

Recursos lingüísticos y terminológicos en Internet



DRA. ISABEL SANTAMARÍA PÉREZ

GRUPO EPA

OBJETIVOS

2

- Consultar y localizar **información** de interés **lingüístico y terminológico**.
- Buscar **textos especializados** para adquirir conocimiento especializado sobre la materia.
- Constituir un **corpus textual especializado**, base del vaciado terminológico.
- Presentar distintos **recursos de consulta lingüística y terminológica** en Internet.

CONTENIDOS

3

- Internet como fuente de información especializada: obtención de corpus
 - Directorios vs. buscadores
- Internet como almacén de recursos lingüísticos y terminológicos

Internet como fuente de información especializada: el corpus

4

Internet



fuente de información lingüística



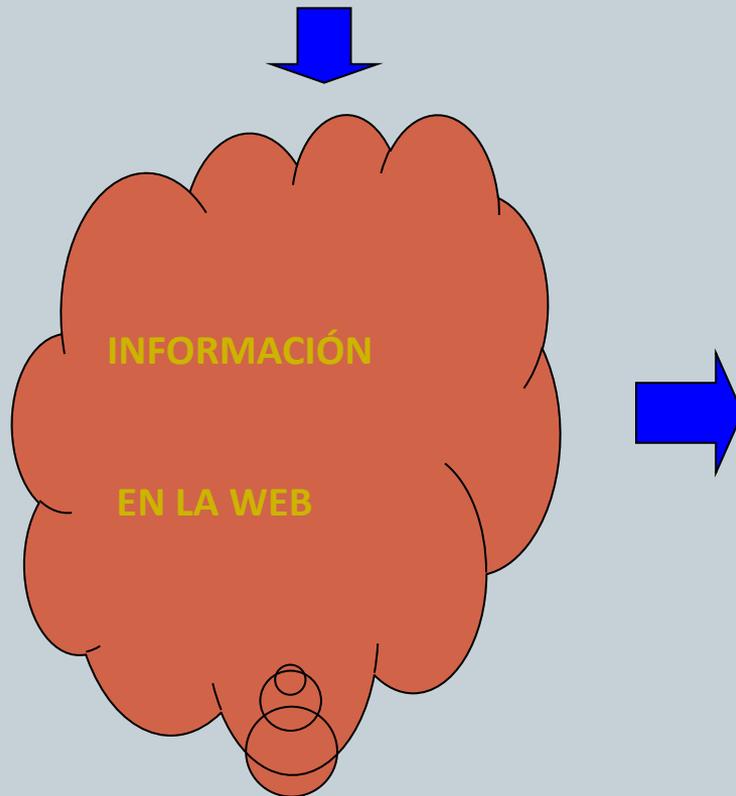
Corpus

- 
- información textual y discursiva
 - información sobre unidades concretas (variantes, definiciones, categoría gramatical)
 - ejemplos de coocurrencias, etc.

Internet como fuente de información especializada: el corpus

5

- Necesidad de información



Criterios o estrategias de búsqueda para la recuperación de información

Herramientas de búsqueda en Internet

Internet como fuente de información especializada: el corpus

6

- ¿Qué tipo de información recuperamos?
 - **Material textual** disponible en Internet que contiene información lingüística.
 - **Productos** finales que contienen información lingüística (bases de datos, glosarios, diccionarios en línea, corpus, etc.).

Directorios vs. motores de búsqueda



BUSCADORES

Índices o directorios

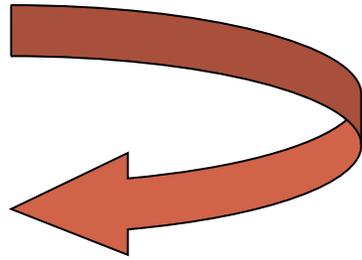
Motores de búsqueda

Metabuscadores



Directorios vs. motores de búsqueda

- Tendencia actual



convertir a los buscadores en **portales** que ofrecen servicios, ventas y productos

Imposibilidad de ubicar a los actuales buscadores en una sola descripción (directorios, motores de búsqueda, etc.)

Directorios vs. motores de búsqueda

9

Directorios

- Búsqueda dirigida
- Funcionan por **navegación**
- Lista de recursos clasificados por temas
- Recogen una pequeña parte de los recursos digitales de la WWW
- Organizan la información manualmente asignando los recursos a una o más clases
- Actualización de la información lenta

Motores de búsqueda

- Consulta libre
- Funcionan por **recuperación de información**
- Programas automáticos que constituyen una BD con todas las páginas web exploradas
- Recogen la totalidad de recursos
- Organizan la información por comparación automática y creación de índices
- Actualización automática y rápida

Indices-directorios

10

Ventajas



- Información clasificada, precisa y de calidad
- Suelen presentar sus enlaces comentados
- Realizada por expertos → calidad
- Permiten buscar por una catalogación temática preexistente
- Es posible navegar por una temática determinada sin concretar la palabra clave

Desventajas

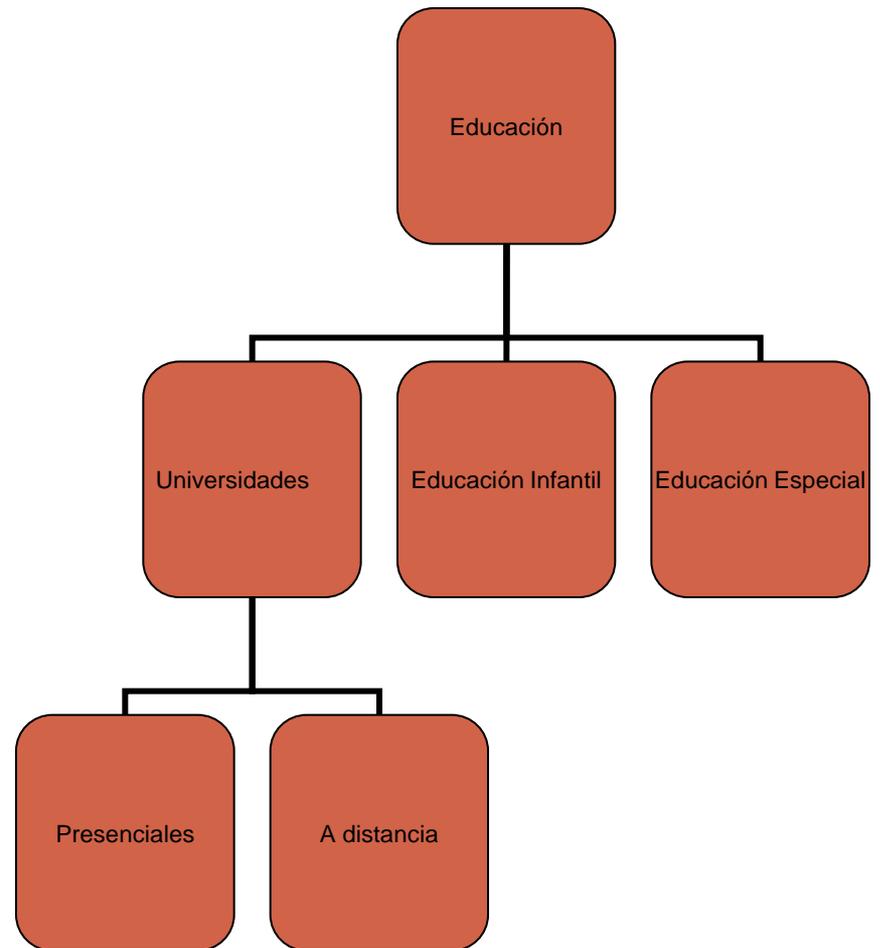


- Resultados menores que con los motores de búsqueda
- Enlaces no siempre actualizados
- Cada directorio tiene su propia clasificación
- Sistema menos intuitivo que en los motores
- En las clasificaciones priman los criterios comerciales

Indices-directorios

11

Clasificación de recursos por categorías y subcategorías temáticas



Motores de búsqueda

12

Ventajas



- Abundante información
- Está todo
- Buscan en toda la red
- Fáciles de usar e intuitivos
- Actualización de los enlaces

Desventajas



- Exceso de información
- Los primeros resultados suelen ser páginas que enlazan con otras
- Presencia de resultados que no interesan

Motores de búsqueda

13

- Recursos **más utilizados** para buscar en Internet
- Se basan en **programas automáticos** que saltan de una página a otra de la web, siguiendo los enlaces que encuentran en cada página para construir una gigantesca base de datos.
- La búsqueda es por **palabras clave**.

Ejemplos

14

- Directorios:
 - Yahoo <http://es.yahoo.com/>
 - Excite <http://www.excite.es/directory>
 - Lycos <http://www.tripod.lycos.es>
- Buscadores:
 - Google <http://www.google.com>
 - AltaVista <http://www.altavista.com>
 - Ozú <http://www.ozu.es>
 - Terra <http://www.terra.es>

Metabuscadores

15

- **Buscador de buscadores**
 - Redirigen las consultas simultáneamente a bases de datos de varios motores y directorios
 - Útiles para consultar datos muy concretos.
 - Algunos permiten realizar una consulta simultánea a todos los buscadores indexados o bien determinar el buscador en concreto al que se desea enviar la interrogación.

- Ejemplos de metabuscadores

- <http://metacrawler.com>

- <http://www.i-une.com>

- <http://www.kartoo.com>

- Otros metabuscadores en:

- ✦ <http://www.cuervoblanco.com/metabuscadores.html>

- ✦ <http://www.buscopio.net>

- ✦ <http://www.internetinvisible.com>

Buscadores especializados

17

- Se estructuran sobre la base de un directorio temático, aunque pueden ofrecer un motor de búsqueda.
- Su atractivo reside en la especialización temática y en la calidad de los vínculos.
- Los recursos que proporcionan son muy variados: portales temáticos, sitios web de asociaciones, páginas personales, etc.

<http://www.scirus.com>

Recuperación de información: operadores booleanos

18

- Permiten expresar nuestras necesidades de información con la máxima precisión.
- Sirven para exigir o excluir conceptos en la búsqueda avanzada y en determinados motores de búsqueda.

Tipos de operadores booleanos

- **AND** (&): localiza documentos que contengan uno y otro términos expresados en la consulta.
- **OR** (I): localiza documentos que contengan uno u otro de los términos especificados en la consulta.
- **NOT** (!): recupera documentos que no contengan un término expresado en la consulta.

Tipos de operadores booleanos

20

- **Frase exacta:** Permite buscar ocurrencias de fragmentos textuales que coincidan exactamente con el contenido de un documento.
 - Se representan:
 - ✦ “transcripción genética”

Tipos de operadores booleanos

21

- **Limitar idioma:** Permite limitar los resultados de la búsqueda a los que se encuentren en una determinada lengua especificada previamente.
- **Limitar a un servidor** (domain: host): Encuentra las páginas con la palabra o frase especificada en el dominio del servidor Web donde se encuentre estas páginas.
 - Domain: .mx (limitamos la búsqueda a ese dominio: México)

Tipos de operadores booleanos

22

- **Limitar a título:** El motor de búsqueda encuentra las páginas que contienen la palabra o frase en el título.
- **Limitar a texto:** el contenido de la página.
- **Limitar a período temporal.**

Tipos de operadores booleanos

23

- **Formatos de documentos en Internet**
 - Páginas .html
 - Documentos de word .doc
 - Archivos .pdf
 - Imágenes .jpg, .gif
 - Comprensión y descomprensión de ficheros .zip

Evaluación de resultados

24

- Autoría
- Actualidad
- Precisión y corrección
- Tratamiento del contenido
- Originalidad
- Enlaces a otros recursos
- Ergonomía
- Citación en otros sitios

Internet como almacén de recursos lingüísticos y terminológicos

25

- a. Recursos lexicográficos
- b. Recursos terminológicos: catálogos de diccionarios especializados
- c. Recursos textuales
- d. Bases de datos terminológicos y neológicos
- e. Centros de referencia y normalización terminológica
- f. Consulta con otros especialistas: listas de discusión y foros...

a) Recursos lexicográficos

26

- Dictionarios y enciclopedias
 - Español
 - <http://www.rae.es>
 - <http://clavecorp.grupo-sm.com>
 - Catalán
 - <http://pdl.iecat.net>
 - Inglés
 - <http://www.m-w.com> (Merriam Webster)
 - <http://www.ldoceonline.com> (Longman)
 - <http://dictionary.cambridge.org>
 - <http://www.hyperdictionary.com>
 - <http://www.oed.com/> (Oxford English Dictionary)
 - Francés
 - <http://atilf.atilf.fr/tlfv3.htm>
 - Alemán
 - <http://www.dwds.de/>

a) Recursos lexicográficos

27

○ Portales

- ✦ <http://www.yourdictionary.com>. Portal con 2.500 diccionarios y gramáticas en 300 lenguas
- ✦ <http://www.onelook.com/browse.shtml>. Portal con cerca de 900 diccionarios
- ✦ <http://www.diccionarios.com>. Portal de Larousse.
- ✦ <http://www.lexicool.com>. Portal con más de 4.000 enlaces de diccionarios y glosarios bilingües y multilingües.
- ✦ <http://elmundo.es>: Portal que ofrece diccionarios generales monolingües y bilingües

b) Recursos terminológicos

- Portales

- Portalingua <http://www.portalingua.info>. Portal dedicado a las lenguas neolatina. Consulta de recursos en línea.
- TermSciences <http://www.termosciences.fr/>. Portal terminológico multidisciplinario. Permite la consulta de términos en francés, inglés, español y alemán
- Eureka <http://www.foreignword.com/eureka/default.asp>. Portal que recoge información sobre bancos de datos terminológicos, glosarios, diccionarios, traducción, etc.
- Latinosapiens <http://www.latinosapiens.info>. Portal de información científica en lenguas neolatina. Contiene artículos, foros, revistas.

b) Recursos terminológicos

29

- Catálogo de glosarios y diccionarios especializados
 - Hiperenciclopèdia <http://www.grec.net/home/cel/mainc.htm>
 - Visual dictionary <http://www.infovisual.info/>
 - Glossarist <http://www.glossarist.com/>
 - Terminology <http://lipas.uwasa.fi/comm/termino/collect/> Collection
 - Vademécum del traductor externo http://www.europarl.europa.eu/transl_es/plataforma/pagina/celter/celula.htm
 - Inventerm <http://www.inventerm.com/>

c) Recursos textuales

30

- **Corpus**
 - BwanaNet
 - ✦ <http://brangaene.upf.es/bwananet/indexes.htm>
 - BNC (British National Corpus)
 - ✦ <http://www.natcorp.ox.ac.uk/>
 - CREA (Corpus de Referencia del español actual)
 - Corpus del español de Mark Davies
<http://www.corpusdelespanol.org/>
 - Internet (Scirus [científico], Google)

d) Bancos de datos terminológicos

31

a) ¿Qué es una BDT?

- Recopilación estructurada y automatizada de unidades terminológicas (UT), acompañadas de información sobre ellas, destinado a responder las necesidades de un grupo definido de usuarios.

d) Bancos de datos terminológicos

32

b) Perfiles de usuarios

- traductores
- especialistas
- normalizadores
- lexicógrafos generales y especializados
- documentalistas
- terminólogos
- asesores y planificadores lingüísticos
- redactores técnicos, periodistas científicos
- ingenieros lingüísticos y lingüistas computacionales

d) Bancos de datos terminológicos

33

c) Tipos de BDT

1 Según los objetivos

- traduccionales
- normalizadores
- especializados

2 Según la temática

- bancos especializados en general: distintas áreas
- bancos muy especializados: una sola área

d) Bancos de datos terminológicos

34

- 4 Según las dimensiones
 - macrobancos
 - minibancos
- 5 Según el número de lenguas
 - monolingües
 - plurilingües propiamente dichos
 - monolingües con equivalencias en otras lenguas
- 6 Según el modo de organizar los datos
 - a partir de documentos
 - a partir de términos

d) Bancos de datos terminológicos

35

7 Según los autores:

- organismo normalizador
- organismo internacional
- empresa multinacional
- servicio lingüístico universitario, etc.

8 Según las posibilidades de acceso:

- por Internet: libre, previa suscripción, de pago.
- CD-ROM

d) Bancos de datos terminológicos y neológicos

- TERMITE (Banco creado por la International Telecommunications Union) <http://www.itu.int/terminology/index.html>
- Terminología de la FAO <http://www.fao.org/faoterm/index.asp?lang=ES>
- IATE <http://iate.europa.eu/>
- UPF_Term <http://upfterm.upf.edu:8080/cat/index.htm>
- TIS (Terminological Information System) <http://tis.consilium.eu.int/utfwebtis/frames/introfsEN.htm>
- Termium http://www.termium.gc.ca/site/accueil_home_f.html
- Cercaterm <http://www.termcat.es>
- Neoloteca <http://www.termcat.es>
- Bobneo <http://obneo.iula.upf.edu/bobneo/>

d) BDT vs. recopilaciones

37

- Ejemplos de recopilaciones:

- *Terminology Collection*

<http://www.uwasa.fi/comm/termino/collect/index.html>

- Lista de diccionarios de *Linguist List*

<http://linguistlist.org/sp/Dict.html>

- Traductor virtual Univ. Vigo

<http://www.uvigo.es/webs/sli/virtual/diccion.htm>

e) Organismos relacionados con la terminología

- **Regionales**
 - ACATERM (www.acaterm.org) ISO (www.iso.ch)
 - TERMCAT (www.termcat.net)
 - UZEI (www.uzei.com). Banco de datos: EUSKALTERM (www1.euskadi.net/euskalterm/indice_c.htm)
 - TERMIGAL (<http://www.cirp.es/>)
- **Nacionales**
 - AENOR (www.aenor.es)
 - AETER (racefyn.insde.es/AETER/indice_AETER.htm)
- **Internacionales**
 - INFOTERM (<http://www.infoterm.info/>)
 - RITERM (www.riterm.net)
 - REALITER (www.realiter.net)
 - ISO (www.iso.org)

f) Listas de discusión, foros y blogs

- The Linguist List. Conjunto de listas (corpora, ELEX, lexicography)
- Relación temática de listas de distribución de RedIris (tecnotrad, traducción, tradumática)
- Lista Zéfir. Lista de distribución para traductores en catalán.
- Tradumática. Lista dedicada a la información, cooperación e intercambio entre profesionales.
- Lista Termilat. Lista de la Unión Latina dedicada a la terminología.
- Blogs: Notas terminológicas, Bases de datos terminológicos, La ciencia y la terminología, Terminología en catalán, etc.

Bibliografía

40

- “Recursos para la terminología a través de Internet: búsqueda y evaluación”, Talleres on line de Terminología, IULA, Universitat Pompeu Fabra.