado generales, emplear términos específicos

$$\texttt{pintura} \text{ vs. } \texttt{pintura rupestre}$$

- Evitar palabras ambiguas:

$$\texttt{reparaciones de pisos} \rightarrow \begin{cases} \texttt{viviendas ?} \\ \texttt{suelos ?} \end{cases}$$

- Emplear sinónimos y palabras relacionadas

$$\texttt{simio} \rightarrow \texttt{mono}, \texttt{chimpancé}, \texttt{gorila}, \text{ etc.}$$

Consejos y sugerencias (cont.)

- El orden puede influir

  `empanada carne` vs. `carne empanada`

  - Palabras más relevantes primero

    `formatear disco linux` vs. `linux formatear disco`

- Pistas o enlaces útiles en páginas poco relevantes a primera vista
- Restringir la búsqueda a los resultados anteriores*
- Limitar el dominio de búsqueda
  - Ej. sólo en `udc.es`
- Limitar los campos de la página en los que buscar
  - Ej. sólo en el título
- Buscar en el buscador *localizado*:

  google.es $\neq$ google.com

Consejos y sugerencias (cont.)

- Examinar las "búsquedas relacionadas" y el autocompletado que te sugiere el buscador
- Búsquedas especializadas:
  - Emplear motores o herramientas especializadas
  - Preguntar en foros especializados, etc.
- **Suele haber mucha más información en inglés**
- Leer la ayuda del buscador

Ejercicios

- Lanza en Google las siguientes consultas:
  - `"el monte everest mide"`
  - `F1 OR "fórmula 1"`
  - `normativa do galego`
  - `normativa do galego` (restringida al dominio `xunta.es`)
  - `"normativa do galego"` (restringida al dominio `xunta.es`)
  - `computación cuántica`
  - `quantum computing`

Operadores avanzados: comodín *

- **Varía mucho entre buscadores y con el tiempo**
- **Altavista (2000)**: serie de caracteres dentro de una cadena
  - Representar variantes

    `libr*`={`libro`, `libros`, `librería`, etc.}
  - Ampliar la búsqueda
- **Google/Yahoo! (2009)**: exactamente 1 palabra completa
  - Conjuntamente con operadores de FRASE

`"chose * flights"`={`"chose several flights"`, `"chose his flights"`, etc.}

`"chose * * flights"`={`"chose several intercontinental flights"`, `"chose among his flights"`, etc.}
  - Buscar citas aproximadas
- **Google (2011)**: 1 o más palabras completas
  - Conjuntamente con operadores de FRASE

`"chose * flights"`={`"chose several flights"`, {`"chose several intercontinental flights"`, etc.}

`"chose * * flights"`={`"chose several intercontinental flights"`, `"chose among several intercontinental flights"`, etc.}
  - Buscar citas aproximadas

Operadores avanzados (cont.): proximidad

- Exige que la otra palabra esté dentro de un radio dado
- Anteriormente se explicitaba mediante operadores como NEAR o ADJ
- Actualmente implícito: la proximidad entre sí de los términos de la consulta dentro del documento aumenta la relevancia

Operadores avanzados (cont.): restringir dominio de búsqueda

- Buscar sólo dentro de un (sub)dominio (ej., udc.es)
- **Google**/**Yahoo**/**Bing**: `site:`

  `consejo de gobierno site:udc.es`

  - Puede combinarse con NOT para

    `consejo de gobierno -site:udc.es`

Operadores avanzados (cont.): restringir campo de búsqueda

- Buscar sólo dentro del texto de un campo determinado de la página (título, texto, etc.)

- **Google/Yahoo!/Bing**:
  - `intitle:` : busca dentro del título de la página
  - `intext:` : busca en el texto de la página
  - `inanchor:` : busca en el texto descriptivo asociado a un enlace
  - `inurl:` (sólo en Google) : busca en la URL de la página

`intitle:ternera`

Operadores avanzados (cont.): paréntesis y anidamientos

- Agrupar expresiones de búsqueda complejas
  - **Altavista:** `(crema AND cacahuete) AND (gelatina OR mermelada)`

$$\rightarrow \begin{cases} \texttt{crema de cacahuete y gelatina} \\ \texttt{crema de cacahuete y mermelada} \\ \ldots \end{cases}$$

- Uso complejo y problemático

Feedback: [Páginas similares]/[Similar pages]

- Buscar páginas similares a la indicada (*relevance feedback*)
  - Disponible en Google

Caché: [En caché] / [Cached]

- **Copia** de la página cuando se indexó
  - Aunque el enlace no esté disponible en ese momento
  - Puede no estar actualizado
- **Muestra correspondencias** con los términos de la búsqueda

Herramientas del idioma

- Seleccionar idioma del interfaz
- Acotar el idioma de búsqueda
- Acotar el país de búsqueda (*localización*)
- Traducir una página devuelta
- Traducir una página dada (URL)
- Traducir un texto

Buscadores temáticos

- Especializados en temas concretos
- Wikis: `http://www.wiki.com/`
- Legislativos:
  - DOG (`http://www.xunta.es/diario-oficial`)
  - BOE (`http://www.boe.es`)
- Publicaciones científicas:
  - ISI Web of Knowledge (`http://isiknowledge.com`)
  - Google Scholar (`http://scholar.google.com`)
  - Microsoft Academic Search (`http://academic.research.microsoft.com/`)
  - CiteSeerX (`http://citeseer.ist.psu.edu`)

Buscadores temáticos (cont.)

- Noticias:
  - Google Noticias España (`http://www.google.es/nwshp?hl=es`)
  - El Pais `http://www.elpais.com`
  - El Mundo `http://www.elmundo.es`
  - La Voz de Galicia `http://www.lavozdegalicia.es`
- Código de programación:
  - Google Code Search (`http://www.google.com/codesearch`)
- Cine y TV:
  - The Internet Movie Database (IMDb) (`http://www.imdb.com/`)
- ...

Metabuscadores (o multibuscadores)

- **No son buscadores** propiamente dichos:
  1. Lanzan la consulta contra varios buscadores a la vez
  2. Recopilan los diferentes resultados
  3. Devuelven una única lista de resultados

- Ventajas:
  - Reducen esfuerzo:
    - Buscar en varios buscadores a la vez
    - Elimina duplicados
  - Cuando hay poca información disponible

- Desventajas:
  - Demasiados resultados
  - No permite aprovechar características concretas

Metabuscadores (o multibuscadores) (cont.)

- On-line:
  - Metacrawler (`http://www.metacrawler.com`)
  - Mamma (`http://www.mamma.com`)
- Software instalable:
  - iMetaSearch
  - WebSeeker

Colecciones de buscadores

- **Repositorios** de buscadores y directorios ("buscadores de buscadores"):
  - Buscopio (`http://www.buscopio.net`)
  - Search Engine Guide (`http://www.searchengineguide.com`)
  - `http://en.wikipedia.org/wiki/List_of_search_engines`

Wikis

- "Familia" de sitios web editables vía web por los propios usuarios (colaborativos)
  - Temáticas
  - Muy fácil referenciar páginas internamente
  - Mecanismos de control de modificaciones
- En alza:
  - Accesibilidad y comodidad
  - Utilidad
  - Colaborativas
- No requieren software especial para su uso (navegador web), pero sí para su creación (servidor)

Wikis (cont.)

- Privadas: ej. documentación interna de empresas
- Públicas:
  - Wikipedia: enciclopedia (http://www.wikipedia.org/)
  - Wiktionary: diccionarios (`http://www.wiktionary.org/`)
  - Wikibooks: libros de texto (`http://www.wikibooks.org/`)
  - Wikimedia Commons: recursos multimedia (`http://commons.wikimedia.org/wiki/Main_Page`)
- Buscador de wikis: `http://www.wiki.com/`
- Para saber más: `http://en.wikipedia.org/wiki/Wiki`

Foros

- Páginas web que permiten el intercambio de mensajes, que quedan "colgados" en la página a disposición de los demás

  `http://www.infojardin.com/foro`

- Jerarquizados por temática y *hilos de discusión* (primer mensaje)

- Acceso web: sólo requiere navegador

Listas de correo

- Servicio de distribución y recepción de mensajes por email, de forma que cuando se envía un mensaje a la dirección de la lista, lo reciben todos los usuarios suscritos a ella
  - La dirección de email de la lista "engloba" todas las direcciones individuales de los suscriptores
- "Foro por email"
- Vía mail: sólo cliente de correo (MS Outlook, Thunderbird, ...)
- Pueden tener moderador
- Dónde buscar:
  - RedIris (`http://www.rediris.es/list/`)
  - CataList (`http://www.lsoft.com/catalist.html`)

Newsgroups (Grupos *Usenet*)

- Grupos temáticos de debate similares a las listas de correo
- Red *Usenet*: luego **Google Groups** (`http://groups.google.es`)
  - Anterior a la web
  - Jerarquización por temática (muy estricta)
  - Repositorio
- Servidor de *news*: software de gestión de los grupos
- Cliente específico lector de noticias (*newsreader*)
  - Ya integrados en algunos clientes de correo (MS Outlook, Mozilla, etc.)
- El usuario se conecta al servidor de *news* y descarga los mensajes llegados desde su última conexión

Grupos de discusión

- Grupos de debate que combinan las funcionalidades anteriores (foros, listas de correo y *newsgroups*) y un directorio temático
  - Yahoo! Groups (`http://es.groups.yahoo.com`)
  - Antes Google Groups Beta, ahora **Google Groups** (`http://groups.google.es`)
    - Incluye antiguo Google Groups (antiguo *Usenet*)
  - Windows Live Groups (`http://groups.live.com`)

Blogs, *Weblogs* o bitácoras

- Página web en las que uno o varios usuarios opinan acerca de todo tipo de temas y experiencias, generalmente en formato de diario y de forma más bien subjetiva e informal (opiniones)

  `http://amis95.blogspot.com`

- Cómo crear un *blog*:
  - Servicios web especializados
    - Blogger (`http://www.blogger.com`)
  - Software especializado
    - Movable Type (`http://www.movabletype.org`)

- Acceso web: sólo requiere navegador
  - Problema: cómo saber si actualizado
  - Solución: RSS

- Buscadores especializados:
  - Google Blog Search (`http://blogsearch.google.com`)

Web feed: Atom y RSS

- Permite la notificación de actualizaciones en contenidos web y su lectura remota sin tener que acceder a las páginas originales
  - Actualizaciones de un blog, noticias de un periódico *on-line*, etc.
- Requiere:
  - Web a leer configurada como fuente RSS/Atom (iconos)
  - Que el usuario use un lector (*reader*) RSS/Atom
    - Servicios web:
      Google Reader: `http://www.google.com/reader`
      iGoogle: `http://www.google.com/ig`
    - Programas instalables: ver
      `http://email.about.com/od/rssreaderswin/tp/top_rss_windows.ht`

Web feed: Atom y RSS (cont.)

- Funcionamiento:
  - El usuario suscribe su lector a una fuente RSS/Atom (vía un servidor RSS/Atom)
  - Cuando una fuente es actualizada, avisa al servidor, y éste al lector, para que así pueda actualizar los contenidos
- Ventajas:
  - Un mismo lector puede estar suscrito a varias fuentes a la vez

Barras de herramientas

- Programas instalables que anexan un nuevo panel al navegador web mediante el cual podemos lanzar búsquedas sin tener que entrar en la web del buscador
- "*Spin-offs*" de los principales buscadores (y de otras webs o aplicaciones):
  - Google (`http://www.google.com/intl/es/`)
  - Yahoo (`http://es.toolbar.yahoo.com/`)
  - Bing (`http://toolbar.discoverbing.com/toolbar/es-ES.html`)
- Funcionalidades extra: filtro de *popups*, traducción, etc.
- Funcionalidad básica (interfaz de búsqueda) ya integrada en algunos navegadores:
  - Explorer, Firefox, etc.
  - CUIDADO!!! a lo mejor por defecto configurado para buscar en `.com` en lugar de `.es`

Buscadores instalables

- Permiten indexar y buscar tus propios contenidos
- *"De escritorio"*: permiten indexar y buscar en tu PC
  - Incluidos en el sistema operativo (Vista, Win7)
  - "*Spin-offs*" de los principales buscadores
    Google Desktop (`http://desktop.google.es`)
  - Conjuntamente con otro software (ej. Nero)
- Motores de búsqueda independientes
  - Algunos también contenidos web
    Ej. Lucene (`http://lucene.apache.org/`)

Webmasters y buscadores

- **Webmaster**: administrador de un sitio web
- Quiero que mi web sea indexada
  - Enviar la URL al buscador (ver su ayuda)
    - Google ($\pm$1 mes): `http://www.google.es/addurl`
  - Esperar que te indexe: necesarios enlaces a tu página (web oculta)
- Qué no quiero que indexe: fichero `robots.txt`

Webmasters y buscadores (cont.)

- Cómo mejorar tu ranking
  - Que la página funcione
  - Evitar contenidos dinámicos
  - Que la referencien en cantidad y calidad (y viceversa)
  - Páginas no demasiado grandes (5-15K)
  - Incluir palabras clave en
    - Título
    - URL
    - Campos `<META>`: etiquetas `description` y `keywords`
    - Principio del texto
    - Imágenes: nombres de sus archivos y etiquetas `alt` asociadas
    - Textos enlaces internos y externos