

Bibliotecas digitales: metainformación y nuevos servicios



Rafael C. Carrasco (carrasco@dlsi.ua.es)

Director Adjunto de la Biblioteca Virtual Miguel de Cervantes

(<http://www.cervantesvirtual.com>)



Esquema

1. La Biblioteca Virtual Miguel de Cervantes.
2. ¿Llegan? Buscadores y demás.
3. ¿Se quedan? Accesibilidad, usabilidad y otras “bilidades” .
4. ¿Vuelven? Servicios y aprovechamiento de la metainformación .

Historia



La Biblioteca Digital Miguel de Cervantes:

- Fue creada en julio de 1999.
- Es fruto de la colaboración entre la Universidad de Alicante y el Banco Santander Central Hispano (Fundación Marcelino Botín).



Objetivos

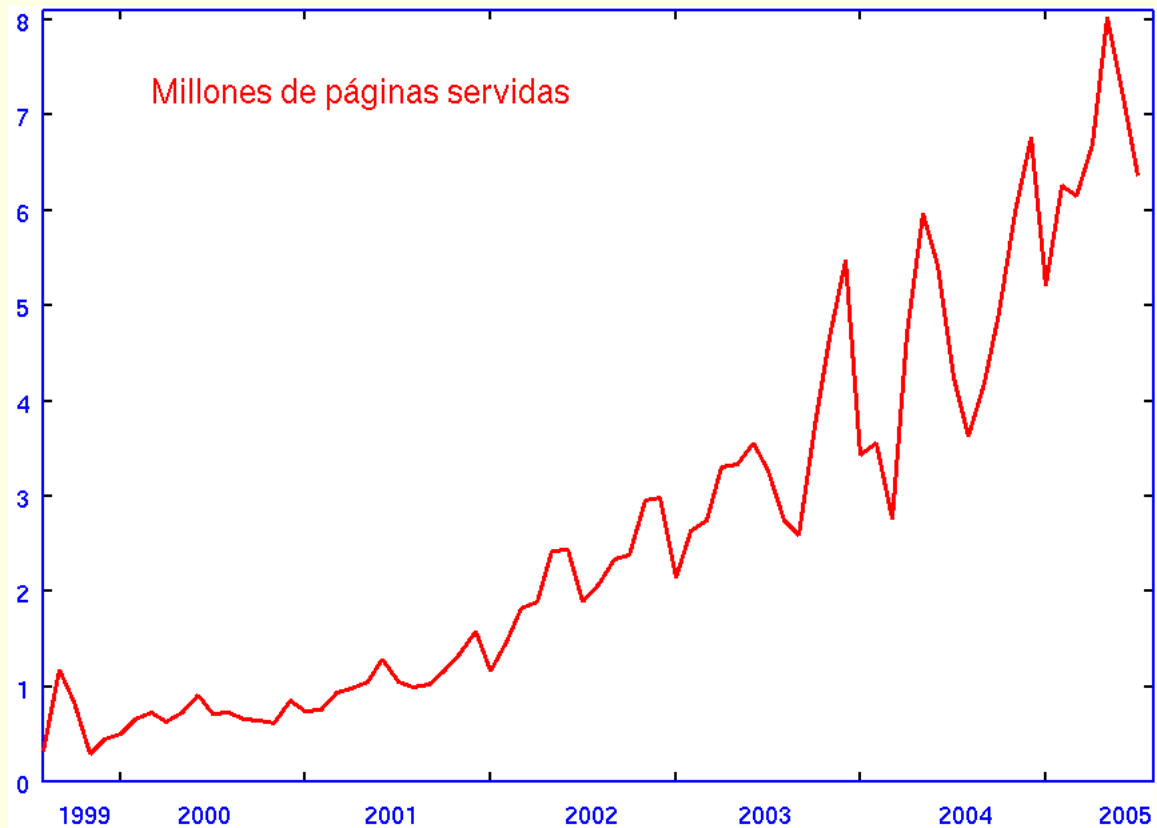
Su **objetivo** fundamental es la difusión en formato digital (fundamentalmente a través de **internet**) de la cultura española e hispanoamericana.

1. Las ediciones son seleccionadas por criterios de interés científico.
2. La mayor parte de los textos está transcrita y supervisada por expertos.
3. Los textos incluyen **metainformación** sobre su estructura.

Repercusión de la biblioteca

Páginas servidas con éxito (anual)	71.441.417
Promedio diario de páginas servidas con éxito	195.729
Promedio de transferencia diaria	8,88 GB

Repercusión de la biblioteca



Buscadores

Los enlaces recibidos importan, pero no son el único factor.

Dominio	Google	alltheweb	PR
cervantesvirtual.com	7.800	38.100	7
www.gutenberg.org	14.400	305.000	8
cvc.cervantes.es	635	14.900	7
www.rae.es	0	3.940	8
rincondelvago.com	7.690	124	6
www.elmundo.es	33.400	1.100.000	8

(Fuente: <http://www.prchecker.info>)

Buscadores



alltheweb [advanced search](#) :: [customize preferences](#) :: [submit site](#) :: [help](#)

ladrones

Results in: [Any Language](#) [English, Spanish](#)

Web | News | Pictures | Video | Audio

1 - 10 of 546,000 Results for **ladrones** Offensive content filter: [On](#) - [Off](#)

Refine your search: (click "+" or "-" to include or exclude terms, and then click the "SEARCH" button)

- los **ladrones**
- ladrones** proa

Web Results ([What's this?](#))

Sociedad General de Autores y Editores
Web Sgae. SERVICIOS INTERNET. Programación Barcelona Bilbao Madrid Oviedo Valencia. El 'Anuario SGAE 2005' compila en 600 páginas las claves de un año de cultura en España
[more hits from: http://www.sgae.es/](http://www.sgae.es/) - 33 KB

ADSL - Terra
(*Promociones vigentes desde el 1 de junio de 2005 hasta el 30 de junio de 2005. Precios sin IVA incluido
[more hits from: http://www.terra.es/adsl](http://www.terra.es/adsl) - 18 KB

Ladrones.
... Basta con que busquemos el término **ladrones** en Google. ... En este caso, si buscamos en Google la palabra **ladrones**, el primer resultado que obtenemos es el de la página ...
[more hits from: http://www.hispamp3.com/noticias/noticia.php?noticia=20040427190728](http://www.hispamp3.com/noticias/noticia.php?noticia=20040427190728) - 26 KB

Trasteando > Google bombing contra la SGAE
... Las pretensiones de esta sociedad de **ladrones** van muchos más lejos, canon a discos duros y supongo que ... han hecho cuando se busque "**ladrones**", se encontrará la SGAE por lo ...
[more hits from: http://nkieto.f2o.org/blog/archivos/000099.php](http://nkieto.f2o.org/blog/archivos/000099.php) - 30 KB



Buscadores

- ¿Qué se demanda?
- ¿Cuál es la oferta?
- Démonos a conocer.
- Conozcamos al cliente.



Buscadores

Los términos más buscados pueden consultarse en sitios como <http://inventory.overture.com/>.

Permite el estudio de la demanda.

Términos	promedio diario
miguel de cervantes	189.7
cervantes	169.5
miguel de cervantes saavedra	87.5
miguel cervantes	59.6
instituto cervantes	34.5
cervantes don quixote	23.3
miguel de cervantes biography	22.5
cervantes marchetti roma walter	16.6
lorna dee cervantes	12.8
biografia de miguel de cervantes saavedra	11.9
cervantes virtual	11.1
biografia de miguel cervantes	10.9
biografia de miguel de cervantes	10.4
miguel cervantes saavedra	10.3
biblioteca cervantes	9.9
hotel cervantes	8.5
cervantes institute	8.2
edgar cervantes	7.7

Buscadores

¿Cuáles son las obras más consultadas?

- El ingenioso hidalgo Don Quijote de la Mancha
- México (Constituciones hispanoamericanas)
- La Celestina
- Diccionario del español usual en México
- La Ética de Aristóteles
- La vida es sueño (Pedro Calderón de la Barca)
- La vida de Lazarillo de Tormes y de sus fortunas y adversidades.
- Gramática de la Lengua Castellana (RAE)
- Poema de Mío Cid (facsímil).



Buscadores

El nivel de competencia pueden conocerse en sitios como <http://www.sitecook.com>.







Permite estudiar la oferta.

Search engine positioning report for cervantesvirtual.com

Analysis complete.

Results for cervantesvirtual.com

[Archive](#) | [Google time series](#) | [Frequently Asked Questions](#)

	search match density	AltaVista	Google	Fast
Total number of pages indexed		45,300	1,010,000	68
Top page position for selected search phrases				
'lengua de signos' [remove]		16	>100	>100
'literatura chicana' [remove]		22	>100	>100
'literatura infantil' [remove]		8	>100	>100
Celestina [remove]		19	>100	>100
Cervantes [remove]		41	>100	>100
Quijote [remove]		63	>100	44

Search Density : Frequency of searches made for this phrase (4 bars max)

Match Density : Number of sites returned for this phrase (4 bars max)

More green bars and fewer red bars are likely to indicate a good search term to target.

Directories

[Open Directory](#) : 218 entries (approx.)

Record for cervantesvirtual.com created on Mon Jun 27 05:23:38 EDT 2005
 next update can be made on the 1st of next month;
 additional search terms can be added at any time



Buscadores

Darse a conocer requiere a veces no sólo darse de alta sino también aportar metainformación: <http://dmoz.org>

Search: "**literatura infantil**"

Open Directory Categories (1-5 of 5)

1. [Kids and Teens: International: Español: Artes: Literatura](#) (21 matches)
2. [Kids and Teens: International: Català: Arts i cultura](#) (3)
3. [World: Català: Arts i cultura: Literatura: Gèneres: Infantil i juvenil](#) (2)
4. [World: Español: Artes: Literatura: Infantil y juvenil](#) (4)
5. [World: Català: Arts i cultura: Literatura: Certàmens](#) (1)

Open Directory Sites (1-20 of 72)

1. [Biblioteca de literatura infantil y juvenil](#) - Proyecto vinculado a la Biblioteca Virtual Miguel de Cervantes, contiene una extensa bibliografía ordenada temáticamente y por autor.
– <http://cervantesvirtual.com/portal/platero/> [Kids and Teens: International: Español: Artes: Literatura](#) (21)
2. [Història de la literatura infantil i juvenil](#) - Breu història de la narrativa infantil, antologies escolars i revistes per a infants i joves durant el modernisme i el noucentisme.
– <http://www.xtec.net/~malons22/personal/infantil.htm> [Kids and Teens: International: Català: Arts i cultura](#) (3)
3. [Història de la literatura infantil i juvenil](#) - Breu història de la narrativa infantil, antologies escolars i revistes per a infants i joves durant el modernisme i el noucentisme.
– <http://www.xtec.net/~malons22/personal/infantil.htm> [World: Català: Arts i cultura: Literatura: Gèneres: Infantil i juvenil](#) (2)
4. [Literatura infantil](#) - Biblioteca electrónica del Instituto Latinoamericano de la Comunicación Educativa, de cuentos y narraciones infantiles.
– <http://lectura.ilce.edu.mx:3000/biblioteca/sites/litin/> [World: Español: Artes: Literatura: Infantil y juvenil](#) (4)



Buscadores

Keywords: cervantes literatura

Domain	Google	G. AOL	IWon	Amazon A9	Dmoz	Yahoo Index	Ink tomi	Alta Vista	Yahoo	MSN	Teoma	Ask Jeeves	Hot Bot	Lycos
cervantesvirtual.com	3	3	3	3	2	2	2	-	-	-	-	-	1	1
cvc.cervantes.es	1	1	1	1	1	-	-	-	-	8	-	-	3	3
rinconcastellano.com	9	9	9	9	7	-	-	-	-	-	-	-	-	-
centraldellibro.com	10	10	10	10	-	-	-	-	-	-	-	-	-	-
elmundo.es	-	-	-	-	-	-	-	-	-	-	3	3	7	7
liceus.com	8	8	8	8	-	-	-	-	-	-	-	-	-	-
csdl.tamu.edu	-	-	-	-	10	3	3	-	-	6	-	-	-	-
iblnews.es	-	-	-	-	-	-	-	-	-	-	4	4	10	10
aache.com	7	7	7	7	-	-	-	-	-	-	-	-	-	-

Fuente: <http://www.seotoolset.com>



Accesibilidad

La biblioteca está siguiendo las normas de accesibilidad del W3C

(<http://www.w3.org/WAI>)

Hay servicios de validación en línea como

(<http://webxact.watchfire.com/>)

Desarrollamos además un proyecto de acceso multimodal con Telefónica I+D.

Accesibilidad

Las normas de accesibilidad son útiles para todos:

1. Mejora la audiencia (indexadores, internacionalización, claridad de presentación, velocidad de acceso,...).
2. Simplifica la gestión (p.ej. actualización).
3. Aumenta la eficiencia (p.ej. carga del servidor).



Facilidad de uso

Para simplificar el uso de la biblioteca se realizaron análisis propios y tests a los usuarios:

- Los usuarios no recuerdan la dirección de la página.

Inclusión de iconos en las cabeceras para página de inicio y para añadir a favoritos.

- Los usuarios no saben por dónde empezar.

Creación de una visita guiada.

- Algunas secciones no explican su cometido.

Introducción de textos explicativos en cada apartado de la biblioteca.

Facilidad de uso

- Dificultad para interpretar los resultados de las búsquedas.
Se presenta un texto que diferencia los resultados obtenidos.
- Algunas obras del catálogo eran enlaces externos.
Apertura de los enlaces externos en ventanas nuevas.
- Páginas con exceso de texto provocan abandonos.
Utilización de un lenguaje claro y conciso.
- Algunas páginas tardan demasiado en aparecer.
Limitación del tamaño de las páginas.
- Exceso de elementos decorativos.
Simplificación del diseño.

Metainformación: ¿qué es?

La metainformación es información sobre la información, esto es, sobre el contenido:

155.2

NUTes Nutlin, Joseph

La estructura de la personalidad / Joseph Nutlin.
Buenos Aires : Kapelusz, 1973.

237 p. : il.- - (Biblioteca de Psicología Contemporánea ; 27)

1.- PSICOLOGIA 2.- PERSONALIDAD

Metainformación



Internet contiene información mal estructurada sintácticamente y sin significado (semántica).

[Comprar carcasa rojo ferrari nokia 3200 en diarioSport.com ...](#) - [[Translate this page](#)]

Comprar carcasa rojo **ferrari** nokia 3200 en diarioSport.com: Compara y Compra las mejores marcas a los mejores precios.

tienda.sport.es/share-cgi/ search.ftcb?k=Carcasa+Rojo+Ferrari+Nokia+3200&f=1&id=23 - 89k - [Cached](#) - [Similar pages](#)

[Comprar nokia 6100 rojo en diarioSport.com: Compara y Compra ...](#) - [[Translate this page](#)]

Comprar - De Compras. Frontal Nokia 6100 Rojo **Ferrari** [Buscar Similares] Blautel Nuevas carcasas de colores vivos para el nokia 6100. ...

tienda.sport.es/share-cgi/search.ftcb?k=Nokia+6100+Rojo&f=1&id=23 - 102k - [Cached](#) - [Similar pages](#)

[Blogs Informativos Telecinco - Fórmula 1 - Alonso se sincera](#) - [[Translate this page](#)]

SUERTE ALONSO ERES EL MEJOR,todo **ALICANTE** esta con tigo. ... Creo que **Ferrari** y Schumacher no se han metido con Alonso como sí ha sucedido al revés. ...

www.blogs.telecinco.es/formula1/ post/2005/06/23/alonso-se-sincera - 33k - [Cached](#) - [Similar pages](#)

[Ordenador portátil ACER FERRARI - Precios10.com](#) - [[Translate this page](#)]

Ordenador Portátil Acer **Ferrari** al mejor precio. Punto de venta Acer. ... Para realizar su pedido pulse **COMPRAR AQUÍ** a la izquierda del producto. ...

www.precios10.com/portatiles/acer_ferrari.shtml - 36k - [Cached](#) - [Similar pages](#)

[ferrari](#) - [[Translate this page](#)]

Todos los anuncio >> **Comprar** / Vender >> coches / motos / bicis ... Vendo peugeot 306 2.0HDI deportivo color rojo **ferrari** 3 puertas | coches / motos ...

www.mundoanuncio.com/categoria/ coches_motos_bicis_20/buscarferrari_1.html - 61k - [Cached](#) - [Similar pages](#)

[::: Vendo Ordenador PC en Alicante venta de Ordenadores de segunda ...](#) - [[Translate this page](#)]

Portatil Acer **Ferrari**, pura potencia y diseño. AMD Athlon™ ... 100 € - Vendo dreamcast+juegos+mandos+accesorios (**alicante**) ...

alicante.campusanuncios.com/ Ofertaventa-Alicante-5-5.html - 42k - [Cached](#) - [Similar pages](#)

[Información de empresas de ALICANTE : HispaVista : La Bolsa ...](#) - [[Translate this page](#)]

FERRANDIZ BARBER MARIA DOLORES 000258765B SLNE. **FERRARI** STUDIO SL ... Comunicación y utilidades, **Comprar** y vender, Información, Ocio ...

labolsa.com/empresa/provincia/alicante/F/ - [Similar pages](#)

Sponsored Links

[Ferrari](#)

Gorras, fotos, camisetas y todo tipo de artículos **Ferrari** en eBay
www.ebay.es

[Ferrari](#)

Todo sobre el mundo del motor
Análisis, noticias, precios, fotos
es.cars.yahoo.com/

[Cheap Flights To Alicante](#)

Huge Discounts on **Alicante** Flights.
A Wide Range of Amazing Deals.
www.flights4less.net

[¿Buscas casa en Alicante?](#)

Es facilisimo.com: encontrar casa, disfrutar un hogar. 125.000 casas
www.facilisimo.com



Metainformación: Dublin Core

Nuestras páginas contienen metainformación según el estándar Dublin Core. (<http://dublincore.org>)

Los buscadores no dan mucha relevancia a la metainformación, ya que no pueden validarla. Parece ayudar si aparece también en el texto y se evitan repeticiones.



Metainformación: Text Encoding Initiative

También tenemos metainformación estructural mediante el vocabulario de marcado de la Text Encoding Initiative para textos literarios y lingüísticos. (<http://www.tei-c.org>)

El vocabulario **TEI** es usado por numerosos proyectos (PERSEUS, Oxford Text Archives, WWP).

Metainformación: TEI

```
<title> Doña Perfecta </title>
<author> Benito Pérez Galdós </author>
<text>
  ....
  <note>
    Doña Perfecta, como mujer de su época, ....
  </note>
</text>
```

1. El autor es identificado con el nombre con el que aparece en el libro.
2. Podemos, por ejemplo, localizar los comentarios del editor sobre el personaje.

Metainformación: TEI

El marcado estructural permite:

1. separar las tareas del editor y del tipógrafo;
2. realizar búsquedas elaboradas: p. ej. Sevilla en un pie de imagen;
3. cambiar el formato o estilo de la presentación para todos los documentos a un tiempo.

Metainformación: TEI

La estructura básica:

```
<TEI.2>  
  <teiHeader> [ TEI Header information ] </teiHeader>  
  <text>  
    <front> [ front matter ... ] </front>  
    <body> [ body of text ... ] </body>  
    <back> [ back matter ... ] </back>  
  </text>  
</TEI.2>
```

Metainformación: TEI

Algunos ejemplos:

- **p**: párrafo en prosa;
- **l**: línea (quizá incompleta) de verso;
- **sp**, **speaker**: discurso o parlamento de un personaje y personaje;
- **hi**: texto destacado;
- **pb**, **lb**: cambio de página o línea;
- **div**, **div1**, ...: secciones del libro.

Metainformación: TEI

Las etiquetas más utilizadas

1	p	450059	26.9 %	21	figure	4807	0.28 %
2	l	402754	24.1 %	22	change	3955	0.23 %
3	hi	249863	14.9 %	23	resp	3955	0.23 %
4	sp	84453	5.06 %	24	respStmt	3955	0.23 %
5	speaker	84277	5.05 %	25	div2	3659	0.22 %
6	pb	67971	4.07 %	26	title	3080	0.18 %
7	cell	41664	2.49 %	27	bibl	2499	0.14 %
8	note	34880	2.09 %	28	language	2474	0.14 %
9	head	32020	1.91 %	29	role	2319	0.14 %
10	foreign	29178	1.74 %	30	castItem	2319	0.14 %
11	stage	23103	1.38 %	31	milestone	1540	0.09 %
12	name	21922	1.31 %	32	div3	1144	0.06 %
13	lg	18731	1.12 %	33	titlePart	1051	0.06 %
14	row	13379	0.802 %	34	front	1051	0.06 %
15	div1	11452	0.686 %	35	docTitle	1051	0.06 %
16	item	11252	0.674 %	36	titlePage	1051	0.06 %
17	q	10323	0.618 %	37	text	1051	0.06 %
18	div0	6568	0.393 %	38	body	1046	0.06 %
19	eDate	4981	0.298 %	39	table	1045	0.06 %
20	lb	4901	0.293 %	40	author	1033	0.06 %

Metainformación estructural



Formulario de búsqueda avanzada en textos

Búsqueda de

Las PALABRAS en

Las PALABRAS en
 Todos

Las PALABRAS en distinto al de la obra
 Todas las PALABRAS

<p>Restringir por</p> <hr/> <p>TÍTULO de la obra <input type="text"/></p> <p>AUTOR de la obra <input type="text"/> Consultar el catálogo</p> <p><input type="radio"/> Escritura de la obra</p> <p>PERÍODO de <input type="text"/> entre <input type="text" value="Previas s. XIV"/> y <input type="text" value="siglo XXI"/></p> <p><input type="radio"/> Edición</p>	<p>Opciones</p> <hr/> <p><input type="radio"/> Buscar frase exacta</p> <p><input checked="" type="radio"/> Buscar palabras sueltas ordenadas</p> <p><input checked="" type="radio"/> Buscar palabras sueltas no ordenadas</p> <p><input type="checkbox"/> Buscar en el interior de las notas del editor</p> <p><input type="checkbox"/> Distinguir mayúsculas/minúsculas</p> <p><input checked="" type="checkbox"/> Ordenar resultados</p>
--	---

Done.

Metainformación estructural



The screenshot shows the 'Test design tool' interface. At the top left is the logo of the Biblioteca Virtual Miguel de Cervantes. The title 'Test design tool' is centered at the top. Below the title are three buttons: 'New search', 'To Test', and 'Save to repository'. A horizontal line separates this header from the 'PREVIEW' section. The 'PREVIEW' section contains a text input field with the instruction 'Select the appropriate form of the following words:' and a list of words: 'oír', 'escuchar', 'sentir', and 'percibir'. Below this are two numbered sentences with dropdown menus for word selection. The first sentence is '1. Y ella va lentamente hasta la puerta. [dropdown] unos instantes: y como no [dropdown] nada, abre un poco el postigo. (Madame Clay, F. Alloza Villagrasa)'. The second sentence is '2. En mi mesa de escribir, los observo, pero no los veo. Y [dropdown] pero no los [dropdown]. (Búsqueda, Saskia Saer)'. Another horizontal line separates the 'PREVIEW' section from the 'OPTIONS' section. The 'OPTIONS' section contains three checkboxes: 'Score: ', 'Visible answers: ', and 'Show reference: '. At the bottom are three buttons: 'Remove phrases', 'Publish as HTML', and 'Publish as PDF'.

BIBLIOTECA VIRTUAL
MIGUEL DE
CERVANTES

Test design tool

New search To Test Save to repository

PREVIEW

Select the appropriate form of the following words:
oír escuchar sentir percibir

1. Y ella va lentamente hasta la puerta. [dropdown] unos instantes: y como no [dropdown] nada, abre un poco el postigo. (*Madame Clay*, F. Alloza Villagrasa).

2. En mi mesa de escribir, los observo, pero no los veo. Y [dropdown] pero no los [dropdown]. (*Búsqueda*, Saskia Saer)

OPTIONS

Score: Visible answers: Show reference:

Remove phrases Publish as HTML Publish as PDF

Metainformación semántica

The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.

Tim Berners-Lee

Metainformación semántica

“web semántica” = significado + reglas deductivas.

La UA tiene código postal 03071

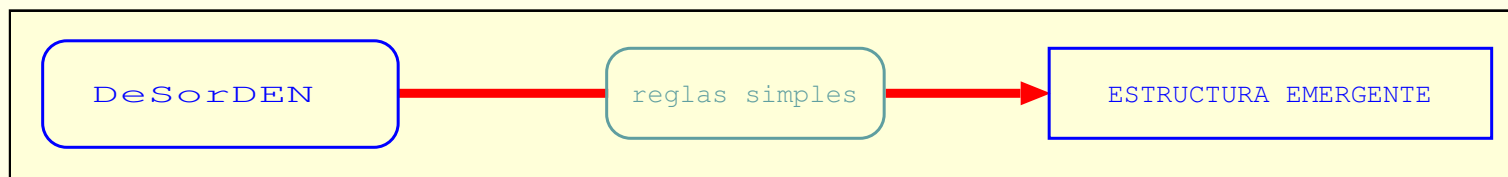
03071 es un código postal de Alicante

L'altet es el aeropuerto más cercano a Alicante

L'altet es el aeropuerto mas cercano a la UA.

Metainformación semántica

Colonia virtual: Mitchel Resnick, “Turtles, termites and traffic jams”





Este documento está disponible en

<http://www.cervantesvirtual.com/documentos/denia.pdf>