



Universidad  
Carlos III de Madrid

Departamento de INFORMÁTICA.

PROYECTO FIN DE CARRERA

# ANALIZADOR SEMÁNTICO: Mujer Olímpica 2012

Autor: Richard Vega Caridad

Titulación: Ingeniería Técnica en Informática de Gestión

Tutor: Almudena Alcaide Raya

Leganés, 28 de julio de 2011



Título: ANALIZADOR SEMÁNTICO: Mujer Olímpica 2012

Autor: Richard Vega Caridad

Director: Almudena Alcaide Raya

## EL TRIBUNAL

Presidente: D<sup>a</sup>. Esther Palomar

Vocal: D. Juan Manuel Tirado

Secretario: D. Guillermo Suárez

Realizado el acto de defensa y lectura del Proyecto Fin de Carrera el día 28 de julio de 2011 en Leganés, en la Escuela Politécnica Superior de la Universidad Carlos III de Madrid, acuerda otorgarle la CALIFICACIÓN de

VOCAL

SECRETARIO

PRESIDENTE



# Agradecimientos

A mi familia por la paciencia y el apoyo sin los que hoy no sería nadie.

A mi tutora Almudena y a todo el laboratorio de SETI por las horas que me han aguantado con ellos, siempre poniéndome las cosas fáciles y ayudándome en todo lo que he necesitado.

A Alex Casas por su aportación y a JL por su clases.



# Resumen

Como parte de las conclusiones del Seminario Permanente “Mujer y Deporte”, celebrado en la Facultad de Ciencias de la Actividad Física y del Deporte de la Universidad Politécnica de Madrid, en las jornadas sobre “El deporte femenino en los medios de comunicación”, celebrado en la semana del 5 al 11 de noviembre de 2009, se establece que en nuestro país “**el éxito de las mujeres en algunos deportes es igual o superior al de los hombres y, a pesar de ello, las mujeres siguen sin tener apenas cabida en los medios de comunicación**”.

Las razones por las que este fenómeno existe son múltiples y complejas aunque se puede señalar como factor principal la diferencia en el poder de financiación y apoyo económico que se da entre las categorías femenina y masculina dentro de un mismo deporte.

El presente proyecto se plantea como trabajo de investigación que pretende mostrar cómo, a través de plataformas tecnológicas como son la Web 2.0., se puede dar impulso al deporte femenino.

Para la consecución de este objetivo se ha desarrollado un algoritmo de búsqueda de noticias que traten sobre los logros de las deportistas femeninas españolas en disciplinas olímpicas y paraolímpicas, para la posterior visualización en una plataforma web. La plataforma web diseñada se encuentra totalmente automatizada no necesitando de ningún *web master* al uso, reduciendo así los costes de mantenimiento y explotación. La aplicación podría ser configurada para el seguimiento de eventos especiales como pueden ser competiciones europeas o las olimpiadas del 2012.

Este buscador de noticias será el **primer buscador semántico en dicha materia**.

En resumen, este proyecto pretende ser una puerta que acerque el deporte femenino a los usuarios de Internet, demostrando así que los logros del deporte español no son sólo masculinos sino que el deporte femenino logra los mismos o incluso mejores éxitos.

**Palabras clave:** algoritmo de búsqueda, analizador semántico, logros del deporte femenino, plataforma web automatizada.





# Abstract

The statement "**the success of women in some sports is equal or superior to men, and despite this, women still have no place in the media**", is part of the conclusions of the Permanent Seminar on "Women and Sport" held in the Faculty of Physical Education and Sport, of the Polytechnic University of Madrid, at the conference on "Women's sports in the media, 2009".

The reasons for this phenomenon are many and complex but one of the main factors highlights the difference in funding possibilities and financial support that exists between male and female categories in the same sport.

This project aims to show how through technology platforms, such as Web 2.0, you can give stimulus to women's sports.

To achieve this goal we have developed a search algorithm that retrieves news of female athletes' achievements in Spanish Olympic and Paralympics disciplines. The results will be displayed in a web platform which will serve as a proof of concept. The web platform is fully automated not needing of any webmaster, thus reducing maintenance and operating costs. The application could be configured to monitor special events, as well as national, European or international competitions such as the Olympic Games in 2012.

**The news search engine is the first semantic search engine in that area.**

In summary, this project will be a door that brings female sport to Internet users, showing that Spanish sport achievers are not only men but also women who achieve the same or better successes.

**Keywords:** searching algorithms, semantic analysis, sportswomen achievements, automated web platform.



# Índice general

<b>CAPÍTULO 1: INTRODUCCIÓN</b> .....	<b>1</b>
<b>1.1. INTRODUCCIÓN</b> .....	<b>3</b>
<b>1.2. DESCRIPCIÓN DEL ÁMBITO DEL PROYECTO</b> .....	<b>6</b>
1.2.1. COBERTURA MEDIÁTICA.....	6
1.2.2. PROBLEMÁTICA.....	7
1.2.3. DELIMITACIÓN DE LA SOLUCIÓN.....	7
<b>1.3. OBJETIVOS</b> .....	<b>7</b>
1.3.1. OBJETIVOS TECNOLÓGICOS.....	8
1.3.2. OBJETIVOS SOCIALES.....	8
1.3.3. OBJETIVOS PERSONALES.....	8
<b>1.4. METODOLOGÍA DE DESARROLLO</b> .....	<b>9</b>
<b>1.5. ESTRUCTURA DE LA MEMORIA</b> .....	<b>9</b>
<b>CAPÍTULO 2: ANÁLISIS</b> .....	<b>11</b>
<b>2.1. ANÁLISIS DE LOS DATOS</b> .....	<b>13</b>
2.1.1. DEFINICIÓN Y ELEMENTOS DE UNA NOTICIA.....	13
➤ Titular.....	13
➤ Epígrafe.....	14
➤ Lead.....	15
➤ Cuerpo de la noticia.....	15
2.1.2. LA EXTENSIÓN DE LA NOTICIA.....	15
2.1.3. LA ESTRUCTURA DE LA NOTICIA WEB.....	16
➤ Título.....	16
➤ Encabezado (lead, blurb).....	16
➤ Cuerpo.....	16
<b>2.2. TÉCNICAS DE EXTRACCIÓN DE INFORMACIÓN WEB</b> .....	<b>17</b>
2.2.1. RASTREO EN LA WEB.....	17
2.2.2. RECURSOS DE INFORMACIÓN.....	18
2.2.2.1. Páginas Web o URL.....	18
2.2.2.2. RSS.....	18
2.2.2.3. Atom.....	18

2.2.3.	HERRAMIENTAS DE RECONOCIMIENTOS DE PATRONES.....	18
2.2.3.1.	<i>Expresiones Regulares</i> .....	18
2.2.3.2.	<i>Hojas de Estilo o CSS</i> .....	19
2.2.3.3.	<i>DOM</i> .....	19
2.2.3.4.	<i>Librerías</i> .....	20
2.2.3.5.	<i>Aplicaciones</i> .....	20
<b>CAPÍTULO 3: DESARROLLO DEL ALGORITMO.....</b>		<b>21</b>
<b>3.1.</b>	<b>INTRODUCCIÓN.....</b>	<b>23</b>
<b>3.2.</b>	<b>FUENTES DE INFORMACIÓN.....</b>	<b>23</b>
3.2.1.	PERIÓDICOS DEPORTIVOS.....	23
➤	Diario AS en su edición digital.....	23
➤	Diario EL MUNDO DEPORTIVO en su edición digital.....	24
➤	Diario MARCA en su edición digital.....	24
3.2.2.	FEDERACIONES DEPORTIVAS.....	24
<b>3.3.</b>	<b>ANÁLISIS GRAMATICAL.....</b>	<b>27</b>
3.3.1.	GRAMÁTICA ESPAÑOLA.....	27
3.3.2.	ANÁLISIS LÉXICO.....	27
3.3.3.	ANÁLISIS SINTÁCTICO.....	31
3.3.4.	ANÁLISIS SEMÁNTICO.....	32
<b>3.4.</b>	<b>DESARROLLO DEL ALGORITMO DE BÚSQUEDA DE NOTICIAS.....</b>	<b>32</b>
3.4.1.	NOTICIAS VÁLIDAS vs. NOTICIAS INVÁLIDAS.....	32
3.4.2.	IMPACTO DE LAS PALABRAS EN LAS NOTICIAS.....	33
3.4.3.	ANÁLISIS SINTÁCTICO DE TÍTULOS DEPORTIVOS.....	36
3.4.4.	LA ELIPSIS.....	36
3.4.5.	VERBOS POSITIVOS Y NEGATIVOS.....	37
3.4.6.	SUJETOS MALICIOSOS.....	37
3.4.7.	REGLAS DEL ALGORITMO.....	38
<b>3.5.</b>	<b>EJEMPLO DE LA EJECUCIÓN DEL ALGORITMO.....</b>	<b>40</b>
<b>3.6.</b>	<b>DISEÑO DE LA BASE DE DATOS.....</b>	<b>41</b>
3.6.1.	MODELO E/R.....	41
3.6.2.	DICCIONARIO DE DATOS.....	42
<b>CAPÍTULO 4: PLATAFORMA WEB.....</b>		<b>43</b>
<b>4.1.</b>	<b>ANÁLISIS DE LA PLATAFORMA WEB.....</b>	<b>45</b>
4.1.1.	ANÁLISIS DE REQUISITOS.....	45
4.1.1.1.	<i>Requisitos de capacidad</i> .....	45
4.1.1.2.	<i>Requisitos de restricción</i> .....	49
4.1.2.	DIAGRAMA DE CASOS DE USO.....	50
4.1.2.1.	<i>Invitado</i> .....	50
4.1.2.2.	<i>Usuario registrado</i> .....	53
4.1.2.3.	<i>Administrador</i> .....	55
<b>4.2.</b>	<b>DISEÑO DE LA PLATAFORMA WEB.....</b>	<b>58</b>
4.2.1.	MÓDULOS PARA MEJORAR LA USABILIDAD DE LA PLATAFORMA WEB.....	58
4.2.1.1.	<i>Multidioma</i> .....	58
4.2.1.2.	<i>Cambio del tamaño de fuente</i> .....	59
4.2.1.3.	<i>Gestor de Eventos</i> .....	59
4.2.1.4.	<i>Login</i> .....	59

4.2.1.5.	<i>Shoutbox o Bandeja de gritos</i> .....	61
4.2.1.6.	<i>Foro</i> .....	62
4.2.1.7.	<i>Estadísticas</i> .....	63
4.2.1.8.	<i>FAQ</i> .....	63
4.2.1.9.	<i>Tienda electrónica</i> .....	64
4.2.1.10.	<i>Mapa del sitio Web</i> .....	65
4.2.2.	PORTADA .....	66
4.2.3.	MENÚ PRINCIPAL (INVITADO).....	67
4.2.4.	MENÚ REGISTRADO .....	68
4.2.5.	MENÚ ESPECIAL/ADMINISTRADOR.....	69
4.2.6.	VISUALIZACIÓN DE LAS NOTICIAS .....	70
4.2.7.	VERSIÓN PARA DISPOSITIVOS MÓVILES .....	72
<b>4.3.</b>	<b>SEGURIDAD EN LA PLATAFORMA .....</b>	<b>73</b>
	<b>CAPÍTULO 5: PLAN DE DESARROLLO .....</b>	<b>79</b>
<b>5.1.</b>	<b>HERRAMIENTAS UTILIZADAS .....</b>	<b>81</b>
5.1.1.	ECLIPSE .....	81
5.1.2.	JAVA.....	81
5.1.2.1.	<i>JDBC</i> .....	81
5.1.3.	MYSQL .....	82
5.1.4.	PHP.....	83
5.1.5.	PHPMYADMIN.....	83
5.1.6.	JOOMLA!.....	84
5.1.7.	WAMP SERVER.....	85
<b>5.2.</b>	<b>ACOPLAMIENTO DE LAS HERRAMIENTAS .....</b>	<b>85</b>
<b>5.2.</b>	<b>CICLO DE VIDA DEL PROYECTO .....</b>	<b>88</b>
<b>5.3.</b>	<b>ESTRUCTURA DE TRABAJO (WBS) .....</b>	<b>88</b>
5.3.1.	I. ANÁLISIS DE REQUISITOS .....	88
5.3.2.	II. DISEÑO .....	89
5.3.3.	III. IMPLEMENTACIÓN Y PRUEBAS UNITARIAS .....	89
5.3.4.	IV. INTEGRACIÓN Y PRUEBAS DEL SISTEMA .....	89
5.3.5.	V. OPERACIONES Y MANTENIMIENTO .....	90
<b>5.4.</b>	<b>ESTRUCTURA DE RECURSOS (RBS).....</b>	<b>90</b>
<b>5.5.</b>	<b>DIAGRAMA DE GANTT .....</b>	<b>92</b>
<b>5.6.</b>	<b>PRESUPUESTO.....</b>	<b>94</b>
	<b>CAPÍTULO 6: CONCLUSIONES.....</b>	<b>95</b>
<b>6.1.</b>	<b>CONCLUSIONES.....</b>	<b>97</b>
6.1.1.	RESULTADOS .....	97
<b>6.2.</b>	<b>LÍNEAS FUTURAS .....</b>	<b>98</b>
	<b>CAPÍTULO 7: BIBLIOGRAFÍA.....</b>	<b>101</b>
<b>7.1.</b>	<b>BIBLIOGRAFIA.....</b>	<b>103</b>
	<b>APÉNDICES.....</b>	<b>105</b>
	<b>APÉNDICE A: CONCLUSIONES DEL SEMINARIO PERMANENTE MUJER Y DEPORTE .</b>	<b>107</b>
	PRESENTACIÓN .....	107

OBJETIVOS.....	107
CONCLUSIONES.....	108
<b>APÉNDICE B: HEROÍNAS FUERA DE FOCO (08.01.2011).....</b>	<b>111</b>
HEROÍNAS FUERA DE FOCO .....	111
<b>APÉNDICE C: GRAMÁTICA ESPAÑOLA.....</b>	<b>115</b>
C.1. LA PALABRA.....	115
C.1.1. El sustantivo.....	115
C.1.2. El pronombre .....	116
C.1.3. El determinante .....	117
C.1.4. El adjetivo .....	117
C.1.5. El adverbio.....	117
C.1.6. El verbo.....	118
C.1.7. Las proposiciones.....	119
C.1.8. Las conjunciones.....	119
C.1.9. Las interjecciones.....	119
C.2. LA PROPOSICIÓN .....	119
C.2.1. Sujeto.....	120
C.2.2. Predicado .....	120
<b>APÉNDICE D: BREVE HISTORIA DE LOS BUSCADORES DE INTERNET.....</b>	<b>121</b>
<b>APÉNDICE E: FUENTES WEB O RSS .....</b>	<b>123</b>
<b>APÉNDICE F: NOTICIA F.C. BARCELONA (22.02.2011).....</b>	<b>125</b>
PALCO DE BARÇA-ZARAGOZA ESTARÁ INTEGRADA POR MUJERES.....	125
<b>APÉNDICE G: CÓDIGO .....</b>	<b>127</b>
G.1. ALGORITMO DE BÚSQUEDA.....	127
G.2. PÁGINA WEB.....	129
G.2.1. Código para la visualización de las noticias.....	129
G.2.2. Código para la traducción de noticias on-line.....	131
G.2.3. Javascript.....	132
<b>APÉNDICE H: LISTADO DE PALABRAS.....</b>	<b>135</b>
H.1. VERBOS.....	135
H.1.1. Verbos positivos .....	135
H.1.2. Verbos negativos .....	136
H.2. LOGROS .....	136

# Índice de figuras

ILUSTRACIÓN 1 PIRÁMIDE INVERTIDA: ESTRUCTURA DE UNA NOTICIA .....	15
ILUSTRACIÓN 2 EJEMPLO DE NOTICIA DEPORTIVA .....	17
ILUSTRACIÓN 3 ÁRBOL DOM DEL HTML .....	20
ILUSTRACIÓN 4 CABECERA PERIÓDICO AS .....	23
ILUSTRACIÓN 5 SECCIÓN NOSOTRAS DEL PERIÓDICO AS .....	23
ILUSTRACIÓN 6 CABECERA DEL PERIÓDICO MUNDO DEPORTIVO .....	24
ILUSTRACIÓN 7 CABECERA DEL PERIÓDICO MARCA .....	24
ILUSTRACIÓN 8 ELIMINACIÓN DE PALABRAS IRRELEVANTES (GOOGLE) .....	28
ILUSTRACIÓN 9 NO ELIMINACIÓN DE PALABRAS IRRELEVANTES (ALTAVISTA) .....	29
ILUSTRACIÓN 10 EJEMPLO DE STEMMING CON EL ALGORITMO DE PORTER .....	30
ILUSTRACIÓN 11 EJEMPLO DE STEMMING EN CASTELLANO .....	30
ILUSTRACIÓN 12 ESTRUCTURA ÁRBOL DEL ANÁLISIS SINTÁCTICO .....	31
ILUSTRACIÓN 13 ALGORITMO DE BÚSQUEDA DEL ANALIZADOR SINTÁCTICO AUTOMÁTICO DE NOTICIAS ...	39
ILUSTRACIÓN 14 EJEMPLO DE NOTICIA VÁLIDA .....	40
ILUSTRACIÓN 15 MODELO E/R .....	42
ILUSTRACIÓN 16 DIAGRAMA DE CASOS DE USO: INVITADO .....	51
ILUSTRACIÓN 17 DIAGRAMA DE CASOS DE USO: REGISTRADO .....	53
ILUSTRACIÓN 18 DIAGRAMA DE CASOS DE USO: ADMINISTRADOR .....	55
ILUSTRACIÓN 19 MULTIDIOMA .....	58
ILUSTRACIÓN 20 CAMBIO DEL TAMAÑO DE FUENTE .....	59
ILUSTRACIÓN 21 ASCENSOR AUTOMÁTICO .....	59
ILUSTRACIÓN 22 CALENDARIO DE EVENTOS .....	59
ILUSTRACIÓN 23 OPCIÓN DEL MENÚ PARA REGISTRARSE .....	60
ILUSTRACIÓN 24 MÓDULO DE REGISTRO .....	60
ILUSTRACIÓN 25 MÓDULO SHOUTBOX .....	62
ILUSTRACIÓN 26 MÓDULO FORO .....	62
ILUSTRACIÓN 27 MÓDULO DE ESTADÍSTICAS .....	63
ILUSTRACIÓN 28 MÓDULO DE FAQ .....	64
ILUSTRACIÓN 29 TIENDA VIRTUAL .....	64
ILUSTRACIÓN 30 MAPA DEL SITIO WEB .....	65
ILUSTRACIÓN 31 BIENVENIDA .....	66
ILUSTRACIÓN 32 MENÚ NOTICIAS .....	67
ILUSTRACIÓN 33 MENÚ OLÍMPICOS .....	67
ILUSTRACIÓN 34 MENÚ PARALÍMPICOS .....	68

## Índice de figuras

ILUSTRACIÓN 35 MENÚ PRINCIPAL USUARIO REGISTRADO.....	68
ILUSTRACIÓN 36 OPCIÓN TU PERFIL .....	68
ILUSTRACIÓN 37 MENÚ PRINCIPAL USUARIO ADMIN .....	69
ILUSTRACIÓN 38 MÓDULO GESTIONAR CALENDARIO .....	70
ILUSTRACIÓN 39 EJEMPLO DE NOTICIAS EN CASTELLANO .....	71
ILUSTRACIÓN 40 ICONOS DE LAS NOTICIAS .....	71
ILUSTRACIÓN 41 EJEMPLO DE NOTICIAS EN CATALÁN.....	72
ILUSTRACIÓN 42 EJEMPLO DE NOTICIAS EN INGLÉS.....	72
ILUSTRACIÓN 43 VERSIÓN PARA PLATAFORMAS MAC .....	73
ILUSTRACIÓN 44 VERSIÓN PARA PLATAFORMAS SMARTPHONE (NO MAC).....	73
ILUSTRACIÓN 45 FORMULARIO DE LOGIN.....	74
ILUSTRACIÓN 46 LOGIN CIFRADO.....	75
ILUSTRACIÓN 47 LOGIN EN CLARO.....	75
ILUSTRACIÓN 48 CAPTCHA EN FORMULARIO DE REGISTRO .....	76
ILUSTRACIÓN 49 USO DE JSECURE.....	77
ILUSTRACIÓN 50 OPCIÓN SSL.....	78
ILUSTRACIÓN 51 ACOPLAMIENTO DE LAS HERRAMIENTAS.....	85
ILUSTRACIÓN 52 DIAGRAMA DE ACTIVIDAD DE UNA PUBLICACIÓN EN JOOMLA! .....	86
ILUSTRACIÓN 53 DIAGRAMA DE ACTIVIDAD DEL MOTOR DE BÚSQUEDA .....	87
ILUSTRACIÓN 54 CICLO DE VIDA .....	88
ILUSTRACIÓN 55 RBS .....	90
ILUSTRACIÓN 56 RECURSOS MATERIALES .....	91
ILUSTRACIÓN 57 ORGANIGRAMA DE RRHH .....	91
ILUSTRACIÓN 58 DIAGRAMA DE GANTT.....	93
ILUSTRACIÓN 59 ESTRUCTURA DE UNA ORACIÓN.....	120
ILUSTRACIÓN 60 EJEMPLO DE PAGERANK .....	122
ILUSTRACIÓN 61 SÍMBOLOS DE ABRIR Y CERRAR NOTICIAS.....	133



# Índice de tablas

TABLA 1 ANÁLISIS DE LAS FEDERACIONES CON DISCIPLINAS OLÍMPICAS DE VERANO .....	25
TABLA 2 FEDERACIONES CON DISCIPLINAS OLÍMPICAS DE INVIERNO .....	26
TABLA 3 FEDERACIONES CON DISCIPLINAS PARALÍMPICAS .....	26
TABLA 4 FEDERACIONES SIN DISCIPLINAS OLÍMPICAS .....	26
TABLA 5 EJEMPLO DE DERIVACIÓN .....	29
TABLA 6 RESULTADO DEL ANÁLISIS DE NOTICIAS VÁLIDAS .....	33
TABLA 7 RESULTADO DEL ANÁLISIS DE NOTICIAS INVÁLIDAS .....	34
TABLA 8 RESUMEN DE REGLAS .....	38
TABLA 9 UR-C-001 .....	45
TABLA 10 UR-C-002 .....	46
TABLA 11 UR-C-003 .....	46
TABLA 12 UR-C-004 .....	46
TABLA 13 UR-C-003 .....	47
TABLA 14 UR-C-006 .....	47
TABLA 15 UR-C-007 .....	47
TABLA 16 UR-C-008 .....	48
TABLA 17 UR-C-009 .....	48
TABLA 18 UR-C-011 .....	48
TABLA 19 UR-R-001 .....	49
TABLA 20 UR-R-002 .....	49
TABLA 21 UR-R-003 .....	49
TABLA 22 UR-R-004 .....	50
TABLA 23 CU-I-001 .....	51
TABLA 24 CU-I-002 .....	52
TABLA 25 CU-I-003 .....	52
TABLA 26 CU-I-004 .....	52
TABLA 27 CU-R-001 .....	53
TABLA 28 CU-R-002 .....	54
TABLA 29 CU-R-003 .....	54
TABLA 30 CU-A-001 .....	55
TABLA 31 CU-A-002 .....	56
TABLA 32 CU-A-003 .....	56
TABLA 33 CU-A-004 .....	56
TABLA 34 CU-A-005 .....	57
TABLA 35 CU-A-006 .....	57
TABLA 36 CU-R-007 .....	58

## Índice de tablas

TABLA 37 LISTA DE TAREAS .....	92
TABLA 38 TABLA DE VERBOS POSITIVOS.....	135
TABLA 39 TABLA DE VERBOS NEGATIVOS.....	136
TABLA 40 TABLA DE LOGROS.....	136

# Glosario

<b>ADO</b> <sup>1</sup>	<i>Asociación de Deportistas Olímpicos</i>
<b>ADOP</b> <sup>2</sup>	<i>Apoyo al Deporte Objetivo Paralímpico</i>
<b>API</b>	<i>Application Programming Interface</i>
<b>BOE</b> <sup>3</sup>	<i>Boletín Oficial del Estado</i>
<b>CMS</b>	<i>Content Management System</i>
<b>COE</b> <sup>4</sup>	<i>Comité Olímpico Español</i>
<b>CSD</b> <sup>5</sup>	<i>Consejo Superior de Deportes</i>
<b>CSS</b> <sup>6</sup>	<i>Cascade Style Sheet</i>
<b>DRAE</b> <sup>7</sup>	<i>Diccionario de la Real Academia de la lengua Española</i>
<b>EURATHLON</b>	<i>Programa europeo de ayuda al deporte.</i>
<b>INEF</b> <sup>8</sup>	<i>Instituto Nacional de Educación Física</i>
<b>GIF</b>	<i>Graphics Interchange Format. Extensión de un fichero de imagen</i>
<b>GNU</b> <sup>9</sup>	<i>Proyecto de software libre</i>

<sup>1</sup> [ADO]<http://www.ado.es>

<sup>2</sup> [http://www.paralimpicos.es/publicacion/8SC\\_PlanAdo/8SC\\_PlanAdo.html](http://www.paralimpicos.es/publicacion/8SC_PlanAdo/8SC_PlanAdo.html)

<sup>3</sup> [BOE]<http://www.boe.es>

<sup>4</sup> [COE]<http://www.coe.es>

<sup>5</sup> [CSD]<http://www.csd.gob.es>

<sup>6</sup> <http://www.w3schools.com/css>

<sup>7</sup> [DRAE01]<http://www.rae.es>

<sup>8</sup> <http://www.inef.upm.es>

<b>GPL</b> <sup>10</sup>	<i>General Public License</i>
<b>HTML</b> <sup>11</sup>	<i>HyperText Marked Language</i>
<b>IDE</b>	<i>Integrate Development Environment</i>
<b>ISO</b> <sup>12</sup>	<i>International Organization for Standardization</i>
<b>JDBC</b>	<i>JAVA DataBase Connectivity</i>
<b>JDT</b>	<i>JAVA Development Toolkit</i>
<b>JPEG</b>	<i>Joint Photographic Experts Group. Extensión de un fichero de imagen</i>
<b>LDAP</b>	<i>Lightweight Directory Protocol. Protocolo ligero de acceso a directories.</i>
<b>MIME</b>	<i>Multipurpose Internet Mail Extensions</i>
<b>PICS</b>	<i>Platform for Internet Content Selection</i>
<b>PNG</b>	<i>Portable Network Graphics. Extensión de un fichero de imagen</i>
<b>RFC</b>	<i>Request For Comments</i>
<b>RSA</b> <sup>13</sup>	<i>Algoritmo de encriptación de clave pública</i>
<b>RSS</b>	<i>Really Simple Syndication</i>
<b>RBS</b>	<i>Resource Breakdown Structure</i>
<b>SEO</b>	<i>Search Engine Optimization</i>
<b>SGML</b>	<i>Standard Generalized Marked Language (ISO 8879:16986)</i>
<b>SQL</b>	<i>Structure Query Language</i>
<b>URL</b> <sup>14</sup>	<i>Uniform Resource Locator</i>
<b>XML</b>	<i>eXtensible Marked Language</i>
<b>W3C</b> <sup>15</sup>	<i>World Wide Web Consortium</i>
<b>WBS</b>	<i>Work Breakdown Structure</i>

<sup>9</sup> <http://www.gnu.org>

<sup>10</sup> <http://www.gnu.org/licenses/gpl.html>

<sup>11</sup> <http://www.w3schools.com/html>

<sup>12</sup> <http://www.iso.org>

<sup>13</sup> <http://www.rsa.com/rsalabs/node.asp?id=2146>

<sup>14</sup> [BER+96] y [FIE+99]

<sup>15</sup> [W3C] <http://www.w3c.es>

# Capítulo 1: Introducción

---

José María Vargas Vila (23/07/1860 - 25/05/1933). Escritor colombiano.

Los que han proclamado la bancarrota de la ciencia, no han hecho sino  
anunciar la derrota de su ignorancia.



## 1.1. INTRODUCCIÓN

Son múltiples las leyes e iniciativas a todos los niveles (tanto a nivel europeo, como estatal o como local), que instan a plantear nuevas formas de pensar y actuar para hacer posible la transformación y el cambio para el logro de una sociedad más igualitaria.

Tanto es así que resultan evidentes los avances sociales de las mujeres en todos los ámbitos, pero a pesar de la aparente igualdad de oportunidades entre hombres y mujeres, existen aún obstáculos que dificultan la igualdad efectiva. El deporte, como actividad plenamente integrada en la sociedad no es ajeno a esta realidad.

A nivel europeo, nació en 1995 el programa **EURATHLON** (96/C222/08) con el objetivo de dotar a las subvenciones comunitarias a favor del deporte, de un marco para establecer criterios objetivos para la concesión de ayudas. Así mismo, se establecieron objetivos específicos para promocionar actividades deportivas, entre otros, el “apoyo de actividades deportivas de finalidad social,... que promuevan la igualdad de mujer y hombres.”

En el año 2000, la Comisión Europea presentó la **Declaración de Niza sobre el deporte**, en cuyo apartado número 8, también se hace referencia a la importancia de la promoción del acceso de hombres y de mujeres a la práctica deportiva de todos los niveles.

En 2004, la Asamblea General de las Naciones Unidas proclamó el año 2005 como **Año Internacional del deporte y de la educación física**, siendo uno de los objetivos establecidos para el “Desarrollo del Milenio” *promover la igualdad de género y el empoderamiento de las mujeres*.

Y en 2007, la Comisión Europea participó en el **Libro Blanco sobre el Deporte**<sup>16</sup> que enfatiza en la importancia social y económica del deporte, y establece que, *“en el marco de su Plan de Trabajo para la Igualdad entre las Mujeres y los hombres 2006-2010, la Comisión impulsará la perspectiva de género en todas las actividades relacionadas con el deporte...”*.

Por otra parte, a nivel nacional, en 1988 tras los Juegos Olímpicos de Seúl, y con vistas a la celebración por primera vez en España de unos Juegos Olímpicos de verano de 1992 en Barcelona, se pone en marcha el **Plan ADO** por parte de la Asociación de Deportistas Olímpicos para apoyar la promoción y desarrollo de los deportistas nacionales de alto rendimiento en deportes olímpicos. Su objetivo principal era brindar a los deportistas medios y recursos necesarios para conseguir una gran actuación en los Juegos Olímpicos de Barcelona'92. Desde entonces, el deporte español ha ido sufriendo una metamorfosis en relación a los éxitos conseguidos; si hasta 1988 tan sólo se habían conseguido 27 medallas (6 de ellas de oro) en 16 Juegos Olímpicos de Verano y 12 Juegos Olímpicos de Invierno, desde la implantación del Plan ADO se ha triplicado la consecución de las medallas (88 medallas con 29 de oro) sin contar los éxitos conseguidos en las diversas competiciones internacionales, mundiales y europeos.

Además, desde el año 2004 los Juegos Paralímpicos también tiene sus propias becas con el plan **ADOP**.

---

<sup>16</sup> [http://ec.europa.eu/sport/club/pdf/whitepaper-short\\_es.pdf](http://ec.europa.eu/sport/club/pdf/whitepaper-short_es.pdf)

## INTRODUCCIÓN

Con estos dos planes de becas se consigue fomentar el deporte español a alto nivel. Pero observando los resultados, se puede comprobar la tendencia del deporte femenino en los éxitos olímpicos: en Seúl'88 no consiguió ninguna medalla, desde entonces con el plan ADO las medallas en Juegos Olímpicos para el deporte femenino suponen un 30 – 40% de las medallas totales.<sup>17</sup>

Además, desde las políticas de impulso del papel de la mujer en la sociedad, han ido surgiendo los sucesivos *Planes de Igualdad* que desde el año 1990 promueven la incorporación de la mujer a todos los ámbitos donde su presencia no era equitativa, así como la prevención y lucha contra la desigualdad entre mujeres y hombres, entre ellos el DEPORTE. Cabe destacar el **IV Plan de Igual de Oportunidades entre Mujeres y Hombres 2003-2006**, que favoreció la aparición de la **ORDEN PRE/525/2005, de 7 de marzo**, por la que se da publicidad al Acuerdo de Consejo de Ministros por el que se aportan medidas para favorecer la igualdad entre mujeres y hombres publicada en el BOE el 8 de marzo de 2005. En concreto el Artículo 6 Deportes dice:

"6.1 Se acuerda crear, dentro del Consejo Superior de Deportes, una unidad con carácter permanente para desarrollar el programa <<Mujer y Deporte>>."

"6.2 Igualmente, a través del Consejo Superior de Deportes se subvencionará a la Comisión Mujer y Deporte del Comité Olímpico Español."

"6.3 Se acuerda la firma de un convenio entre el Consejo Superior de Deportes y el Instituto de la Mujer para promover el deporte femenino."

"6.4 Se encomienda al Consejo Superior de Deportes que en todas las campañas o exposiciones para el fomento del deporte se aplique el principio de paridad."

"6.5 Se acuerda que el Consejo Superior de Deportes promueva estudios de investigación, análisis y estadísticas dirigidos a fomentar la igualdad de género en el deporte."

En este contexto y bajo el marco de la ORDEN PRE/525/2005, de 7 de marzo se crean los **Programas Mujer y Deporte**, en el seno del Consejo Superior de Deportes, en la Subdirección General de Promoción Deportiva y Deporte Paralímpico.

"El principal objetivo de estos programas es fomentar la participación de las mujeres en el ámbito de la actividad física y el deporte en igualdad de condiciones que los hombres, superando las barreras u obstáculos que aún hoy día dificultan la realidad de una práctica con equidad de género en este ámbito de la vida.

---

<sup>17</sup> Datos obtenidos del COE.



El fomento de la participación de las mujeres ha de ser para todas las edades, modalidades deportivas, grados de implicación/competición y niveles profesionales (puestos técnicos, de gestión, administración y dirección).

Para poder abarcar este objetivo tan amplio es indispensable la coordinación con múltiples Organismos (Instituto de la Mujer, Federaciones Deportivas, Universidades, Comunidades Autónomas, Comisión Mujer y Deporte del COE, Ayuntamientos, empresas, etc...).

Como principales líneas de actuación cabe destacar la convocatoria anual de Subvenciones para Programas específicos de la Mujer y Deporte (seguimiento y soporte), el impulso de eventos deportivos en las Comunidades Autónomas, la firma de Convenios de Colaboración con otros Organismo, la organización y participación en Congresos, Seminarios, cursos, así como el planteamiento y realización de estudios de investigación."

Años más tarde, se da un paso más con la **Ley Orgánica 3/2007, de 22 de marzo, para la Igualdad Efectiva de Mujeres y Hombres**; así como el **Plan Estratégico de Igualdad de Oportunidades 2008-2011** en abril de 2008 promovido por el, recién creado por aquél entonces, **Ministerio de Igualdad** presidido en un primer momento por D.<sup>a</sup> Bibiana Aído. Más en concreto, el Artículo 29 Deportes dice:

"Todos los programas públicos de desarrollo del deporte incorporarán la efectiva consideración del principio de igualdad real y efectiva entre mujeres y hombres en su diseño y ejecución.

El Gobierno promoverá el deporte femenino y favorecerá la efectiva apertura de las disciplinas deportivas a las mujeres, mediante el desarrollo de programas específicos en todas las etapas de la vida y en todos los niveles, incluidos los de responsabilidad y decisión."

Con todos estos cambios se aprecia cómo han ido aumentando la promoción del deporte femenino con jornadas exclusivamente para ello, como puede ser *el triatlón de la mujer de la Real Federación Española de Triatlón*<sup>18</sup>, *la carrera de la mujer*<sup>19</sup> o portales web exclusivos para ellas como puede ser *Canal Mujer y Agua de la Real Federación Española de Natación*<sup>20</sup> o

---

<sup>18</sup> <http://www.eltriatlondelamujer.org>

<sup>19</sup> <http://www.carreradelamujer.com>

<sup>20</sup> [http://www.rfen.es/publicacion/18SC\\_Canal\\_mujer/143SS\\_Inicio.asp](http://www.rfen.es/publicacion/18SC_Canal_mujer/143SS_Inicio.asp)

## DESCRIPCIÓN DEL ÁMBITO DEL PROYECTO

mediante la creación de comisiones como la *Real Federación Española de Tenis de Mesa*<sup>21</sup> creando jornadas específicas para dar a conocer su deporte.

Todas estas iniciativas requieren una difusión que no se encuentra en los medios de televisión, es por ello, la necesidad del uso de una herramienta que poco a poco se ha ido haciendo básica en esta sociedad, Internet; una herramienta muy útil que llega a miles de casas y de mujeres. En este sentido, el proyecto pretende ser una puerta que acerque el deporte femenino a la gente mediante un portal web con noticias de deportistas españolas femeninas demostrando así que los logros del deporte español no son sólo masculinos sino que el deporte femenino logra los mismos éxitos en el deporte.

## 1.2. DESCRIPCIÓN DEL ÁMBITO DEL PROYECTO

### 1.2.1. COBERTURA MEDIÁTICA

Si se realiza un estudio de la cobertura o atención mediática que recibe el deporte femenino por parte de la prensa se observa que es casi nulo; basta con comprobar los grandes periódicos deportivos nacionales y se aprecia como tan sólo uno de ellos dispone una sección dedicada al mundo del deporte femenino a nivel internacional, por el contrario el resto de ellos le dedican pequeñas secciones en algún deporte concreto tipo fútbol o baloncesto, o si no, se ve relegado a las secciones “Otros deportes” o “Más deporte”.

Tal y como se recoge en el Seminario permanente “Mujer y Deporte” celebrado en la facultad de Ciencias de la Actividad Física y del Deporte (INEF) de la Universidad Politécnica de Madrid en las jornadas sobre “*El deporte femenino en los medios de comunicación*” durante la semana del 5 al 11 de noviembre de 2009 [véase apéndice A], el deporte en España es absorbido por completo por el deporte masculino y más concretamente por el fútbol dejando de lado el resto de deportes y disciplinas, cosa que se refleja en los medios de comunicación.

En lo referente a la difusión en los medios de comunicación electrónicos, existen algunos portales web referentes al deporte femenino, sirvan como ejemplo:

- [www.mujoydeporte.org](http://www.mujoydeporte.org): portal promovido por el COE y su comisión de Mujer y Deporte siguiendo todo el panorama nacional.
- [www.deportefemeninotambien.com](http://www.deportefemeninotambien.com): asociación que promueve el deporte femenino en Madrid en todas las categorías apoyada por las distintas federaciones deportivas madrileñas.
- [www.amecopress.com](http://www.amecopress.com): periódico promovido por la Asociación Española de Mujeres Profesionales en los Medios de Comunicación que cubre todo tipo de noticias relacionadas con el mundo de la mujer.

---

<sup>21</sup> [http://www.rfetm.com/?pagina=ver.php&id\\_noticia=840&id\\_subcategoria2=](http://www.rfetm.com/?pagina=ver.php&id_noticia=840&id_subcategoria2=)

- [www.solodeportefemenino.com](http://www.solodeportefemenino.com): programa SDF de radio digital creada por la Universidad Europea de Madrid que da cobertura a todo el deporte femenino apoyado por el CSD lleva en antena desde el 17 de septiembre de 2009.

Un ejemplo de esta poca repercusión es el artículo de Raúl Cosín del pasado 8 de enero titulado *“Heroínas fuera de foco”*<sup>22</sup> [véase Apéndice B] para el diario ABC en el dónde relata la situación del baloncesto español aprovechando la celebración de la Copa de la Reina 2011 en Valencia tan solo tres meses después de lograr la mayor hazaña del baloncesto femenino en nuestro país, una medalla de bronce en un mundial de baloncesto.

### 1.2.2. PROBLEMÁTICA

La poca atención recibida es fruto del poco interés que despierta el deporte femenino en el público en general, y en consecuencia genera menos difusión y mediatización, provocando que sea mucho menos rentable. En particular, mantener un portal web especializado requiere, además de mucho trabajo y esfuerzo, una gran inversión de capital económico y patrocinadores que se unan al proyecto.

Además, no es solo a nivel de los medios de comunicación, sino que los datos de estudios más recientes reflejan que todavía existen grandes diferencias en cuanto a la participación y representación femenina en las distintas esferas del deporte. Es, por tanto, primordial promover la equidad en el deporte entre hombres y mujeres, y fomentar a las mujeres para que disfruten del deporte y de los beneficios que éste conlleva.

Existe pues la necesidad de, por un lado, desarrollar iniciativas que fomenten, promuevan y acerquen el deporte femenino a nivel nacional a todo el público en general y, por otro lado, que estas iniciativas sean de bajo coste, minimizando así el coste económico y posibilitando la viabilidad de estos proyectos.

### 1.2.3. DELIMITACIÓN DE LA SOLUCIÓN

Ante esta problemática, uno de los retos del proyecto es minimizar ese alto capital económico y, por ello, la gran característica del portal web que se desarrollará será la retroalimentación de todas aquellas noticias susceptibles de fomentar el deporte femenino de una manera automática.

Para conseguir esta retroalimentación, se pretende crear una tarea de ejecución automática que busque esas noticias mediante un algoritmo propio de selección.

## 1.3. OBJETIVOS

En este proyecto pueden dividirse los objetivos en distintos ámbitos, objetivos tecnológicos, objetivos sociales y objetivos personales.

---

<sup>22</sup> <http://www.abc.es/20110108/deportes/abcp-reunen-valencia-20110108.html>

### 1.3.1. OBJETIVOS TECNOLÓGICOS

- a) El objetivo principal del proyecto es el diseño e implementación de una metodología propia para la selección de las distintas noticias publicadas en Internet y que hagan referencia a la consecución de éxitos por parte de las deportistas españolas. Para ello, se utilizará un algoritmo de búsqueda y se creará un analizador semántico o de búsqueda intencionada que filtre las noticias de interés.

En particular, **se desarrollará el primer analizador semántico automático de noticias que contenga logros de deportistas femeninas españolas en disciplinas olímpicas.**

- b) Además, se pretende que ésta metodología propia de selección de noticias permita que el algoritmo sea flexible, esto es, se puedan parametrizar los tipos y criterios de búsqueda y análisis de forma que pueda ser aplicado este algoritmo en otros campos y/o materias.
- c) Para poder visualizar los resultados obtenidos por parte del motor de búsqueda, se desarrollará un demostrador sobre una plataforma web, de modo que el algoritmo sea el motor fundamental que sostenga de manera automática todo el contenido de dicha plataforma web.

### 1.3.2. OBJETIVOS SOCIALES

- a) El principal objetivo social es, aprovechando el demostrador web creado para la visualización de los resultados, fomentar el deporte femenino en la sociedad, de modo, que la plataforma web creada de a conocer los logros y hazañas de las deportistas españolas en los deportes olímpicos a nivel, no solo nacional sino también internacional.
- b) La característica principal de este portal web, cómo se ha descrito en los objetivos tecnológicos, será la alimentación automática de contenidos evitando así, de esta manera, la existencia de la figura de un *webmaster* al uso y reduciendo costes de mantenimiento, gestión y operación.
- c) De este modo, y siguiendo las recomendaciones de la *Guía para la incorporación de la perspectiva de género a la gestión deportiva local de 2011*<sup>23</sup>, se motivará la implantación de este tipo de herramientas en portales de pocos recursos económicos (ayuntamientos y otros organismos de género).
- d) Para poder crear un vínculo entre los visitantes de la página web y la propia página, se pretende que la página web sea una plataforma dinámica que dé cabida a la participación de los usuarios a través de comentarios en la propia página.

### 1.3.3. OBJETIVOS PERSONALES

- a) A nivel personal, el alumno realizador de este proyecto pretende aprender nuevas tecnologías (tecnologías no aprendidas durante la realización de la carrera) que

---

<sup>23</sup> [GUI+11]

supongan un reto y a la vez un desarrollo personal y profesional de cara a su futuro más inmediato.

#### 1.4. METODOLOGÍA DE DESARROLLO

Este proyecto se plantea como un proyecto de investigación en el que se abarca el desarrollo de un nuevo algoritmo de búsqueda intencionada o semántica de noticias, así como el desarrollo de un demostrador que sirva de prueba de concepto. El primer acometido se ha desarrollado partiendo de un algoritmo sencillo de análisis sintáctico y léxico sobre titulares de noticias deportivas. En etapas posteriores se añadieron las caracterizaciones necesarias para proporcionar intencionalidad a la búsqueda dotando al algoritmo de la capacidad de discernir entre “logro” y “fracaso”.

Para el segundo acometido, se ha optado por un ciclo de vida en cascada. Se ha elaborado una lista básica de requisitos que han permitido el desarrollo de un demostrador software sencillo, basado en un gestor de contenidos estándar. Aunque el demostrador cuenta con algunas funcionalidades añadidas, no esenciales para la prueba de concepto, éstas forman parte de las herramientas integradas en el gestor de contenidos y no se han incluido como parte del estudio.

#### 1.5. ESTRUCTURA DE LA MEMORIA

El resto de la memoria sigue la estructura que se indica a continuación. Los capítulos 2 y 3 se centran en el trabajo de investigación y desarrollo de un analizador semántico de noticias. En el capítulo 4 se detalla el desarrollo del demostrador software o prueba de concepto. Por último, los capítulos 5, 6 y 7 están dedicados a describir el presupuesto, las conclusiones y la bibliografía, respectivamente.

- **Capítulo 2:** Se realiza un análisis del estado del arte relacionado con el proyecto, análisis de los datos y técnicas de extracción.
- **Capítulo 3:** Explica el desarrollo del algoritmo de búsqueda implementado para la extracción de la información.
- **Capítulo 4:** Explica el desarrollo de la plataforma web que mostrará la información recogida a través del algoritmo de búsqueda.
- **Capítulo 5:** Explica del plan de desarrollo llevado a cabo para la realización del proyecto, tales como recursos utilizados, organización de tareas y presupuestos.
- **Capítulo 6:** Conclusiones extraídas de la realización de este proyecto, a la vez que se indican las posibles líneas de trabajo que se pueden llevar a cabo en un futuro.
- **Capítulo 7:** Relación detallada de toda la bibliografía consultada para la ejecución del proyecto.
- **Capítulo 8:** Se recogen todos aquellos documentos en forma de apéndice que aportan información al proyecto.



# Capítulo 2: Análisis

---

Max Karl Ernest Ludwig Planck (23/04/1858 – 04/10/1947). Físico alemán.

Los descubrimientos científicos y el conocimiento, han sido realizados únicamente por aquellos que se han consagrado a ello, sin perseguir ninguna finalidad práctica.





## 2.1. ANÁLISIS DE LOS DATOS

### 2.1.1. DEFINICIÓN Y ELEMENTOS DE UNA NOTICIA

Una noticia es el relato o redacción (cada cual con sus propias reglas de construcción y enunciación) que refiere a un hecho novedoso o atípico –o la relación entre hechos novedosos y/o atípicos-, ocurrido dentro de una comunidad o un determinado ámbito específico, y que por estas condiciones de novedad y poco ordinario hace que merezca ser conocido y/o divulgado.<sup>24</sup>

La estructura de la noticia impresa suele presentar alguno de estos elementos:

#### ➤ *Titular*

Es un texto muy breve, claro y preciso que recoge una síntesis (generalmente en una o dos líneas) de lo que se informa posteriormente. Es el elemento más visible y debe referirse al contenido (por connotación o denotación).

En sus orígenes el texto noticioso ocupaba toda la página del periódico sin encabezamientos y los titulares se limitaban a servir de separación entre las diferentes noticias. Con la llegada del periodismo informativo (2ª mitad del S. XIX) los periódicos empezaron a ordenar sus contenidos y a presentarlos de manera más atractiva, diferenciando las diversas noticias e introduciendo titulares más complejos y llamativos. Éstos se convertirán en el reclamo para atraer al público e interesar a los lectores.

El titular es muy importante, porque a veces es lo único que se alcanza a leer y en muchas ocasiones es lo único que se recuerda de una noticia aunque se haya leído en su totalidad. Todo titular debe cumplir tres funciones: ser atractivo (llamar la atención de lector), informativo (dar cuenta del contenido de la noticia) y ser objetivo (exponer el contenido de la noticia) o subjetivo (exponer la opinión del autor o un aspecto segmentado de la noticia).

Elementos del titular:

- *Cintillo*: sirve para vincular distintas informaciones que se relacionan temáticamente. Orienta al lector en la tarea de la lectura. Ejemplo: Deportes, Internacional, Sociedad, Cultura...
- *Antetítulo*: precede al título y complementa aspectos informativos de la noticia que no aparecen en el titular. Se escribe en un cuerpo de letra menor que el título y con un tipo de letra diferente. También llamado *Volanta*.
- *Título*: es el elemento fundamental del encabezamiento. Resume la entradilla o primer párrafo de la noticia.
- *Subtítulo*: amplía algunos detalles fundamentales apuntados en el título o en el antetítulo. También llamado *Copete* o *Bajada*.
- *Ladillo*: es un pequeño título que se coloca dentro de la columna de texto y que aparece justificado a un lado. Se coloca para separar los distintos párrafos de la

---

<sup>24</sup> [ARS+04]

## ANÁLISIS DE LOS DATOS

noticia. Suele ser bastante corto y no debe repetir palabras que se hayan utilizado antes en el título, antetítulo o subtítulo.

- *Sumario*: titulares que pretenden llamar la atención sobre aspectos del cuerpo de la noticia que no se incluyen en el encabezamiento. Son muy utilizados en revistas gráficas y de información general.

Existen tres tipos de titulares:

- *Informativos*: identifican la acción y al protagonista.
- *Expresivos*: no persiguen íntegramente informar sobre un hecho, sino que intentan impactar a los lectores. Suelen de una palabra, aparecen primera página y son muy frecuentes en la prensa deportiva.
- *Apelativos*: utilizan el lenguaje para llamar la atención sobre un hecho del que no se informa en profundidad. Son más de propios de prensa sensacionalista y de sucesos.

Por herencia, en función de su sistema jerárquico también se clasifican en:

- *De intervalo abierto*: son los titulares en los cuales únicamente está presente el título.
- *De intervalo abierto a la derecha*: son aquellos titulares en los que solo aparece el título pero incluyen una especie de guía llamados señaladores deícticos (informan del tiempo, espacio y persona).
- *De intervalo abierto a la izquierda*: son aquellos titulares insertados dentro de la noticia que pretenden relajar la lectura y llamar la atención sobre un punto concreto de la noticia. Son conocidos también como intertítulos.
- *Con continuidad*: son aquellos titulares en los cuales se distingue claramente el título y el lead.
- *Con semicontinuidad inferior*: aquellos titulares que tienen subtítulo.
- *Con semicontinuidad superior*: aquellos titulares que tienen antetítulo.

Otras categorías:

- *Titulares temáticos*: mencionan genéricamente el tema sobre el que trata la noticia. Son titulares informativos, pero sólo tratan un elemento de la noticia sin aportar datos complementarios.
- *Titulares de actos de la palabra*: están basados en declaraciones (tanto orales como escritas) de personajes de actualidad. Pueden ser de tres tipos:
  - *Titulares con cita textual*: reproducen literalmente una declaración.
  - *Titulares en forma indirecta*: recogen las declaraciones sintetizándolas, reelaborándolas y condensándolas para transmitir la idea que tiene el protagonista.
  - *Titulares mixtos*: utilizan citas directas e indirectas. El periodista no utiliza la cita completa, pero sí algunas palabras.

### ➤ *Epígrafe*

Suele estar ubicado en la página siguiente a la noticia y anterior al prólogo. Ofrece información sobre las fotografías y/o infografías y/o gráficos; por lo que sólo aparece si existe material fotográfico en la noticia.

➤ **Lead**

El lead (del inglés ‘liderar’ o ‘conducir’) es el núcleo esencial de la noticia. No es un resumen sino una síntesis, con una gran economía de palabras. Su función es dar cuenta inmediata y clara de lo fundamental de ella, e interesar y captar la atención para que el público siga leyendo, viendo o escuchando. Suele corresponder con el primer párrafo de la noticia.

➤ **Cuerpo de la noticia**

Se da la información completa. La información va de mayor a menor importancia.

Según Melvin Mencher en su libro *“News Reporting and Writing”*<sup>25</sup> identificaba estos elementos anteriormente descritos y definía una *“Pirámide invertida”* similar a la que se representa en la siguiente ilustración:



Ilustración 1 Pirámide invertida: Estructura de una noticia

Esta *“Pirámide invertida”* constituye una referencia a la hora de redactar los artículos periodísticos.

### 2.1.2. LA EXTENSIÓN DE LA NOTICIA

La extensión de la noticia y de todos los demás textos publicados en línea no puede ser demasiado largos, ya que los lectores en Internet rechazan instintivamente textos demasiado extensos.

---

<sup>25</sup> [MEN+02]

Los artículos pensados para la edición en papel, cuando se trasladan sin adaptación al ordenador, resultan, por lo general, demasiado largos y obligan al lector a la engorrosa tarea de avanzar en el texto a lo largo de varias pantallas.

Leer en la pantalla no es lo mismo que leer en papel. La lectura en el computador cansa. Es por esta razón que la estructura de la noticia para Internet debe ser de los más concisa, concreta y organizada posible, para permitirle al lector el llamado "escaneo", es decir, el conocimiento, a través de un vistazo, de los detalles sustanciales de la información.

### 2.1.3. LA ESTRUCTURA DE LA NOTICIA WEB

La correcta presentación de una noticia en Internet está relacionada con la técnica de la pirámide invertida, es decir, un tipo de redacción que organiza las ideas en una secuencia completamente opuesta a la utilizada tradicionalmente en otros medios como son los artículos científicos y académicos, que utiliza el esquema de la pirámide invertida justo al contrario; sirva de ejemplo este mismo proyecto o cualquier artículo de una revista científica como puede ser *Nature*<sup>26</sup> o *Science*<sup>27</sup>.

#### ➤ *Título*

Cada noticia consta del título que debe ser el resumen más exacto y fiel, pero además suficientemente llamativo del texto que sigue a continuación para atrapar la atención del usuario.

#### ➤ *Encabezado (lead, blurb)*

Es un párrafo inicial que guía al lector en el conocimiento del hecho y la puerta de entrada a la noticia. El encabezado contiene la esencia de la información, de manera que cualquier lector pueda tener la noción cabal de lo ocurrido con solo leer dicho párrafo. En este párrafo, no hay que tratar de incluir las respuestas a las seis preguntas: **¿qué?, ¿quién?, ¿cuándo?, ¿cómo?, ¿dónde? y ¿por qué?**, ya que es demasiada información para tan poco espacio. Según las reglas del periodismo tradicional, la respuesta a las seis preguntas debe estar incluida en los tres primeros párrafos del texto.

#### ➤ *Cuerpo*

Constituye un desarrollo racional y coherente de lo propuesto en el encabezado, generalmente siguiendo el orden de importancia de los datos (la pirámide invertida): primero lo más relevante y luego lo de menos interés. Para facilitar a los usuarios la lectura en la pantalla, en las noticias largas es indispensable utilizar los subtítulos. Debido a que ningún lector posee la información completa acerca de todos los detalles de un acontecimiento, en caso de que alguna noticia esté relacionada con alguna otra, es necesario recordarle siempre al lector los datos básicos de lo ocurrido antes para darle continuidad al tema o utilizar los vínculos que constituyen una de las características principales y ventajosas de la escritura para la Web. De esta forma, se puede ampliar la información sin necesidad de ser repetitivos.

---

<sup>26</sup> Revista científica sobre biología, <http://www.nature.com>

<sup>27</sup> Revista científica sobre ciencia, <http://www.sciencemag.org>

NOTICIAS

Lunes, 23 de Mayo, 2011 Femenino

TÍTULO

**Marta Silva, nombrada 'Jugadora Universitaria del Año' en USA**

Reconocimiento de la Asociación Nacional de Entrenadores de Golf de Estados Unidos

SUBTÍTULO



Marta Silva

Marta Silva ha sido nombrada 'Jugadora Universitaria del Año' en Estados Unidos por la Asociación Nacional de Entrenadores de Golf de ese país, un reconocimiento que le llega a la jugadora española pocas horas después de conseguir la cuarta plaza en la Final Universitaria USA celebrada en el Traditions Golf Club (Texas), ganada por Austin Ernst, de la Universidad de Louisiana State.

LEAD

"Constituye un gran honor, por supuesto. Estoy muy orgullosa porque este premio muestra todo lo que he realizado a lo largo del año, que ha sido realmente bueno", ha manifestado la gallega Marta Silva, quien asimismo ha sido incluida en el primer equipo All America 2011 y ganado la Golfstat Cup por ser la golfista universitaria con mejor media de golpes de todo Estados Unidos en la presente temporada.

Marta Silva ascendió a mitad de año al número 1 del Ranking universitario estadounidense tras imponerse con brillantez en el Liz Murphey Collegiate Classic, torneo celebrado en la Universidad de Georgia, donde milita desde la temporada 2008-2009.

CUERPO

Marta Silva se convirtió con ello en la primera golfista española en alcanzar este honor en el marco de una sobresaliente trayectoria que incluye en el pasado reciente otras cuatro victorias, todas ellas en 2009: UNLV Spring Rebel Invitational, SEC Women's Championship, Cougar Classic Crowns y Eat A Peach Collegiate.

Su regularidad –71.51 golpes de media– se ha visto plasmada por unos resultados que le han situado seis veces entre las cinco primeras en los diez torneos disputados esta temporada.

Con anterioridad a su acceso a la Universidad de Georgia, asimismo en Estados Unidos pero en su etapa en el instituto, la jugadora compostelana ganó también otros cuatro torneos que se acumulan a un palmarés ilustrado además, entre otras cosas, por el Campeonato de España Cadete 2005, los Campeonatos de España Absoluto y Junior 2010 y los subcampeonatos de Europa Absoluto Individual y por Equipos 2010.

Marta Silva retorna ahora a España para disputar el Campeonato de España Individual Femenino en Pedreña del 26 al 29 de mayo, donde defenderá el título conquistado el año pasado en La Manga.

Ilustración 2 Ejemplo de noticia deportiva

## 2.2. TÉCNICAS DE EXTRACCIÓN DE INFORMACIÓN WEB

En este apartado se recogen distintas técnicas y/o herramientas que se pueden utilizar para obtener información relativa a las noticias desde Internet. Primeramente, se describirán técnicas básicas y generales que se pueden emplear y a continuación se mostrarán otras aplicaciones y/o librerías concretas que consiguen el mismo fin.

### 2.2.1. RASTREO EN LA WEB

El proceso de recuperación de información, en inglés *Web scraping* o *Web crawling* (también llamado *Web harvesting* o *Web data extraction*), consiste en un software denominado *bot* que se ejecuta *a modo de demonio* simulando una exploración del contenido de las páginas web por parte de un usuario.

Este *bot*, también llamado *crawler* o *spider*, mediante diversas técnicas extrae información de interés para ser primeramente almacenada y más tarde analizada. De este modo, con un simple clic se es capaz de indizar una red de páginas enteras.

Sirva de ejemplo, quizás el *bot* más famoso a día de hoy en Internet, el *Googlebot*<sup>28</sup> de Google, que se encarga de determinar los sitios que debe rastrear, la frecuencia con la que debe hacerlo y las páginas a indexar en cada sitio. Cabe destacar, que hace un año

<sup>28</sup> <http://www.google.com/support/webmasters/bin/answer.py?hl=es&answer=182072>

apareció una nueva versión de este *bot* denominado *Google Cafféine*<sup>29</sup> que mejora la velocidad de búsqueda de su predecesor.

### 2.2.2. RECURSOS DE INFORMACIÓN

A continuación, se muestran diversos formatos de información.

#### 2.2.2.1. Páginas Web o URL

Los *bot* anteriormente mencionados están basados en recorrer listas de páginas web, es decir, listas de *URL (Uniform Resource Locator)*. Las *URL* básicamente son código *HTML (HyperText Marked Language)* que es un lenguaje de marcado utilizado para la elaboración de páginas web. Este lenguaje es una extensión del *SGML (Standard Generalized Marked Language)* declarado en la ISO 8879.

#### 2.2.2.2. RSS

Las *RSS*<sup>30</sup> [véase Apéndice E] es la forma más común de redifusión de contenido web, es un formato *XML (eXtensible Marked Language)* que también es una extensión del *SGML*.

#### 2.2.2.3. Atom

Es otra forma de redifusión de contenido web similar a los *RSS* que también usa el formato *XML*.

De este modo, se puede apreciar que existen diversas fuentes o recursos de información en Internet pero que todos se basan en un lenguaje de marcado o etiquetas, esto significa, que manipulando los *bot* existentes para las *URL* se pueden crear *bot* para las *RSS* o los *Atom*, pues lo único que cambia es el nombre del *tag* o etiqueta.

### 2.2.3. HERRAMIENTAS DE RECONOCIMIENTOS DE PATRONES

#### 2.2.3.1. Expresiones Regulares

Las expresiones regulares<sup>31</sup> son patrones que describen una cadena utilizando una determinada sintaxis y que todos los lenguajes de programación incorporan mecanismos para su representación y manejo.

El uso de las expresiones regulares requiere de un estudio previo del origen a examinar, en el caso del lenguaje de marcado es necesario estudiar qué *tag* son susceptibles de interés para elaborar una expresión regular que capture la información.

Un ejemplo puede verse en el siguiente código, la validación de una dirección de correo electrónico en PHP.

```
$string = "richard.vega@alumnos.uc3m.es";  
if (preg_match(
```

---

<sup>29</sup> <http://googleblog.blogspot.com/2010/06/our-new-search-index-caffeine.html>

<sup>30</sup> [RSS]

<sup>31</sup> [FIE+06], [AHO+08] y [SAN+88]

```

'/^[^0-9][a-zA-Z0-9_]+(.[a-zA-Z0-9_]+)*[@][a-zA-Z0-9_]+(.[a-zA-Z0-9_]+)*[.][a-zA-Z]{2,4}$/' , $string) {
    echo "Mail correcto";
}

```

A la hora de extraer información, el principal inconveniente que tiene es la posibilidad de existir errores en la formación de las fuentes de información debido a etiquetas mal formadas, por lo que a pesar de ser una herramienta muy potente de reconocimiento de patrones puede resultar difícil de ajustar debido a estos errores.

### 2.2.3.2. Hojas de Estilo o CSS

Las *hojas de estilo* en cascada o *CSS* es el lenguaje usado para definir la presentación de un documento estructurado escrito en un lenguaje de marcado. El *W3C* es el encargado de mantener el estándar de *CSS*.

El *CSS* es un fichero plano que define la presentación de zonas, elementos o etiquetas de un lenguaje de marcado que son interpretadas por el navegador. De este modo, el programador encapsula su página web acercándose al modelo vista-controlador separando la vista de la página del contenido.

El siguiente ejemplo crea una clase llamada *button* en un *CSS*, que define tres características el color de fondo, el color del borde y el tipo de letra.

```

.button{
    background:#F762C7;
    font-weight:bold;
    border-color:#C43094;
}

```

Ahora, desde *HTML* se crea un *input* y que pertenezca a la clase *button* que se quiere usar, de modo que a la hora de representarse en un navegador web este *input* cogerá los valores especificados en la clase *button* definida en el *CSS*.

```

<input class="button" name="search" type="button"
value="Search"/>

```

De esta forma, es posible obtener información de la página web pero es necesario que el *CSS* sea claro, conciso y homogéneo, y que además el *HTML* esté bien formado.

### 2.2.3.3. DOM

El *DOM* o Modelo de Objetos del Documento es una interfaz de programación que proporciona un estándar de objetos para representar documentos de lenguajes de etiquetado como *HTML* o *XML*. De este modo, los programas pueden crear, acceder y/o modificar el contenido, la estructura y el estilo de los documentos de una manera mucho más cómoda y sencilla.

Si se toma como ejemplo el código *HTML*, el *DOM* lo que busca es crear una estructura árbol que quedaría del siguiente modo:



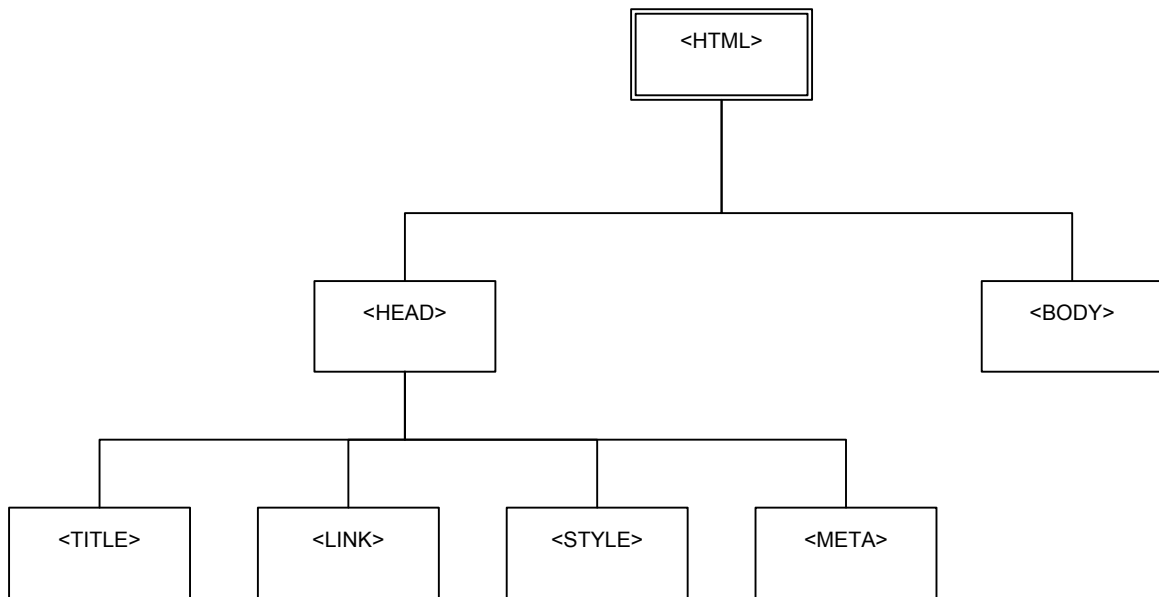


Ilustración 3 Árbol DOM del HTML

De este modo el acceso a todas las partes del código *HTML* se hace navegando a través de una simple estructura de tipo árbol.

#### 2.2.3.4. Librerías

En los lenguajes de programación existen librerías que facilitan la extracción de información, sirven unos ejemplos como SimpleXML de PHP, o Nokigiri en Perl, TinyXML en C++.

#### 2.2.3.5. Aplicaciones

Existen también pequeñas extensiones para los navegadores que facilitan la extracción de datos. En el navegador *Firefox* existe por el ejemplo la extensión *Firebug* que permite inspeccionar cualquier elemento de una página HTML aunque en este caso la extracción del dato sería manual, lo mismo que *WebDeveloper*.

Si se quiere automatizar, existe la posibilidad de usar la extensión *iMacros* de *Firefox*. Consiste en realizar una grabación de un conjunto de acciones de modo que el software será capaz de reproducirlo estas acciones en cualquier momento.



# Capítulo 3: Desarrollo del Algoritmo

---

Julio Verne (08/02/1828 - 24/03/1905). Escritor francés.

La ciencia se compone de errores, que a su vez son los pasos hacia la verdad.



### 3.1. INTRODUCCIÓN

En la actualidad existen portales dedicados al deporte femenino español que tienen un equipo detrás que da soporte con las noticias. En el caso de este proyecto se va a evitar ese equipo de personas y se van a conseguir las noticias de una manera automática, siguiendo de alguna manera la ideología inicial del periódico americano *The Huffington Post*<sup>32</sup> que no disponía de corresponsales a lo largo y ancho del mundo pero cubría la información de todo el mundo agregando lo que otros hacen a modo de blog.

### 3.2. FUENTES DE INFORMACIÓN

Para obtener la información, se contemplan tres posibles escenarios que se analizan a continuación.

#### 3.2.1. PERIÓDICOS DEPORTIVOS

Según el [DRAE01] un periódico es una publicación que sale diariamente. El caso que concierne al proyecto se centrará en los periódicos deportivos dado que el tema central del periódico será el deporte y no la información general como pueden ser los periódicos generalistas.

A nivel nacional existen tres grandes periódicos deportivos, mostrados a continuación alfabéticamente:

##### ➤ *Diario AS en su edición digital*



Ilustración 4 Cabecera periódico AS<sup>33</sup>

Estudiando los contenidos de su página se obtiene la grata sorpresa de disponer una sección dedicada exclusivamente al deporte femenino que es referenciada siempre desde la página principal del periódico.



Ilustración 5 Sección NOSOTRAS<sup>34</sup> del periódico AS

La parte menos interesante para el proyecto es el ámbito de las noticias, pues son noticias de deportistas internacionales provocando que no haya demasiadas noticias de relevancia para el objetivo del proyecto.

<sup>32</sup> Periódico americano que nació como un blog que agregaba noticias y comentarios. Recientemente ha sido comprado por AOL por 231 millones de dólares. <http://www.huffingtonpost.com>

<sup>33</sup> Diario AS <http://www.as.com>

<sup>34</sup> Diario AS, sección NOSOTRAS <http://www.as.com/mas-deporte/nosotras/>

➤ *Diario EL MUNDO DEPORTIVO en su edición digital*



Ilustración 6 Cabecera del periódico MUNDO DEPORTIVO<sup>35</sup>

Estudiando los contenidos de su página se observa que no tiene ninguna sección dedicada al deporte femenino, tan sólo aparecen referencias en determinados momentos pero no tienen un seguimiento específico.

➤ *Diario MARCA en su edición digital*



Ilustración 7 Cabecera del periódico MARCA<sup>36</sup>

Estudiando los contenidos de su página, se observa que dispone de varias secciones para los deportes más populares:

- En el caso del fútbol, dispone de una sección para el fútbol femenino.
- En el caso del baloncesto, dispone de una sección para las competiciones gestionadas por la Federación Española de Baloncesto (FEB) entre las que se incluye el baloncesto femenino.
- En el caso de motor, tenis, ciclismo, balonmano, atletismo y golf se entremezclan las noticias masculinas y femeninas.
- Dispone de una sección cajón desastre denominada Más Deportes que aglutina el resto de deportes incluyendo noticias de deportistas femeninas.

Siguiendo a diario las noticias, se puede observar que no se da la misma cobertura a todos los deportes.

### 3.2.2. FEDERACIONES DEPORTIVAS

Debido a la poca repercusión existente en los periódicos, hay que buscar otra alternativa y se decide por la búsqueda de información en prensa más especializada, es decir, páginas web que hablen exclusivamente de un único deporte a nivel nacional. Por desgracia, al no ser todos los deportes populares no tienen páginas con prensa especializada.

Ante esta última negativa se opta por localizar las páginas de las distintas federaciones deportivas.

A continuación, se muestra en una tabla un estudio con la información de las distintas federaciones españolas de deportes según la fuente de información que se puede usar.

<sup>35</sup> Diario MUNDO DEPORTIVO <http://www.mundodeportivo.com>

<sup>36</sup> Diario MARCA <http://www.marca.com>

Tabla 1 Análisis de las federaciones con disciplinas olímpicas de verano

DEPORTE	Sección Noticias	RSS
Atletismo	SI	NO
Bádminton	SI	NO
Baloncesto	SI	SI
Balonmano	SI	NO
Boxeo	NO	NO
Ciclismo	SI	NO
Esgrima	SI	NO
Fútbol	SI	NO
Gimnasia	SI	NO
Golf	SI	SI
Halterofilia	NO	NO
Hípica	SI	NO
Hockey	SI	NO
Judo	NO	NO
Lucha	SI	NO
Natación	SI	NO
Piragüismo	SI	NO
Remo	SI	NO
Rugby	SI	NO
Taekwondo	SI	SI
Tenis	SI	SI
Tenis de mesa	SI	NO
Tiro con Arco	NO	NO
Tiro Olímpico	NO	NO
Triatlón	SI	NO
Vela	SI	SI
Voleibol	SI	NO

Como se puede observar, el método de difusión de noticias a través de RSS no está extendido y son pocas las que disponen de ello, incluso hay páginas que no queda muy claro dónde aparecen las noticias.

Al tratarse el proyecto de un prototipo, sólo se han contemplado para el estudio las federaciones que disponen a día de hoy de alguna disciplina deportiva en los Juegos Olímpicos de verano, por lo que otras federaciones como la de deportes de invierno, deportes de hielo, deportes paralímpicos, patinaje, pesca, salvamento y socorrismo, espeleología, colombofilia, caza, etc.... no han sido incluidas en el estudio.

Debido a esto, es factible añadir estas federaciones que no han sido incluidas en el estudio, por ello, se ha incluido en la visualización una sección para las federaciones de deportes paralímpicos, como ejemplo de sencillez del método y posibilidades que ofrece la metodología diseñada.

A continuación, se muestran tres tablas con todas esas federaciones de deportes que no han sido incluidos en el estudio. Por un lado las federaciones con disciplinas olímpicas en los juegos de invierno, otra tabla de federaciones con disciplinas paralímpicas y por último federaciones que no disponen de ninguna disciplina olímpica.

Tabla 2 Federaciones con disciplinas olímpicas de invierno

DEPORTE	Sección Noticias	RSS
Deportes de Hielo	SI	NO
Deportes de Invierno	SI	NO

Tabla 3 Federaciones con disciplinas paralímpicas

DEPORTE	Sección Noticias	RSS
Discapacidad física	SI	NO
Discapacidad intelectual	SI	NO
Ciegos	SI	NO
Sordos	SI	NO
Paralíticos cerebrales	SI	NO

Tabla 4 Federaciones sin disciplinas olímpicas

DEPORTE	Sección Noticias	RSS
Actividades subacuáticas	NO	NO
Aeronáutica	SI	NO
Ajedrez	NO	NO
Automovilismo	NO	NO
Baile deportivo	NO	NO
Béisbol y Sófbol	NO	NO
Billar	SI	NO
Bolos	NO	NO
Caza	SI	SI
Colombofilia	NO	NO
Colombicultura	SI	NO
Dominó	NO	NO
Espeleología	NO	NO
Esquí náutico	NO	NO
Fútbol americano	SI	NO
Galgos	SI	NO
Kárate	SI	NO
Kickboxing	NO	NO
Montaña y Escalada	SI	NO
Motociclismo	SI	NO
Motonáutica	SI	NO
Orientación	SI	NO
Pádel	SI	NO
Patinaje	SI	NO
Pelota	NO	NO
Pesca y Casting	SI	NO
Petanca	NO	NO
Polo	SI	NO
Salvamento y Socorrismo	NO	NO
Squash	SI	NO
Surf	SI	NO
Tiro a vuelo	NO	NO

### 3.3. ANÁLISIS GRAMATICAL

Para el diseño del algoritmo se ha optado por analizar la gramática española y sus estructuras para localizar el tipo de noticias especificadas en el proyecto, noticias de mujeres deportistas españolas que ensalcen el deporte español y la operativa que tiene el mayor buscador del momento, Google, en las búsquedas [véase Apéndice D].

#### 3.3.1. GRAMÁTICA ESPAÑOLA

Según [RAE09] una proposición consta de dos partes [véase Apéndice C], un sujeto y un predicado, siendo el sujeto quién realiza la acción del verbo con el que concuerda en género y número, y el predicado es aquello que se dice del sujeto.

El sujeto consta de un sintagma nominal cuyo núcleo es un sustantivo, mientras que el predicado consta de un sintagma verbal cuyo núcleo es un verbo acompañado de diversos complementos.

El predicado puede ser de dos tipos, predicado nominal si el núcleo es un verbo copulativo (ser, estar o parecer) acompañado de un complemento atributo, o predicado verbal si el verbo no es de tipo copulativo, en este caso irá acompañado de otros complementos: directo, indirecto, de lugar...

#### 3.3.2. ANÁLISIS LÉXICO

Una de las herramientas que utilizan los buscadores es el análisis léxico. Esto es, examinar la proposición con la finalidad de reconocer y clasificar los distintos componentes léxicos. Este análisis se lleva a cabo mediante dos técnicas:

- 1) *Eliminación de palabras irrelevantes*: en este paso se eliminan todas aquellas palabras que no aporten un significado a la búsqueda, y esto se hace dependiendo de su categoría gramatical [véase Apéndice C]. Esas categorías sin aporte son las denominadas categorías cerradas: artículos, conjunciones, determinantes, interjecciones, preposiciones y pronombres. Es decir, solo interesan los adjetivos, adverbios, sustantivos y verbos.

El siguiente gráfico ilustra dos grandes buscadores, uno que sí realiza esta eliminación de palabras, Google, y otro que no, Altavista.



Ilustración 8 Eliminación de palabras irrelevantes (Google)

Como se puede apreciar en la imagen, los resultados ofrecidos en la búsqueda por Google no aparece ninguna palabra irrelevante y resalta en los resultados las palabras relevantes.

Por otro lado, en la siguiente imagen del buscador Altavista<sup>37</sup> se puede apreciar como no realiza ese análisis y resalta en los resultados cualquier palabra que se encuentre en la cadena de búsqueda, esto provoca como es el caso resultados poco interesantes para el usuario.

---

<sup>37</sup> www.altavista.com





Ilustración 9 NO eliminación de palabras irrelevantes (AltaVista)

- 2) Lematización o Stemming: este paso consiste en encontrar la raíz, “stem” en inglés, de las palabras de modo que se amplía el rango de búsqueda a la familia de la palabra, iniciándose así una combinación o derivación del lexema con los distintos sufijos y prefijos.

El ejemplo más sencillo en castellano son las conjugaciones verbales y las formas no personales de los verbos, pues todos ellos tienen la misma raíz.

Tabla 5 Ejemplo de derivación

Raíz	Infinitivo	Presente	Preterito Perfecto	Futuro	Condicional
<b>And-</b>	And- ar	And- o	And- uve	And -aré	And -aría

Computacionalmente hablando, esta derivación se realiza mediante expresiones regulares. Además, se han desarrollado distintos algoritmos para obtener y analizar las familias de las palabras, sirvan de ejemplo:

- Algoritmo de Porter: trunca la palabra para obtener la raíz.
- SPLIT: detecta los prefijos y sufijos que forman la palabra y obtiene la raíz más probable.
- STON.

En la siguiente imagen, se puede apreciar un ejemplo, en inglés, existente en una página web accesible a cualquier público.

## Snowball - Demo

Enter some words to stem, using the English stemming algorithm, and then click on Submit:


wonderful excitment

### Results

```
wonderful -> wonder
excitment -> excit
```

Ilustración 10 Ejemplo de stemming con el algoritmo de Porter<sup>38</sup>

Sin embargo, un ejemplo es castellano podría ser el siguiente algoritmo desarrollado por el Grupo de Estructura de Datos y Lingüística Computacional de la Universidad de la Laguna<sup>39</sup>.



Grupo de Estructuras de Datos  
y Lingüística Computacional

## Resultados de la lematización

---

Resultado del reconocimiento de **lentísimo**

---

• Forma canónica: lento

• Categoría: adjetivo usado también como sustantivo masculino

• Flexión: superlativo ( masculino singular )

Ilustración 11 Ejemplo de stemming en castellano<sup>40</sup>

<sup>38</sup> Lematizador de lengua inglesa bajo el algoritmo de Porter (<http://snowball.tartarus.org/demo.php>)

<sup>39</sup> GEDLC Universidad La Laguna (<http://www.gedlc.ulpgc.es>).

<sup>40</sup> Lematizador de lengua castellana  
(<http://www.gedlc.ulpgc.es/investigacion/scogeme02/lematiza.htm>)

### 3.3.3. ANÁLISIS SINTÁCTICO

El análisis sintáctico puede describirse como el intento de construir un árbol de análisis sintáctico para la proposición que se quiere reconocer. Existen dos tipos distintos:

- *Analizador sintáctico descendente*: intenta construir el árbol de análisis sintáctico empezando desde la raíz y descendiendo hacia las hojas o, lo que es lo mismo, intenta obtener una derivación por la izquierda para una cadena de entrada comenzando desde la raíz y creando los nodos del árbol en orden previo.
- *Analizador sintáctico ascendente*: intenta construir el árbol de análisis sintáctico empezando desde las hojas y ascendiendo hacia la raíz o, lo que es lo mismo, intenta obtener una reducción desde una cadena hasta llegar al axioma principal.

En los motores de búsqueda, una vez realizado el análisis léxico, y a partir del mismo, se construirá el árbol sintáctico que se corresponde con cada una de las frases que se van a almacenar en la base de datos para realizar la búsqueda.

De modo que, si se realizara la búsqueda de la frase: “*La persona anciana escuchaba atentamente*”, tras el análisis léxico la frase quedaría: “*persona anciana escuchar atentamente*”. Si ahora se realiza el análisis sintáctico descendente usando la estructura de un árbol de derivación, queda lo siguiente:

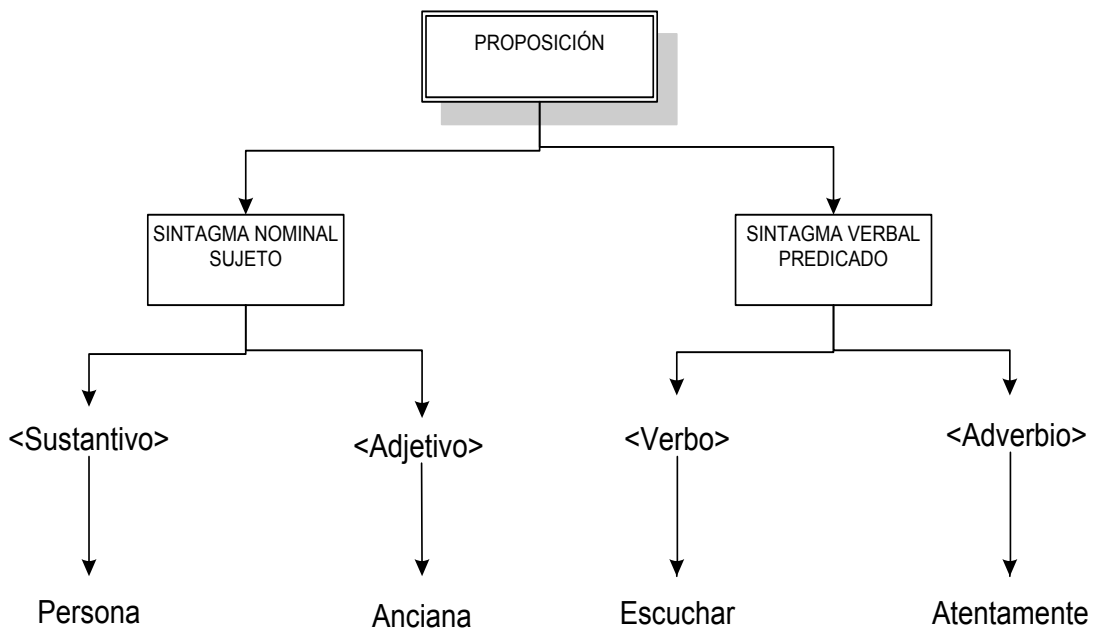


Ilustración 12 Estructura árbol del análisis sintáctico

Tras hacer este análisis sintáctico, el motor de búsqueda toma las hojas del árbol y realiza la búsqueda basándose en dichas hojas.

### 3.3.4. ANÁLISIS SEMÁNTICO

La semántica se encarga de analizar los aspectos del significado o la interpretación del significado de las palabras. Para ello, se analiza el referente de las palabras (lo que denotan, su significado básico según el DRAE) y el sentido de la palabra (el significado figurado, subjetivo o personal). Para hacer un análisis semántico automático se parte del árbol sintáctico realizado en el análisis sintáctico y detectar así posibles restricciones o limitaciones semánticas.

Al realizar un análisis semántico, surgen dos problemas la homonimia y la polisemia. La polisemia se produce cuando una palabra tiene distintas acepciones, mientras que la homonimia es la distinta significación de una palabra cuando las distintas acepciones de la palabra vienen motivadas por distintos orígenes etimológicos. Un ejemplo simple sería:

Hinojo<sup>41</sup>: planta (del latín *feneculum*)/rodilla (del latín *genuculum*)

Cabo<sup>42</sup>: extremo/rango militar/accidente geográfico...

En la actualidad, existen diversos analizadores semánticos automáticos que tratan de analizar el contenido y las opiniones de las redes sociales (twitter, facebook...) y añaden el determinado contexto personal, esto es, la intencionalidad, subjetividad, sentimiento... en las búsquedas. En este ámbito, Google patentó el "*Aggregating Context data for programmable search engine*"<sup>43</sup> que consiste en una metodología propia para realizar búsquedas contextuales.

Analizando el problema planteado en este proyecto desde un punto de vista semántico, se puede observar cómo las búsquedas a realizar por el algoritmo son búsquedas contextualizadas en un ámbito muy concreto. Dada ésta excepcionalidad y unicidad de contexto, es necesario realizar un solución a medida pues en la actualidad no existe ningún software que solucione estas búsquedas tan precisas, es por ello que, el buscador de noticias a realizar será el **PRIMER BUSCADOR de LOGROS de DEPORTISTAS FEMENINOS en ESPAÑA.**

## 3.4. DESARROLLO DEL ALGORITMO DE BÚSQUEDA DE NOTICIAS

Una vez visto como funciona un buscador de noticias en Internet toca analizar el buscador de noticias que se va a implementa en este proyecto.

### 3.4.1. NOTICIAS VÁLIDAS vs. NOTICIAS INVÁLIDAS

Para ello, el primer paso es recoger una base de datos amplia de noticias, dividiéndolas en válidas y no válidas a juicio del analista para estudiar la frecuencia de las palabras y junto con su categoría gramatical. Basándose en los requisitos del buscador, el analista define qué es una noticia válida y qué es una noticia inválida.

---

<sup>41</sup> [http://buscon.rae.es/drae/SrvltGUIBusUsual?TIPO\\_HTML=2&TIPO\\_BUS=3&LEMA=hinojo](http://buscon.rae.es/drae/SrvltGUIBusUsual?TIPO_HTML=2&TIPO_BUS=3&LEMA=hinojo)

<sup>42</sup> [http://buscon.rae.es/drae/SrvltGUIBusUsual?TIPO\\_HTML=2&TIPO\\_BUS=3&LEMA=cabo](http://buscon.rae.es/drae/SrvltGUIBusUsual?TIPO_HTML=2&TIPO_BUS=3&LEMA=cabo)

<sup>43</sup> <http://patents.com/us-20070038601.html>

El analista define una noticia válida como aquella noticia cuyo protagonista principal sea femenino y cuya acción sea un logro o éxito deportivo a nivel nacional o internacional tanto individual como colectivo.

Ejemplos de noticias válidas serían los siguientes titulares encontrados en las ediciones digitales de periódicos deportivos:

“Marina Alabau vuelve a ganar en la clase RS:X de vela”  
 “El equipo español femenino de kumite gana la plata”  
 “Y Duane remata el oro con un ¡bronce!”

Por otro lado, ejemplos de noticias no válidas a juicio del analista serían los siguientes titulares:

“Gómez Noya se impone en la cuarta etapa del Mundial”  
 “Mireia Belmonte, fuera de la final de 4 estilos”  
 “Madrid será escenario del Máster Nacional de Tenis en Silla de Ruedas”

Como se puede observar, los titulares válidos cumplen los requisitos de protagonista principal de la noticia sea femenino (Marina Alabau, equipo español femenino y Duane da Rocha) y obtienen un logro (victoria y medalla). Mientras tanto, en las noticias inválidas se observa como en el primer caso falla el protagonista de la noticia (Gómez Noya), en la segunda noticia falla el logro (fuera de la final) y en la última noticia fallan ambas cosas.

### 3.4.2. IMPACTO DE LAS PALABRAS EN LAS NOTICIAS

Una vez recopilado una cantidad importante de noticias (cien para las noticias válidas y otras cien para las noticias inválidas), durante los meses de octubre y noviembre de 2010, de las distintas fuentes descritas en los capítulos anteriores, se procede a analizar un estudio del impacto de las palabras en las noticias buscando algún tipo de regla significativa que ayude a discriminar las noticias.

Para poder llevar a cabo este análisis se dispone de un software implementado en una de las prácticas realizadas en la titulación.

Los resultados obtenidos en las noticias consideradas válidas son:

Tabla 6 Resultado del análisis de noticias válidas

<b>Palabras totales:</b>	<b>4774</b>
<b>Palabras distintas:</b>	<b>1385</b>
<b>De</b>	<b>316</b>
<b>La</b>	<b>261</b>
<b>En</b>	<b>229</b>
<b>Y</b>	<b>147</b>
<b>A</b>	<b>123</b>
<b>Que</b>	<b>98</b>

Los resultados obtenidos en las noticias consideradas no válidas son:

Tabla 7 Resultado del análisis de noticias inválidas

<b>Palabras totales:</b>	<b>4877</b>
<b>Palabras distintas:</b>	<b>1632</b>
<b>De</b>	356
<b>El</b>	222
<b>La</b>	203
<b>En</b>	186
<b>Del</b>	98
<b>Y</b>	96

Debido a la gran cantidad de palabras distintas se ha optado por marcar cuáles son las cinco noticias con mayor frecuencia de aparición. Como se puede observar en ambas tablas las palabras corresponden a esas categorías gramaticales que no aportan ningún significado. Por lo que el estudio, a primera vista, no ha generado ningún resultado positivo.

A continuación, se muestra un análisis más detallado buscando palabras de categorías más interesantes en base a una posible función sintáctica dentro de la noticia.

#### DETERMINANTES:

- ✓ El número de "LA" es superior al de "LE" en las buenas, justo al revés que en las malas.
- ✓ "LAS" aparece más veces en las buenas que en las malas; en las buenas "LOS" aparece más que "LAS".
- ✓ "UNA" sale más en las malas, y menos veces que "UN" en ambas.
- ✓ "UN" sale más en las buenas, y más veces que "UNA" en ambas.
- ✓ "UNAS" sale una vez en las malas y nada en las buenas.
- ✓ "UNOS" sale una vez en las buenas y nada en las malas.

#### SUJETOS - SUSTANTIVOS

- ✓ "ESPAÑA" sale el doble en las buenas.
- ✓ "ESPAÑOLAS" solo en las buenas.
- ✓ "ESPAÑOLA" sale el doble en las buenas.
- ✓ "ESPAÑOL" casi el doble en las malas.
  
- ✓ "CHICAS" solo en las buenas.
- ✓ "FEMENINA" sale el doble en las buenas.
- ✓ "FEMENINO" sale el doble en las buenas.
- ✓ "FEMENINAS" solo en las buenas.
- ✓ "FEMENINOS" solo en las buenas.
- ✓ "SELECCIÓN" igual en ambas.
- ✓ "SELECCIONES" cinco veces más en las buenas.
- ✓ "EQUIPO" sale el doble en las buenas.
- ✓ "EQUIPOS" sale seis veces más en las buenas.
  
- ✓ "TENISTAS" solo en las buenas.
- ✓ "JUGADOR" solo en las malas.

- ✓ "JUGADORA" solo en las malas.
- ✓ "JUGADORAS" solo en las malas.
- ✓ Nombres de mujeres deportistas.

OBJETOS DIRECTOS

- ✓ "FINAL" aparece cuatro veces más en las buenas.
- ✓ "FINALES" sale el doble en las buenas.
- ✓ "COPA" sale el doble en las buenas.
- ✓ "MUNDO" misma cantidad.
- ✓ "MUNDIAL" más en las malas que en las buenas.
- ✓ "EUROPA" solo en las buenas.
- ✓ "EUROPEO" más en las malas.
- ✓ "CAMPEONATO" solo en las buenas.
- ✓ "CAMPEONATOS" cuatro veces más en las buenas.
- ✓ "OLIMPICO" solo en las buenas.
- ✓ "OLIMPICOS" más en las malas.
  
- ✓ "ORO" solo en las buenas.
- ✓ "PLATA" solo en las buenas.
- ✓ "BRONCE" solo en las buenas.
- ✓ "MEDALLA" solo en las buenas.
  
- ✓ "MEDALLISTA" solo en las buenas.
- ✓ "FINALISTAS" solo en las buenas.
- ✓ "PRIMERA" más en las malas.
- ✓ "PRIMERO" una vez en las malas, ninguna en las buenas.
- ✓ "SEGUNDA" más en las buenas.
- ✓ "SEGUNDO" solo en las malas.
- ✓ "TERCERA" solo en las buenas.
- ✓ "CUARTA" el triple en las malas.
- ✓ "PREMIOS" solo en las buenas.
- ✓ "TITULO" cinco veces más en las buenas.
- ✓ "PODIO" igual en ambas.
- ✓ "VICTORIA" igual en ambas.
- ✓ "LIDER" igual en ambas.
- ✓ "DEPORTISTA" el doble en buenas.
- ✓ "DESTACAR" solo en las malas.
  
- ✓ "CAMPEONA" solo en las buenas.
- ✓ "CAMPEONES" tres veces más en las buenas.
- ✓ "CAMPEON" igual en ambas.
- ✓ "CAMPEONAS" solo en las buenas.
  
- ✓ "MEJOR" sale el doble en las buenas.
- ✓ "MARCA" sale el doble en las buenas.
  
- ✓ "BUENAS" solo en las buenas.
- ✓ "EXCELENTE" solo en las malas.
- ✓ "RAPIDA" solo en las buenas.
- ✓ "SEMIFINALES" cuatro veces más en las buenas.
- ✓ "ESTRELLAS" el doble en las buenas.

### VERBOS

- ✓ "GANA" solo en las buenas.
- ✓ "GANAS" solo en las buenas.
- ✓ "GANAR" solo en las malas.
- ✓ "GANARA" solo en las malas.
- ✓ "VENCEDOR" solo en las malas.
- ✓ "CONSIGUIO" solo en las buenas.

Con este análisis solamente ayuda a esclarecer las distintas categorías gramaticales y sus posibles funciones sintácticas dentro del titular de la noticia. Es por ello que el siguiente paso es analizar sintácticamente los titulares de las noticias.

### 3.4.3. ANÁLISIS SINTÁCTICO DE TÍTULOS DEPORTIVOS

Debido a que los titulares de las noticias deben ser breves, lo más exacto y fieles posibles además de llamativos tal y como se comentó en el capítulo anterior; un titular, en definitiva, es una proposición por lo que su estructura es la siguiente:

SUJETO + VERBO + COMPLEMENTOS (ATRIB/OD/OI/LUGAR...)

España es campeona del mundo

Partiendo de esta estructura básica yuxtaponiéndola con el tipo de noticias válidas para el buscador se obtienen las siguientes reglas básicas:

- **R1:** El SUJETO debe ser una deportista o grupo de deportistas femeninas.
- **R2:** El VERBO debe ser positivo.
- **R3:** El COMPLEMENTO debe ser un logro, luego será un OBJETO DIRECTO o ATRIBUTO.

En el momento que no se cumpla una de las tres reglas, la noticia no será válida.

Analizando todos los titulares recogidos (200, de los cuales 100 son positivos y 100 son negativos) siguiendo estas tres reglas básicas para cumplir el objetivo, se observa que hay noticias válidas que no cumplen estas reglas, estudiándolas se descubre la aparición de una figura retórica: la elipsis.

### 3.4.4. LA ELIPSIS

La **elipsis** consiste en la supresión de algún elemento lingüístico del discurso sin contradecir las reglas gramaticales. Esto implica que uno o más de los tres elementos anteriormente citados puede desaparecer y provoca las siguientes posibles estructuras:

SUJETO, + ATRIBUTO/OD

España, campeona del mundo

ATRIBUTO/OD + SUJETO

Campeonato del mundo para España

VERBO + ATRIBUTO/OD

Somos campeones del mundo



## ATRIBUTO/OD

Campeones del mundo

Con estas nuevas estructuras aparecen las siguientes nuevas reglas:

- **R4:** Deben aparecer SUJETO, VERBO y ATRIBUTO/OD, en este orden.
- **R5:** Deben aparecer SUJETO y ATRIBUTO/OD en este orden.
- **R6:** Deben aparecer ATRIBUTO/OD y SUJETO en este orden.

La tercera y cuarta opción, elipsis del sujeto o elipsis de sujeto y verbo, son tipos de titulares que no han aparecido en las fuentes consultadas pero que sí son titulares válidos mucho más frecuentes de prensa escrita que prensa electrónica debido a que son más llamativos y se utilizan como reclamo comercial en la portada de los periódicos para obtener mejores resultados en las ventas.

### 3.4.5. VERBOS POSITIVOS Y NEGATIVOS

Tras realizar el pertinente nuevo análisis con todas las reglas hasta ahora recopiladas, aparece el siguiente titular:

SUJETO + VERBO + OD

España pierde el campeonato del mundo

SUJETO, + OD

Como se puede observar, el titular corresponde al primer tipo de estructura pero por el contrario es reconocido por el segundo tipo de estructura, esto es debido a la NO existencia de un verbo positivo lo que implica ir a buscar la siguiente estructura que sí es válida. Para poder evitar esto se añade una nueva regla para el verbo.

- **R7:** Si no hay verbo positivo, no puede existir un verbo negativo.

En el apéndice H se podrá observar una lista de verbos positivos y otra lista de verbos negativos.

### 3.4.6. SUJETOS MALICIOSOS

De este modo se consiguen todas las noticias “válidas”. Leyendo la noticia se observa que no pertenece a un equipo femenino, sino a un equipo masculino. Esto hace florecer un nuevo problema, esta vez en los sujetos y es la existencia de “*sujetos malicioso*” esto es debido a la existencia de los denominados *sustantivos colectivos* que hacen referencia a un grupo de personas, animales o cosas. En el caso que atañe a este proyecto existen varios sustantivos de este tipo tales como: Selección española, Equipo español, España, la Roja...

Dado que ya están analizadas todas las palabras del titular de la noticia no queda más remedio que profundizar dentro de la noticia y analizar el contenido de la misma, de este modo surge una nueva regla:

- **R8:** Ante un sustantivo colectivo, analizar el contenido de la noticia en busca de palabras clave: *femenino, femenina, chica(s), ...*

Las palabras a buscar en esta última regla, no han sido elegidas al azar, sino que tomando como referencia el estudio inicial de la frecuencia de las palabras, se buscan palabras de género femenino que sean de interés tales como: *femenino, femenina, chica, chicas* o todos aquellos sujetos de nombres propios de mujeres deportistas.

### 3.4.7. REGLAS DEL ALGORITMO

En definitiva, se han generado una serie de reglas que conformarán un algoritmo para optimizar la búsqueda de las noticias válidas en las páginas de referencia. Mediante la siguiente tabla, se muestra un pequeño resumen de las reglas obtenidas durante el proceso anteriormente descrito:

Tabla 8 Resumen de reglas

<b>R1</b>	Existencia de sujeto femenino
<b>R2</b>	Existencia de verbo positivo
<b>R3</b>	Si no hay verbo positivo, ausencia de verbo
<b>R4</b>	Existencia de un logro (complemento)
<b>R5</b>	Estructura: Sujeto + Verbo + Complemento
<b>R6</b>	Estructura: Sujeto + Complemento
<b>R7</b>	Estructura: Complementos + Sujeto
<b>R8</b>	Ante la existencia de un sujeto malicioso, analizar el contenido de la noticia en busca de palabras clave

Para una mejor percepción y seguimiento de dichas reglas, se muestra a continuación [véase Ilustración 13] un diagrama de flujo con las reglas definiendo así el algoritmo de búsqueda a implementar. Además, en el Apéndice G se puede encontrar el código de las funciones clave en el algoritmo de búsqueda.

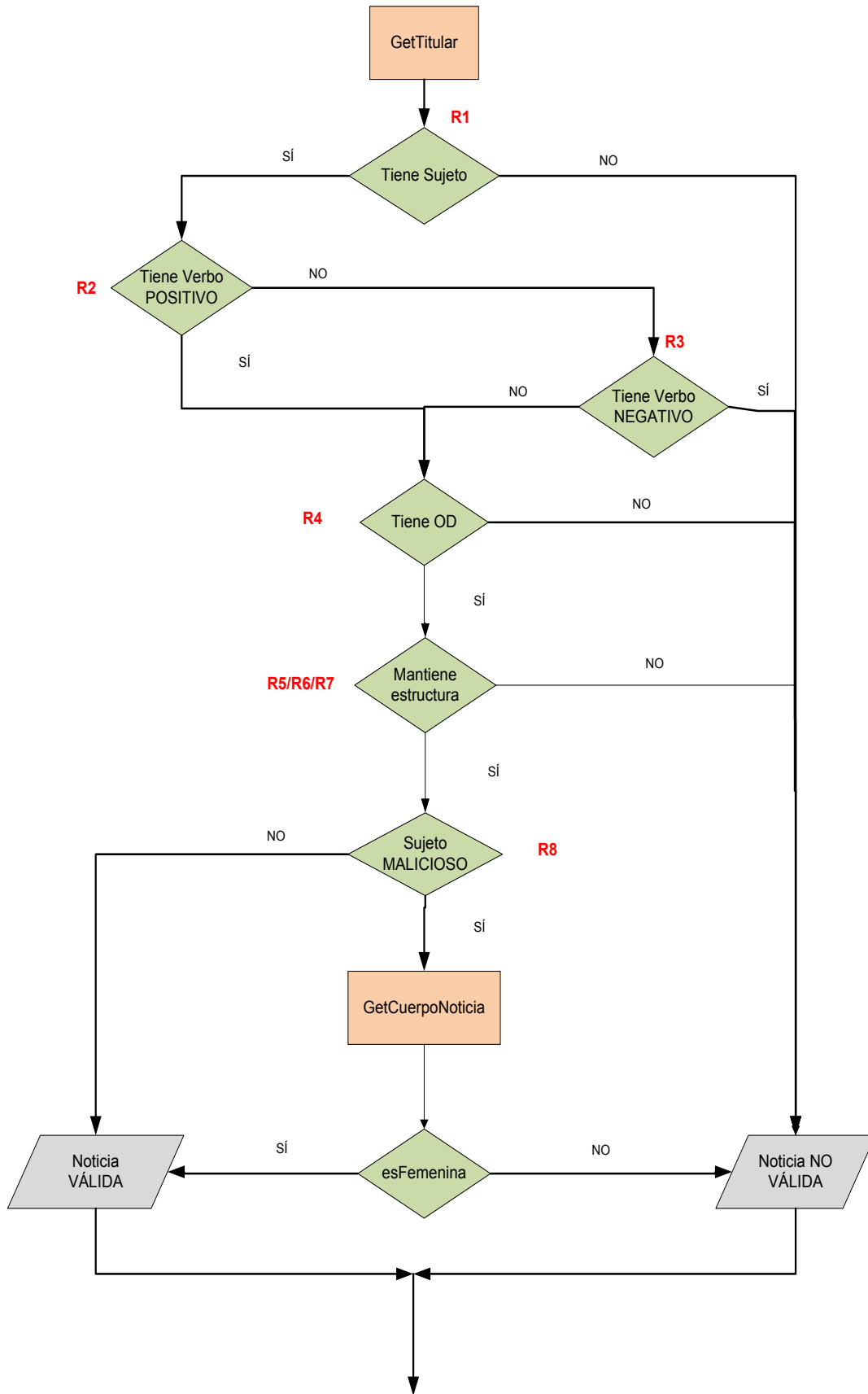


Ilustración 13 Algoritmo de búsqueda del analizador sintáctico automático de noticias

### 3.5. EJEMPLO DE LA EJECUCIÓN DEL ALGORITMO

Para ver el funcionamiento del algoritmo, se muestra el ejemplo de una de las noticias mostradas en la página. En la sección de noticias de la Federación Española de Luchas Olímpicas y Disciplinas Asociadas aparece la siguiente noticia:

#### PLATA PARA MINERVA MONTERO EN EL AUSTRIAN LADIES OPEN



Finalizó el torneo internacional de Austria con una medalla de plata a cargo de la canaria Minerva Montero en 55 Kg. que consiguió tras ganar a Escocia, Canadá (Andrea Ross) y Canadá (J. Barrer) en semifinales, en la final no pudo con la también canadiense (B. Laverdure), además lucharon por el bronce Sara Sánchez en 48 Kg., Mari Mar Serrano en 51 Kg. y Aurora fajardo en 63 Kg. pero no consiguieron la victoria clasificándose en el quinto puesto. Ahora el equipo seguirá entrenando en Austria esta semana con otras selecciones nacionales. Felicidades Minerva.

06/06/11

Ilustración 14 Ejemplo de noticia válida

El algoritmo, al analizar el código fuente de la página web, encuentra lo siguiente:

```
<td class="contentheading" width="100%">
  <a
    href="http://www.felucha.com/index.php?option=com_content&task=view&id=1078&Itemid=83" class="contentpagetitle">
    <strong> <font style="font-size:12px">PLATA PARA MINERVA MONTERO
    EN EL AUSTRIAN LADIES OPEN</font> </strong>
  </a>
</td>
```

El algoritmo aísla el titular de la noticia: “PLATA PARA MINERVA MONTERO EN EL AUSTRIAN LADIES OPEN” y lo empieza a analizar. Siguiendo la nomenclatura de reglas descrita en el apartado anterior:

**R1:** Localizar un sujeto femenino; “**MINERVA MONTERO**”.

**R2:** Localizar un verbo positivo; *no tiene*.

**R3:** Localizar un verbo negativo; *no tiene*.

**R4:** Localizar un logro (OD); “**PLATA**”.

**R5:** Estructura Sujeto + Verbo + OD; *no tiene*.

**R6:** Estructura Sujeto + OD; *no tiene*.

**R7:** Estructura OD + Sujeto; **SI TIENE**.

**R8:** Localizar sujeto malicioso; *no tiene*.

**NOTICIA VÁLIDA.**

## 3.6. DISEÑO DE LA BASE DE DATOS

### 3.6.1. MODELO E/R

Para poder obtener las noticias se necesita una lista de fuentes, y según el estudio anterior es necesario una entidad denominada PÁGINA. Debido a la existencia de dos tipos de fuentes: RSS o URL surge la jerarquía en base al TIPO de la página, creándose así una entidad para las URL y otra para las RSS.

Por otro lado, según se desprende del algoritmo se necesitan diversas listas de palabras según su categoría gramatical, pero estas listas de palabras no tienen ningún tipo de relación entre ellas por lo que se realiza una tabla por cada una de las categorías. Las tablas son:

- SUJETO: lista de palabras sujeto.
- VERBO: lista de palabras de verbos.
- OD: lista de palabras de logros de los deportistas.

En el caso del SUJETO, para optimizar la búsqueda se decide realizar una jerarquía dependiendo del DEPORTE creándose así una entidad por deporte y de este modo la lista de sujetos a buscar será lo más afinada posible al deporte en cuestión, teniendo únicamente los de dicho deporte.

En el caso del VERBO, al existir dos tipos se opta por crear otra jerarquía, esta vez según el significado del verbo, si es POSITIVO o no, lo que da lugar a dos entidades, VERBOSP que albergará los verbos de carácter positivo y VERBOSN con los verbos de carácter negativo.

Además de estas tablas, será necesario una entidad que almacene las noticias válidas según el algoritmo, denominada NOTICIA.

A continuación, se muestra de una manera gráfica el modelo entidad-relación y se detalla cada campo.

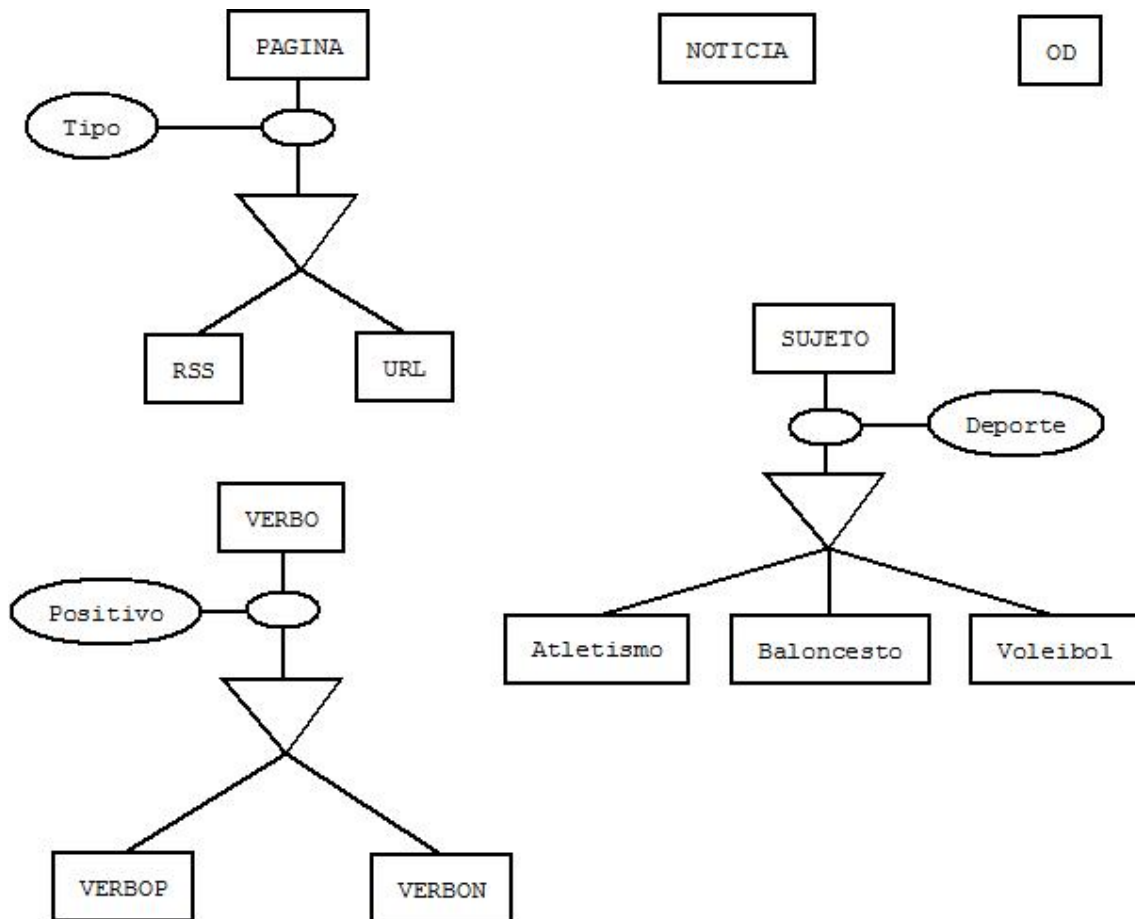


Ilustración 15 Modelo E/R

### 3.6.2. DICCIONARIO DE DATOS

Las entidades URL y RSS:

- *Id*: identificador que será clave primaria de la tabla. Tipo numérico. Clave primaria.
- *Titular*: dirección donde se va a buscar la información. Tipo texto.
- *Deporte*: contendrá el deporte de la página. Tipo texto.
- *ER*: expresión regular para poder crear la dirección del contenido de la noticia. Tipo texto.

Las entidades SUJETO, VERBO y OD

- *Palabra*: palabra a buscar. Tipo texto. Clave primaria.

La entidad NOTICIA:

- *Id*: identificador que será clave primaria de la tabla. Tipo numérico.
- *Titular*: dirección donde se va a buscar la información. Tipo texto. Clave primaria.
- *Referencia*: dirección de la noticia. Tipo texto.
- *Deporte*: deporte de la noticia. Tipo texto.
- *Contenido*: contenido de la noticia. Tipo texto.
- *Fecha*: fecha de publicación de la noticia. Tipo fecha.

# Capítulo 4: Plataforma Web

---

Sir Arthur Charles Clarke (16/12/1917 - 19/03/2008). Escritor y científico británico.

Las leyes de Clarke:

1. Si un anciano y distinguido científico dice que algo es posible, es casi seguro que tenga razón; pero si dice que es imposible, muy probablemente estará equivocado.
2. La única forma de hallar los límites de lo posible es yendo más allá de ellos, hasta lo imposible.
3. Toda tecnología suficientemente avanzada no puede distinguirse de la magia.





## 4.1. ANÁLISIS DE LA PLATAFORMA WEB

### 4.1.1. ANÁLISIS DE REQUISITOS

El objetivo de la especificación de requisitos es definir de manera clara y precisa todas las funcionalidades y restricciones del sistema que se desea construir, de tal manera que, una vez construido, se pueda comprobar el cumplimiento de esas funcionalidades y restricciones.

Los requisitos se engloban en dos grandes categorías:

- ✓ **Requisitos de capacidad:** Representan lo que necesitan los usuarios para resolver un problema o lograr un objetivo.
- ✓ **Requisitos de restricción:** Son las restricciones impuestas por los usuarios sobre cómo se debe resolver el problema o cómo se debe alcanzar el objetivo.

Cada requisito debe ser definido con una serie de atributos, que le añadan toda la información necesaria para su seguimiento posterior y su clasificación. Estos atributos se describen a continuación:

- **Identificador:** cada requisito debe estar identificado de forma única. El formato es el siguiente: *UR-C-XXX* para los requisitos de capacidad y *UR-R-XXX* para los requisitos de restricción, donde *XXX* es un número de tres dígitos que indica el orden en que se registró el requisito y que es único para cada uno de los tipos (de capacidad y de restricción).
- **Necesidad:** Campo que indica si el requisito es imprescindible o no. Puede tomar uno de los siguientes valores: Esencial, Deseable, Opcional.
- **Prioridad:** Indica la preferencia que se debe dar al requisito con respecto al resto durante el desarrollo del sistema. Puede tomar uno de los siguientes valores: Alta, Media o Baja.
- **Estabilidad:** Indica la posibilidad de que el requisito sea variado en un futuro por el cliente. Puede tomar uno de los siguientes valores: Alta, Media o Baja.
- **Fuente:** Indica de dónde proviene el requisito.

En este análisis, en las siguientes secciones, sólo se detallarán los requisitos de necesidad esencial y los casos de uso básicos para el demostrador desarrollado.

#### 4.1.1.1. Requisitos de capacidad

En el siguiente apartado se muestra los requisitos de capacidad del sistema.

Tabla 9 UR-C-001

Identificador: UR-C-001			
<b>Descripción</b>	Ejecución del algoritmo de forma automática.		
<b>Necesidad</b>	<input checked="" type="checkbox"/> Esencial	<input type="checkbox"/> Deseable	<input type="checkbox"/> Opcional
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja

<b>Estabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Fuente</b>	Alumno.		

Tabla 10 UR-C-002

<b>Identificador: UR-C-002</b>			
<b>Descripción</b>	Ejecución del algoritmo varias veces al día.		
<b>Necesidad</b>	<input checked="" type="checkbox"/> Esencial	<input type="checkbox"/> Deseable	<input type="checkbox"/> Opcional
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Estabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Fuente</b>	Alumno.		

Tabla 11 UR-C-003

<b>Identificador: UR-C-003</b>			
<b>Descripción</b>	Existencia de una plataforma web que muestre los resultados de la ejecución del algoritmo.		
<b>Necesidad</b>	<input checked="" type="checkbox"/> Esencial	<input type="checkbox"/> Deseable	<input type="checkbox"/> Opcional
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Estabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Fuente</b>	Alumno.		

Tabla 12 UR-C-004

<b>Identificador: UR-C-004</b>			
<b>Descripción</b>	Visibilidad de la plataforma web en cualquier navegador.		
<b>Necesidad</b>	<input checked="" type="checkbox"/> Esencial	<input type="checkbox"/> Deseable	<input type="checkbox"/> Opcional
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Estabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Fuente</b>	Alumno.		

Tabla 13 UR-C-003

<b>Identificador: UR-C-005</b>			
<b>Descripción</b>	Permitir registro en la plataforma web para crear una comunidad de usuarios.		
<b>Necesidad</b>	<input checked="" type="checkbox"/> Esencial	<input type="checkbox"/> Deseable	<input type="checkbox"/> Opcional
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Estabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Fuente</b>	Alumno.		

Tabla 14 UR-C-006

<b>Identificador: UR-C-006</b>			
<b>Descripción</b>	Existencia de, al menos, tres tipos distintos de usuario en la plataforma web: administrador, registrado y visitante.		
<b>Necesidad</b>	<input checked="" type="checkbox"/> Esencial	<input type="checkbox"/> Deseable	<input type="checkbox"/> Opcional
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Estabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Fuente</b>	Alumno.		

Tabla 15 UR-C-007

<b>Identificador: UR-C-007</b>			
<b>Descripción</b>	Mostrar de una manera ordenada los titulares de las noticias encontradas en el algoritmo de búsqueda en la plataforma web.		
<b>Necesidad</b>	<input checked="" type="checkbox"/> Esencial	<input type="checkbox"/> Deseable	<input type="checkbox"/> Opcional
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Estabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Fuente</b>	Alumno.		

Tabla 16 UR-C-008

<b>Identificador: UR-C-008</b>			
<b>Descripción</b>	Permitir al usuario leer el contenido de la noticia en la plataforma web.		
<b>Necesidad</b>	<input checked="" type="checkbox"/> Esencial	<input type="checkbox"/> Deseable	<input type="checkbox"/> Opcional
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Estabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Fuente</b>	Alumno.		

Tabla 17 UR-C-009

<b>Identificador: UR-C-009</b>			
<b>Descripción</b>	Crear una tienda web a modo de catálogo para ser visible por los usuarios.		
<b>Necesidad</b>	<input checked="" type="checkbox"/> Esencial	<input type="checkbox"/> Deseable	<input type="checkbox"/> Opcional
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Estabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Fuente</b>	Alumno.		

Tabla 18 UR-C-011

<b>Identificador: UR-C-010</b>			
<b>Descripción</b>	Permitir a usuarios registrados publicar comentarios en la plataforma web.		
<b>Necesidad</b>	<input checked="" type="checkbox"/> Esencial	<input type="checkbox"/> Deseable	<input type="checkbox"/> Opcional
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Estabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Fuente</b>	Alumno.		

#### 4.1.1.2. Requisitos de restricción

En el siguiente apartado se muestra los requisitos de restricción del sistema.

Tabla 19 UR-R-001

<b>Identificador: UR-R-001</b>			
<b>Descripción</b>	Autenticación en la plataforma web para usar la funcionalidad. Todos los usuarios.		
<b>Necesidad</b>	<input checked="" type="checkbox"/> Esencial	<input type="checkbox"/> Deseable	<input type="checkbox"/> Opcional
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Estabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Fuente</b>	Alumno.		

Tabla 20 UR-R-002

<b>Identificador: UR-R-002</b>			
<b>Descripción</b>	Dotar de seguridad a la autenticación en la plataforma web.		
<b>Necesidad</b>	<input checked="" type="checkbox"/> Esencial	<input type="checkbox"/> Deseable	<input type="checkbox"/> Opcional
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Estabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Fuente</b>	Alumno.		

Tabla 21 UR-R-003

<b>Identificador: UR-R-003</b>			
<b>Descripción</b>	Dotar de seguridad en el registro de alta en la plataforma web.		
<b>Necesidad</b>	<input checked="" type="checkbox"/> Esencial	<input type="checkbox"/> Deseable	<input type="checkbox"/> Opcional
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Estabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Fuente</b>	Alumno.		

Tabla 22 UR-R-004

Identificador: UR-R-004			
<b>Descripción</b>	Denegar el acceso a la zona de administración de la plataforma web a todo usuario sin privilegios.		
<b>Necesidad</b>	<input checked="" type="checkbox"/> Esencial	<input type="checkbox"/> Deseable	<input type="checkbox"/> Opcional
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Estabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Fuente</b>	Alumno.		

#### 4.1.2. DIAGRAMA DE CASOS DE USO

En el siguiente apartado se van a mostrar los diagramas de casos de uso junto con su especificación. Para la especificación se usará una tabla con los siguientes atributos:

- Identificador: Identifica al caso de uso de forma única. Debe seguir el formato *CU-A-XXX*, siendo *A* la inicial del actor y *XXX* un valor número único para cada caso de uso.
- Nombre: Breve especificación textual del caso de uso.
- Actores: Tipo de usuarios del sistema que inicia el caso de uso.
- Objetivo: Finalidad del caso de uso.
- Precondiciones: Estado previo que se debe cumplir para poder realizar la operación.
- Postcondiciones: Estado en el que queda el sistema tras realizar una operación.
- Escenario básico: Especifica la manera en la que interactúa un actor con el sistema y cuál es la respuesta que el sistema le ofrece.
- Escenarios alternativos: Condiciones excepcionales que afectan al escenario y respuestas del sistema antes estas situaciones.

El proyecto tiene dos partes claramente diferencias, por un lado el algoritmo de búsqueda y por otro la plataforma web dónde mostrar los resultados. Además, uno de los requisitos es la existencia de, al menos, tres tipos distintos de usuarios (invitado, registrado y administrador). A continuación, se van a detallar los casos de uso según los distintos actores participantes en el sistema.

##### 4.1.2.1. Invitado

El usuario invitado será aquel usuario que no esté identificado en la plataforma web. Al no estar registrado, el usuario invitado podrá identificarse en la plataforma web para acceder a su funcionalidad, darse de alta como nuevo usuario registrado y visualizar las noticias. En la siguiente imagen, se muestra el diagrama de casos de uso para el usuario invitado.

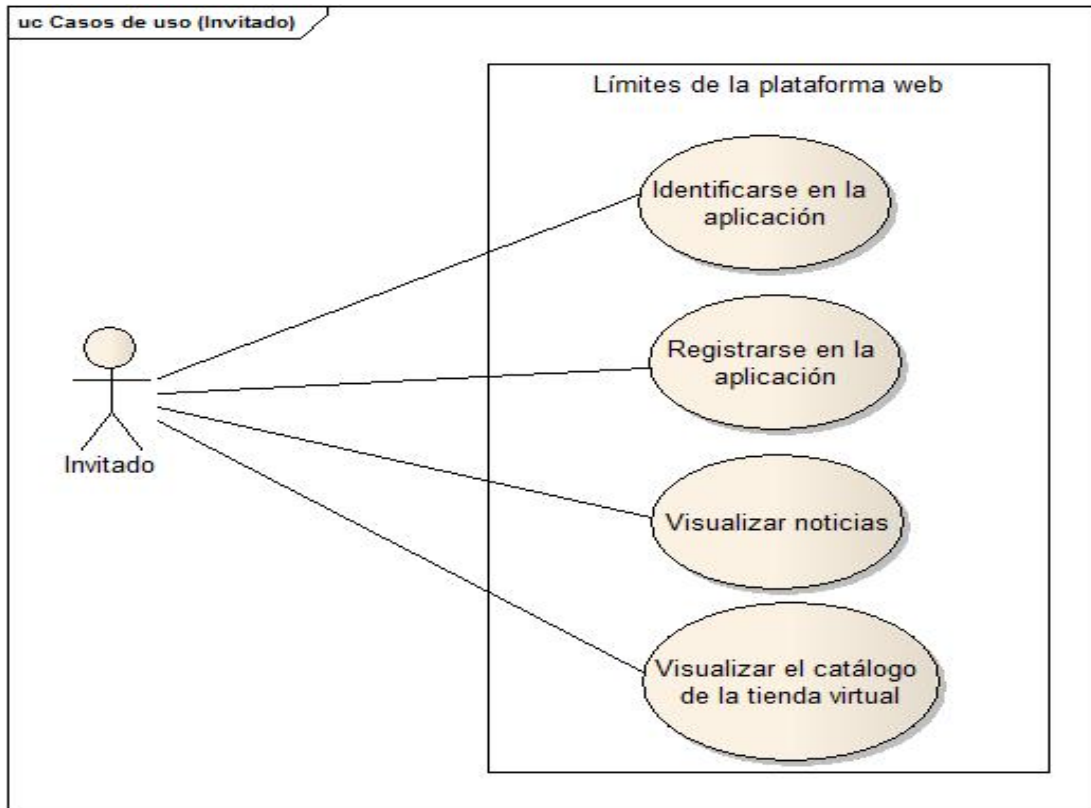


Ilustración 16 Diagrama de casos de uso: Invitado

La descripción de los casos de uso es la siguiente:

Tabla 23 CU-I-001

<b>Identificador: CU-I-001</b>	
<b>Nombre</b>	Identificarse en la aplicación.
<b>Actores</b>	Invitado.
<b>Objetivo</b>	Identificarse en la plataforma web.
<b>Precondiciones</b>	Ninguna.
<b>Postcondiciones</b>	Usuario identificado.
<b>Escenario básico</b>	<ol style="list-style-type: none"> <li>1. El actor accede a la plataforma web.</li> <li>2. El actor introduce sus credenciales de acceso.</li> <li>3. La plataforma le da la bienvenida.</li> </ol>
<b>Escenario alternativo</b>	Ninguno.

Tabla 24 CU-I-002

<b>Identificador: CU-I-002</b>	
<b>Nombre</b>	Registrarse en la aplicación.
<b>Actores</b>	Invitado
<b>Objetivo</b>	Dar de alta un nuevo usuario registrado
<b>Precondiciones</b>	No estar registrado.
<b>Postcondiciones</b>	Usuario dado de alta en la plataforma.
<b>Escenario básico</b>	<ol style="list-style-type: none"> <li>1. El actor accede a la plataforma web.</li> <li>2. El actor rellena un formulario con los datos de registro.</li> <li>3. La plataforma envía un correo electrónico para la confirmación del registro.</li> </ol>
<b>Escenario alternativo</b>	Ninguno.

Tabla 25 CU-I-003

<b>Identificador: CU-I-003</b>	
<b>Nombre</b>	Visualizar noticias.
<b>Actores</b>	Invitado
<b>Objetivo</b>	Poder ver las noticias recogidas en la plataforma.
<b>Precondiciones</b>	No estar registrado.
<b>Postcondiciones</b>	Ninguna.
<b>Escenario básico</b>	<ol style="list-style-type: none"> <li>1. El actor accede a la plataforma web.</li> <li>2. El actor elige la opción adecuada para ver las noticias.</li> </ol>
<b>Escenario alternativo</b>	Ninguno.

Tabla 26 CU-I-004

<b>Identificador: CU-I-004</b>	
<b>Nombre</b>	Visualizar el catálogo de la tienda virtual.
<b>Actores</b>	Invitado.
<b>Objetivo</b>	Visualizar el catálogo de la tienda virtual.
<b>Precondiciones</b>	Ninguna.



<b>Postcondiciones</b>	Ninguna.
<b>Escenario básico</b>	1. El actor accede a la plataforma web. 2. El actor accede a la tienda virtual..
<b>Escenario alternativo</b>	Ninguno.

#### 4.1.2.2. Usuario registrado

El usuario registrado será aquel usuario que se encuentre identificado y, en consecuencia, disponga de privilegios para poder disfrutar de más funcionalidades en la plataforma web.

A un usuario registrado, a parte de la visualización de las noticias, se le permite ser partícipe de la plataforma web pudiendo dejar comentarios en una sección habilitada para ello.

De este modo, este tipo de actores, dispone de dos casos de uso como se muestra a continuación en el siguiente diagrama de casos de uso:

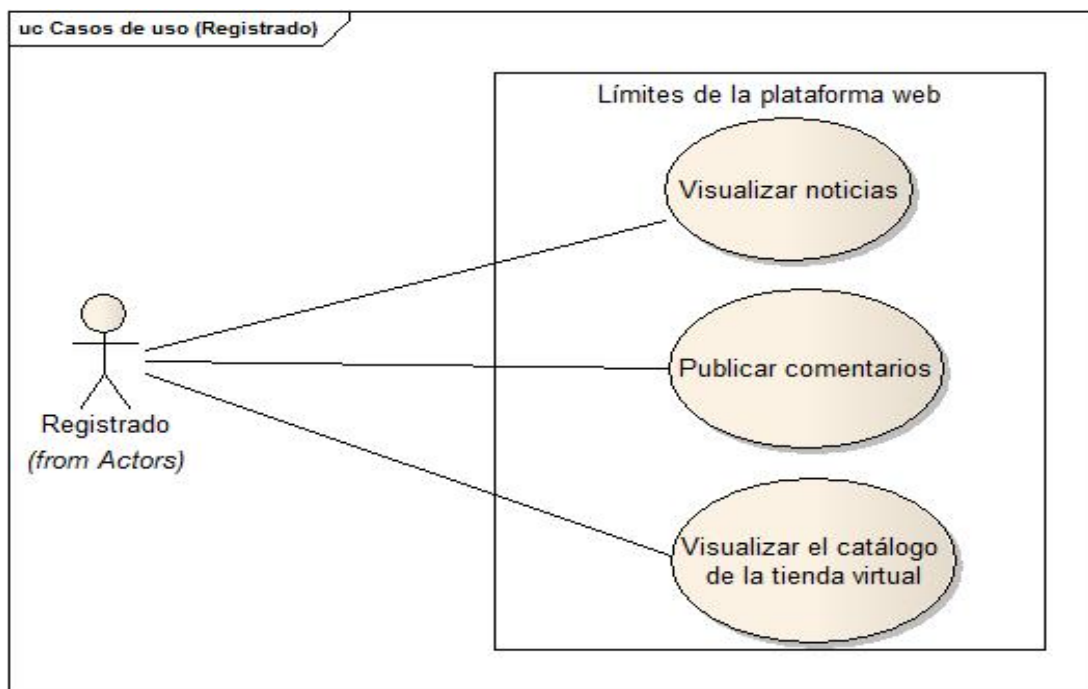


Ilustración 17 Diagrama de casos de uso: Registrado

La especificación de estos casos de uso es la siguiente:

Tabla 27 CU-R-001

<b>Identificador: CU-R-001</b>	
<b>Nombre</b>	Visualizar noticias.
<b>Actores</b>	Registrado.

<b>Objetivo</b>	Poder ver las noticias recogidas en la plataforma.
<b>Precondiciones</b>	Registrado.
<b>Postcondiciones</b>	Ninguna.
<b>Escenario básico</b>	<ol style="list-style-type: none"> <li>1. El actor accede a la plataforma web.</li> <li>2. El actor elige la opción adecuada para ver las noticias.</li> </ol>
<b>Escenario alternativo</b>	Ninguno.

Tabla 28 CU-R-002

<b>Identificador: CU-R-002</b>	
<b>Nombre</b>	Publicar comentarios.
<b>Actores</b>	Registrado.
<b>Objetivo</b>	Publicar comentarios en la plataforma.
<b>Precondiciones</b>	Ser usuario registrado.
<b>Postcondiciones</b>	Ninguna.
<b>Escenario básico</b>	<ol style="list-style-type: none"> <li>1. El actor accede a la plataforma web.</li> <li>2. El actor realiza un comentario en la zona habilitada.</li> </ol>
<b>Escenario alternativo</b>	Ninguno.

Tabla 29 CU-R-003

<b>Identificador: CU-R-003</b>	
<b>Nombre</b>	Visualizar el catálogo de la tienda virtual.
<b>Actores</b>	Registrado.
<b>Objetivo</b>	Visualizar el catálogo de la tienda virtual.
<b>Precondiciones</b>	Ser usuario registrado.
<b>Postcondiciones</b>	Ninguna.
<b>Escenario básico</b>	<ol style="list-style-type: none"> <li>1. El actor accede a la plataforma web.</li> <li>2. El actor accede a la tienda virtual..</li> </ol>
<b>Escenario alternativo</b>	Ninguno.

### 4.1.2.3. Administrador

El usuario administrador será aquel usuario con poder absoluto tanto en el algoritmo de búsqueda como en la plataforma web.

En este caso, el usuario administrador tiene todos los privilegios sobre el algoritmo como la plataforma. De modo que el diagrama de casos de uso queda como se aprecia en la siguiente figura:

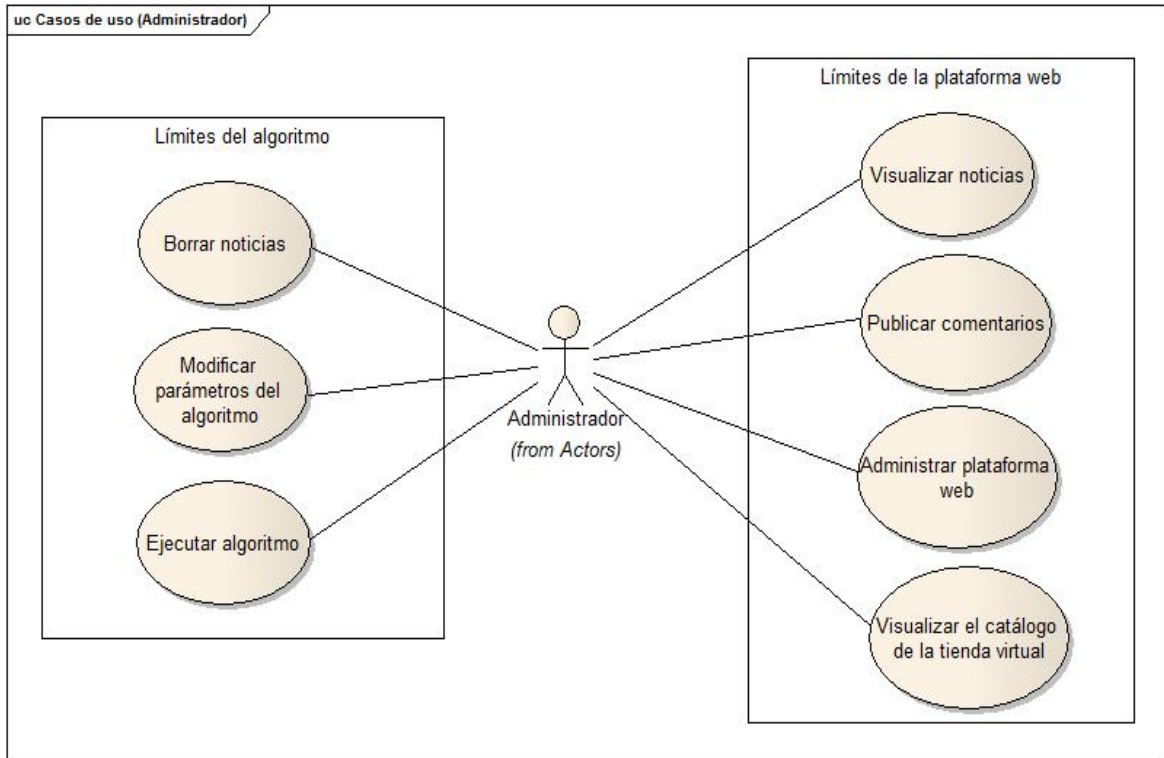


Ilustración 18 Diagrama de casos de uso: Administrador

La especificación de estos casos de uso es la siguiente

Tabla 30 CU-A-001

Identificador: CU-A-001	
<b>Nombre</b>	Borrar noticias.
<b>Actores</b>	Administrador.
<b>Objetivo</b>	Eliminar noticias incorrectas.
<b>Precondiciones</b>	Rol de administrador.
<b>Postcondiciones</b>	Ninguna.
<b>Escenario básico</b>	<ol style="list-style-type: none"> <li>1. El actor accede a la base de datos.</li> <li>2. El actor elimina las noticias incorrectas.</li> </ol>
<b>Escenario alternativo</b>	Ninguno.

Tabla 31 CU-A-002

<b>Identificador: CU-A-002</b>	
<b>Nombre</b>	Modificar parámetros del algoritmo.
<b>Actores</b>	Administrador.
<b>Objetivo</b>	Cambiar los parámetros de búsqueda del algoritmo.
<b>Precondiciones</b>	Rol de administrador.
<b>Postcondiciones</b>	Ninguna.
<b>Escenario básico</b>	<ol style="list-style-type: none"> <li>1. El actor accede a la base de datos.</li> <li>2. El actor modifica los parámetros de búsqueda.</li> </ol>
<b>Escenario alternativo</b>	Ninguno.

Tabla 32 CU-A-003

<b>Identificador: CU-A-003</b>	
<b>Nombre</b>	Ejecutar el algoritmo.
<b>Actores</b>	Administrador.
<b>Objetivo</b>	Ejecutar el algoritmo de manera manual.
<b>Precondiciones</b>	Rol de administrador.
<b>Postcondiciones</b>	Ninguna.
<b>Escenario básico</b>	<ol style="list-style-type: none"> <li>1. El actor inicia de forma manual el algoritmo.</li> </ol>
<b>Escenario alternativo</b>	Ninguno.

Tabla 33 CU-A-004

<b>Identificador: CU-A-004</b>	
<b>Nombre</b>	Visualizar noticias.
<b>Actores</b>	Administrador.
<b>Objetivo</b>	Poder ver las noticias recogidas en la plataforma.
<b>Precondiciones</b>	Registrado.
<b>Postcondiciones</b>	Ninguna.
<b>Escenario básico</b>	<ol style="list-style-type: none"> <li>1. El actor accede a la plataforma web.</li> </ol>

	2. El actor elige la opción adecuada para ver las noticias.
<b>Escenario alternativo</b>	Ninguno.

Tabla 34 CU-A-005

<b>Identificador: CU-A-005</b>	
<b>Nombre</b>	Publicar comentarios.
<b>Actores</b>	Administrador.
<b>Objetivo</b>	Publicar comentarios en la plataforma.
<b>Precondiciones</b>	Ser usuario registrado.
<b>Postcondiciones</b>	Ninguna.
<b>Escenario básico</b>	<ol style="list-style-type: none"> <li>1. El actor accede a la plataforma web.</li> <li>2. El actor realiza un comentario en la zona habilitada.</li> </ol>
<b>Escenario alternativo</b>	Ninguno.

Tabla 35 CU-A-006

<b>Identificador: CU-A-006</b>	
<b>Nombre</b>	Administrar plataforma web.
<b>Actores</b>	Administrador.
<b>Objetivo</b>	Administrar la plataforma web y realizar posibles cambios.
<b>Precondiciones</b>	Rol de administrador.
<b>Postcondiciones</b>	Ninguna.
<b>Escenario básico</b>	<ol style="list-style-type: none"> <li>1. El actor accede a la zona de administración de la plataforma con sus credenciales.</li> <li>2. El actor modifica la plataforma.</li> </ol>
<b>Escenario alternativo</b>	Ninguno.

Tabla 36 CU-R-007

<b>Identificador: CU-R-007</b>	
<b>Nombre</b>	Visualizar el catálogo de la tienda virtual.
<b>Actores</b>	Administrador.
<b>Objetivo</b>	Visualizar el catálogo de la tienda virtual.
<b>Precondiciones</b>	Rol de administrador.
<b>Postcondiciones</b>	Ninguna.
<b>Escenario básico</b>	<ol style="list-style-type: none"> <li>1. El actor accede a la plataforma web.</li> <li>2. El actor accede a la tienda virtual..</li> </ol>
<b>Escenario alternativo</b>	Ninguno.

## 4.2. DISEÑO DE LA PLATAFORMA WEB

Para el diseño de la plataforma se han tenido en cuenta los objetivos del proyecto. El objetivo principal es mostrar todas esas noticias que recopila el algoritmo por lo que deberá existir una sección denominada noticias y, obviamente, deberá estar ordenada por fecha de publicación. Además, dado que se conoce la procedencia de la noticia se va a crear una sección para poder filtrar las noticias según sea la procedencia, en este caso, la procedencia es la URL o RSS del que procede, o sea, la federación deportiva del deporte en cuestión.

A partir de aquí, toda la funcionalidad añadida son extras que se añaden para convertir la página en una gran plataforma global del deporte femenino.

### 4.2.1. MÓDULOS PARA MEJORAR LA USABILIDAD DE LA PLATAFORMA WEB

#### 4.2.1.1. Multidioma

Dado que el objeto de búsqueda son mujeres deportistas españolas y la pluralidad de idiomas en España se opta por realizar la plataforma en formato multidioma, de modo que el usuario pueda elegir el idioma mediante el componente JoomFish. En la actualidad se ha configurado para castellano y catalán, como idiomas oficiales de España, y se ha añadido el inglés para dotar de internacionalidad a la plataforma. Además, cabe recordar que el algoritmo de búsqueda es flexible, por lo que se puede adoptar a cualquier idioma en la búsqueda de noticias. El usuario verá en la esquina superior izquierda de la pantalla lo siguiente:



Ilustración 19 Multidioma

Este complemento actualmente se encuentra desactivado en espera de asentar la página en un hosting y dominio adecuados. Para activarlo, sería necesario la inclusión de los scripts de traducción [véase Apéndice G].

#### 4.2.1.2. Cambio del tamaño de fuente

Para mejorar la usabilidad de la plataforma y en beneficio de personas con deficiencias visuales, se ha incluido el módulo Fontsize para que se pueda modificar el tamaño de la letra en modo online por parte del usuario. Este módulo se encuentra en la esquina superior derecha, justo al lado del selector de idioma.



Ilustración 20 Cambio del tamaño de fuente

Además, en la parte inferior derecha se ha colocado un enlace para volver al principio de la página de una manera automática sin tener que usar el scroll o ascensor del navegador.



Ilustración 21 Ascensor automático

#### 4.2.1.3. Gestor de Eventos

Para que los visitantes puedan estar informados de los próximos eventos del deporte femenino a nivel mundial se ha incluido una sección denominada Eventos que consiste en un calendario con todas esas fechas del deporte femenino importantes a lo largo del año. Para ello, se ha necesitado el componente JEvents. El usuario, en la parte derecha verá el siguiente menú:



Ilustración 22 Calendario de eventos

#### 4.2.1.4. Login

Un paso más para crear una plataforma web, es la creación de una comunidad web para ello, existe la posibilidad de registrarse en la plataforma. El usuario verá en el menú principal una opción denominada *Usuario* que desplegará un submenú con la opción *Abrir/Cerrar sesión* que le llevará al formulario de identificación y la posibilidad de registrarse en la plataforma o recordar sus datos personales.

En la siguiente imagen se puede apreciar esta opción del menú:

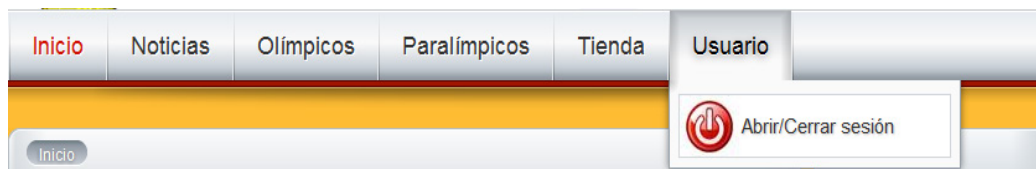


Ilustración 23 Opción del menú para registrarse

Por otro lado, en la siguiente imagen se ilustra el módulo de registro, dónde el usuario podrá logarse, recordar tanto usuario como contraseña recibiendo un correo electrónico y crear una cuenta nueva pinchando en la opción *Regístrese*.

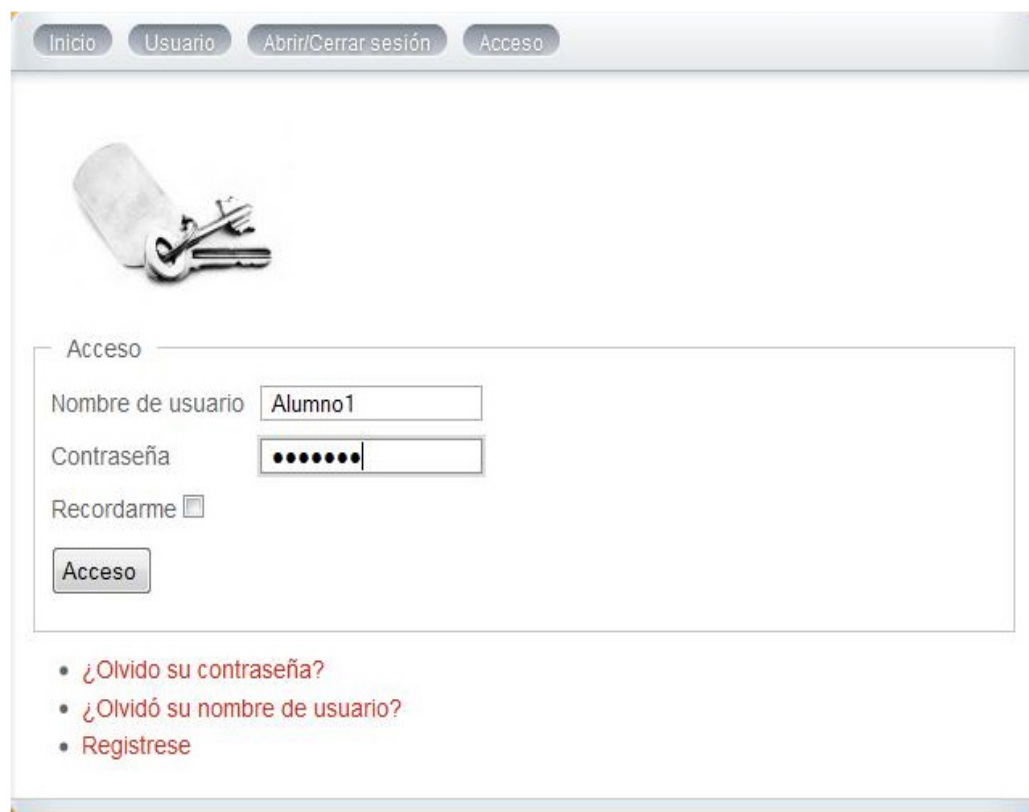


Ilustración 24 Módulo de registro

Cabe destacar que Joomla! trae implementado tres tipos distintos de autenticación en la plataforma con su respectivo módulo:

- Sobre la base de datos de la plataforma.
- Sobre un cuenta de Google.
- Sobre un directorio de direcciones tipo LDAP.
- Sobre un directorio tipo OpenID

En este caso, la opción elegida ha sido la primera, de modo que es la propia plataforma quién almacena en su base de datos la información requerida por parte de los usuarios.

Además, tiene un gestor de roles de los usuarios, permitiendo habilitar distintas funciones según el rol del usuario. Existen roles tanto para el front-end de la



plataforma como del back-end. Como se verá más adelante, se aprovecharán estos distintos tipos de roles para habilitar o deshabilitar, publicar o no publicar determinada información según el rol del usuario.

Los roles que utiliza Joomla!, ordenados de menos a más privilegios, son:

- Front-End:
  - Invitado: usuario no registrado que solo puede ver la parte pública de la plataforma
  - Registrado: usuario que se registra y puede ver los contenidos de la zona privada.
  - Autor: usuario que es capaz de crear artículos o contenidos que quedan pendientes de evaluación por parte de un usuario con rol superior.
  - Editor: usuario que, además de las funciones de autor, puede revisar artículos o contenidos de otros autores sin poder publicar.
  - Publicador o Editor Jefe: usuario que, además de todas las funciones anteriores, autoriza las publicaciones en la plataforma.
- Back-End:
  - Gestor: tiene acceso al gestor multimedia, al gestor de contenidos, la ayuda del sistema y la edición de menús existentes y creación de elementos del menú.
  - Administrador: a parte de las funciones del gestor, tiene permisos para administrar usuarios (salvo el del superadministrador), crear y editar los menús, gestionar el contenido (papelera y estadísticas) e instalar y/o desinstalar componentes, extensiones y/o módulos salvo las plantillas y el administrador de idiomas.
  - Superadministrador: tiene todos los poderes sobre la plataforma. Es el único con capacidad de manipular usuarios, registros y permisos.

De esto modo, la gestión de la plataforma se hace muy sencilla y ofrece posibilidades de delegación en ayudantes.

Cabe destacar, que algunos módulos adoptan su propia jerarquía de roles simplificando los de Joomla! en invitado, registrado, especial y administrador.

#### 4.2.1.5. Shoutbox o Bandeja de gritos

Con este registro que se acaba de comentar, se permite el acceso a otras funcionalidades, como puede ser un foro o la bandeja de gritos creada en la parte izquierda de la pantalla bajo el nombre de Shoutbox, de modo que si no se está **registrado** es imposible escribir en dicha bandeja. Cabe destacar que administrar esta bandeja de gritos no requiere tanto esfuerzo como puede ser el uso de un foro. Para ayudar a administrar este módulo, se ha dado poderes al rol de **publicador** en adelante, a modo de *Community Manager*.

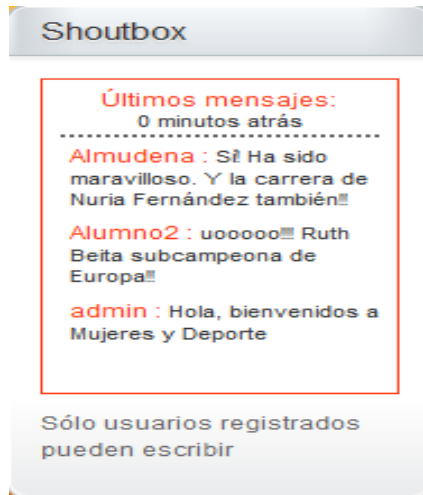


Ilustración 25 Módulo shoutbox

#### 4.2.1.6. Foro

En el caso de foro, se muestra como se ha integrado en la plataforma pero está deshabilitado. El uso o no del foro va a depender del mantenimiento de la plataforma, puesto que la administración de un foro es algo que requiere mucho tiempo y mucho esfuerzo. Se muestra en la plataforma como posibilidad o alternativa a la bandeja de gritos, solo accesible y visible si se está registrado. Para la implementación del foro se ha usado el módulo Kunena.



Ilustración 26 Módulo foro

Este módulo se encuentra actualmente oculto en la página web para todo tipo de usuarios. En caso de querer activarlo, tan solo sería necesario marcarlo como visible a usuarios de tipo *Registrado*. Cabe destacar que el uso de un foro implica la existencia de nuevos roles del tipo *Moderador*.

#### 4.2.1.7. Estadísticas

Para poder tener un seguimiento del tráfico generado en la red, para hacer esa labor de *traffic analyst* tan de moda en los últimos años, se ha creado una sección de estadísticas con información de las visitas aprovechando JoomStats. Además, en el back-end también se muestra en la pantalla principal para un seguimiento del súper-administrador de la plataforma.

A continuación, se muestra el módulo de estadísticas. Este módulo solo está visible para los administradores y usuarios especiales.



Estadísticas 📊	
Visitantes hoy:	0
Visitantes ayer:	0
Visitantes del mes:	1
Visitantes mes anterior:	0
Visitantes del año:	1
Visitantes del año anterior:	0
Visitantes totales:	1
Máx. visitantes al día:	1
Día de más visitas:	02-03-2011
Máx. visitantes al mes:	1
Mes de más visitas:	03-2011
Impresiones hoy:	0
Impresiones ayer:	0
Impresiones este mes:	10
Impresiones totales:	10
Bots hoy:	0
Activo desde:	02-03-2011
Navegadores Web:	
100% FireFox 3.6.14	

Ilustración 27 Módulo de estadísticas

#### 4.2.1.8. FAQ

Ante la posibilidad de delegar las funciones en otras personas con desconocimiento sobre la plataforma, se ha creado una sección denominada FAQ (Frequently Asked Questions) sobre problemas posibles y cómo contactar con el administrador de la plataforma para futuros cambios en ella. A continuación, se muestra el módulo de FAQ que estará visible para los administradores y usuarios especiales.

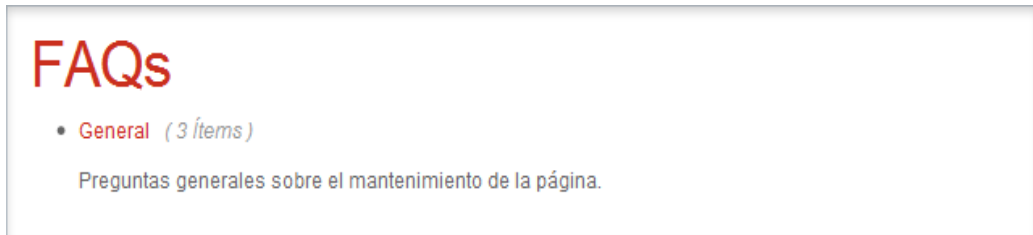


Ilustración 28 Módulo de FAQ

#### 4.2.1.9. Tienda electrónica

Otro de las posibilidades que se ofrece, es la creación de una tienda virtual electrónica. Para la presentación del proyecto, ha sido deshabilitada todo tipo de compra por lo que se encuentra en modo catálogo, esto es, visualización de los productos. Para ello se ha utilizado el componente VirtueMart. Como ejemplo, se han realizado una serie de productos relacionados con la temática de la mujer y el deporte.

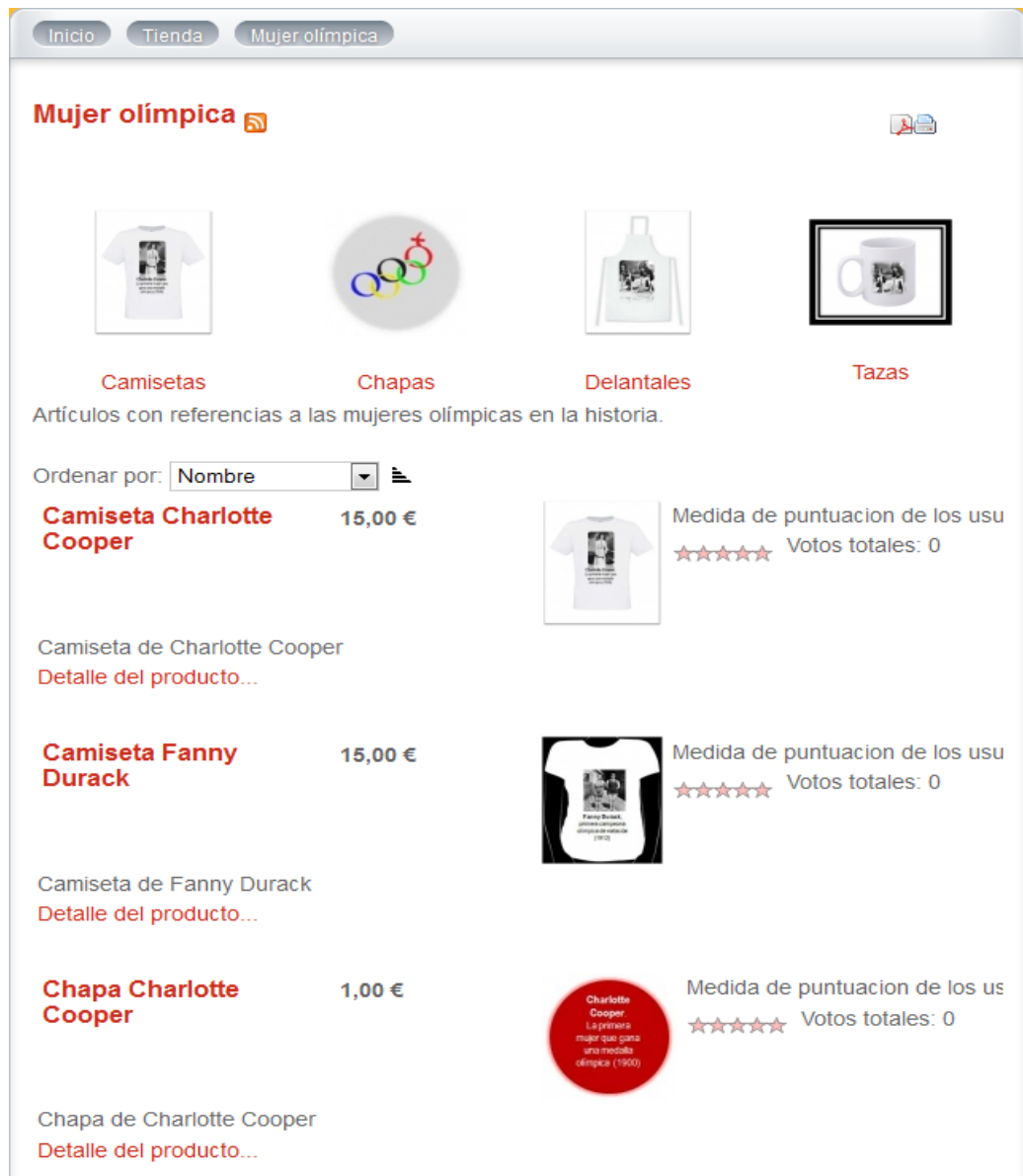


Ilustración 29 Tienda virtual

#### 4.2.1.10. Mapa del sitio Web

Para la realización del sitio web se ha utilizado el módulo Xmap que se integra a la perfección con toda la plataforma, incluyendo la tienda electrónica incluso con los debates existentes que pudieran existir en el foro. Se muestra un pequeño trozo del resultado del mapa web.



Ilustración 30 Mapa del sitio web

El enlace al mapa web está localizado en la parte inferior central de la página, justo encima del pie de la página con el Copyright y los derechos de autor. Cabe destacar que este módulo ayuda a una mejora del posicionamiento en los buscadores tipo Google, Yahoo!...

### 4.2.2. PORTADA

Una vez visto estos pequeños módulos existentes en la plataforma web, se pasa a mostrar de verdad la plataforma. Lo primero, es observar cuál es la primera imagen que recibe el usuario al acceder a la plataforma, algo importante puesto que se debe causar una buena impresión; eso se hace mediante una bienvenida como se ilustra en la siguiente imagen.

Lunes, 11 Julio 2011

A A a

WEB AUTOMATIZADA: MUJER OLÍMPICA

Inicio Noticias Olímpicos Paralímpicos **Tienda** Usuario

Inicio

## Bienvenid@s

Bienvenid@s a la página web del proyecto **PLATAFORMA WEB AUTOMATIZADA: Mujer Olímpica 2012**.

Con este proyecto se pretende dar un impulso al deporte de la mujer en España. Para ello, en esta página web se podrán encontrar los últimos logros conseguidos por nuestras deportistas españolas a nivel nacional e internacional a lo largo de su preparación para los próximos Juegos Olímpicos que se celebrarán en la ciudad inglesa de Londres en el verano de 2012.

Este página no necesita de un webmaster, se encuentra totalmente automatizada. El motor de búsqueda realizado solo muestra los logros de nuestras deportistas olímpicas. Si algún usuario conociera de alguna deportista cuyo logro no se refleje en nuestras noticias, por favor notifíquelo a través del *shoutbox*.

Este proyecto ha sido realizado por Richard Vega Caridad y dirigido por Almudena Alcaide Raya en colaboración con la Universidad Carlos III de Madrid.

Últimos mensajes:  
5 días, 4 horas atrás  
Bloke : Ya está habilitado el registro. Una vez os registréis recibiréis un correo electrónico con un enlace de confirmación.  
Almudena : Sabemos que existen problemas con el registro. Se solucionarán lo antes posible. Mientras tanto no dejéis de leer los logros

Sólo usuarios registrados pueden escribir

Eventos

Julio 2011						
L	M	X	J	V	S	D
27	28	29	30	1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

Agradecimientos | Copyright | **Mapa del sitio web**  
Desarrollo: Richard Vega Caridad.

Dirección del proyecto: Almudena Alcaide Raya - Universidad Carlos III de Madrid.  
Copyright © 2011 Mujer Olímpica 2012. Todos los derechos reservados.  
Joomla! es un software libre publicado bajo la licencia GNU/GPL.

Ilustración 31 Bienvenida

En la imagen anterior, se pueden apreciar todos los módulos comentados hasta ahora que sean públicos: selección de idioma, cambio del tamaño de fuente, shoutbox, calendario de eventos, tienda electrónica y mapa del sitio web. Además, se observa en el pie de la página

el Copyright y los derechos de autor, y un menú principal que habilita las distintas funciones, este menú variará según el rol que disponga el usuario.

### 4.2.3. MENÚ PRINCIPAL (Invitado)

A continuación, se van a ir mostrando las distintas opciones existentes en el menú principal de la plataforma. La primera opción es la de NOTICIAS que dispone de tres elecciones, las noticias de HOY, las noticias de AYER y las noticias del resto de días ANTERIORES.



Ilustración 32 Menú NOTICIAS

La siguiente opción, es OLÍMPICOS, dónde se despliega un menú con las distintas federaciones de las que se obtienen las noticias.



Ilustración 33 Menú OLÍMPICOS

La siguiente opción, es PARALÍMPICOS, que de modo análogo a OLÍMPICOS despliega un menú con las distintas federaciones de las que se obtiene las noticias.



Ilustración 34 Menú PARALÍMPICOS

La última opción pública TIENDA, direcciona la página al módulo de la tienda virtual comentado anteriormente.

#### 4.2.4. MENÚ REGISTRADO

En el caso de usuario registrado en el menú principal aparece la opción USUARIO que contiene una única posibilidad, TU PERFIL.

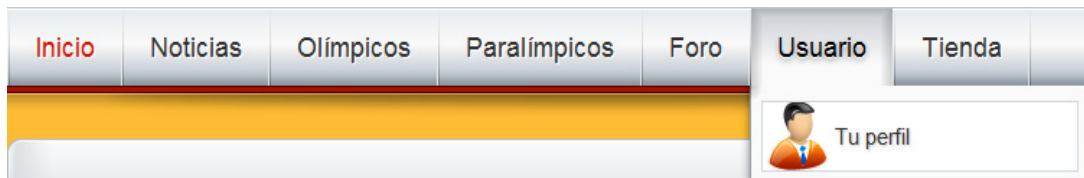


Ilustración 35 Menú principal usuario registrado

La opción de TU PERFIL muestra la siguiente información:

Modificar sus detalles

Nombre de usuario: Alumno2

Su nombre:

E-mail:

Contraseña:

Verifique su contraseña:

Idioma para la parte frontal: Español (spanish formal Internacional) ▼

Zona horaria: (UTC +01:00) Amsterdam, Berlin, Brussels, Copenhagen, Madrid, París, horario de invierno Es

Ilustración 36 Opción TU PERFIL



#### 4.2.5. MENÚ ESPECIAL/ADMINISTRADOR

Las opciones especiales del menú principal que aparecen en este rol son la FAQ y las opciones correspondientes a los módulos de crear eventos y ver las estadísticas. Los módulos de FAQ y estadísticas ya han sido comentados y mostrados anteriormente.

El menú queda de la siguiente forma:



Ilustración 37 Menú principal usuario ADMIN

La opción INSERTAR EVENTO proporciona al usuario gestionar el calendario de los eventos a realizar. Al pinchar sobre esta opción aparece en pantalla un calendario que tiene el siguiente aspecto:

	Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo	
9	28	1 Mundial de Esqu ... Torneo WTA	2 Mundial de Esqu ... Torneo WTA	3 Mundial de Biat ... Mundial de Esqu ... Torneo WTA	4 Europeo indoor ... Mundial de Biat ... Mundial de Esqu ... Torneo WTA	5 Europeo indoor ... Mundial de Biat ... Mundial de Esqu ... Torneo WTA	6 Campeonato de E ... Carrera de la m ... Europeo indoor ... Mundial de Biat ... Mundial de Esqu ... Torneo WTA	
	10	7 Indian-Wells Mundial de Biat ...	8 Día de la mujer ... Europeo de salt ... Indian-Wells Mundial de Biat ...	9 Europeo de salt ... Indian-Wells Mundial de Biat ...	10 Europeo de salt ... Indian-Wells Mundial de Biat ...	11 Europeo de salt ... Indian-Wells Mundial de Biat ...	12 Campeonato de E ... Europeo de salt ... Indian-Wells Mundial de Biat ...	13 Europeo de salt ... Indian-Wells Mundial de Biat ...
		11	14 Indian-Wells	15 Indian-Wells	16 Indian-Wells	17 Eurocopa femeni ... Indian-Wells Mundial de fond ...	18 Copa del Mundo ... Indian-Wells Mundial de fond ...	19 Copa del Mundo ... Final Four Indian-Wells Mundial de fond ...
12			21 Europeo indoor ... Mundial de Pati ... Torneo WTA	22 Europeo indoor ... Mundial de Pati ... Torneo WTA	23 Europeo indoor ... Mundial de Cicl ... Torneo WTA	24 Eurocopa femeni ... Europeo indoor ... Mundial de Cicl ... Torneo WTA	25 Copa de la Rein ... Europeo indoor ... Mundial de Pati ... Torneo WTA	26 Copa de la Rein ... Europeo indoor ... Mundial de Pati ... Torneo WTA
	13		28 Europeo de Luch ...	29 Europeo de Luch ...	30 Europeo de Luch ...	31 Europeo de Luch ...	1	2

Añadir evento

Ilustración 38 Módulo gestionar calendario

Con la opción multidioma todos estos menús mostrados anteriormente se traducen automáticamente al idioma seleccionado.

#### 4.2.6. VISUALIZACIÓN DE LAS NOTICIAS

Para la visualización de las noticias se realiza mediante la ejecución de una consulta a la base de datos y se representa en una tabla todo ello a través de SCRIPT de pHp [véase Apéndice G] combinado con HTML y MySQL. Para realizar esto en Joomla!<sup>44</sup> se aprovecha el plugin Jumi.

<sup>44</sup> [MAR+11]

Los datos que se muestran en la tabla son la fecha de publicación, el titular de la noticia y en caso de no ser específico de una federación se muestra a que federación pertenece. Además, se muestran tres iconos para poder mostrar y ocultar el contenido de la noticia si lo tuviera, y poder navegar a la página web de la noticia.

Un ejemplo de las noticias es:










10-08-2010	<b>Marina García, a semifinales en los 100 braza</b>	Natación	  
10-08-2010	<b>Y Duane remata el oro con un ¡bronce!</b> Duane da Rocha. No hace mucho podía haber dejado la natación, pero dentro sentía una llamada que le decía que algún día tendría su recompensa. Ese día ha sido esta tarde más allá de las seis, cuando la española ha tocado la pared en la tercera plaza, ¡bronce! ¡bronce europeo! Una carrera muy dura, con las inglesas como motos y con otras cinco nadadoras en busca de una medalla que viaja a España, a Málaga, a Mijas,... España, dos de dos hoy, oro y bronce. Un gran día para todos.	Natación	  
19-03-2010	<b>Bárbara Hernando, récord de España de pentatlón (4.381)</b>	Atletismo	  

Ilustración 39 Ejemplo de noticias en castellano

Como se puede apreciar en la figura anterior, inicialmente se muestra tan solo el titular, y a la derecha aparecen tres iconos, el primero de ellos un símbolo de suma (+) para expandir la noticia y que aparezca el contenido de la misma como tal y como se muestra la figura. El segundo de ellos es un símbolo de resta (-) para contraer el contenido de la noticia. Y por último, una flecha hacia la derecha que al pulsar sobre ella abrirá una nueva ventana al usuario para navegar hasta la página web dónde se encuentra la noticia publicada. Esta funcionalidad se hace mediante unas pequeñas funciones en Javascript [véase Apéndice G]. No se encuentra habilitado en todas las noticias debido a los derechos de autor de las fuentes de información.



Ilustración 40 Iconos de las noticias

Al tener la página en multidioma y tener la base de datos en castellano, a la hora de obtener las páginas en otro idioma es necesario añadir un pequeño script intermedio antes de mostrar la tabla en pantalla. Este script [véase Apéndice G] consiste en pasar la noticia por el API de Google Translator para que devuelva la noticia en el idioma requerido.

A continuación, se va a mostrar un ejemplo completo de la opción multidioma, poniendo la página en otro lenguaje, mostrando los menús y viendo el resultado de ejecución de una de las federaciones.



Ilustración 41 Ejemplo de noticias en catalán

Otro ejemplo, en inglés sería:



Ilustración 42 Ejemplo de noticias en inglés

#### 4.2.7. VERSIÓN PARA DISPOSITIVOS MÓVILES

Con el aumento de las tecnologías móviles y el uso de los denominados Smartphone, se ha implementado versiones para este tipo de dispositivos. Su funcionamiento es el siguiente: la página web detectará cuál es el tipo de dispositivo del visitante PC, plataforma MAC, otra tipo de Smartphone (Symbian, Blackberry) y cargará la plantilla acorde al dispositivo.

A continuación, se muestra las imágenes correspondientes a la plataforma MAC [véase Ilustración 43] y la correspondiente a otro tipo de Smartphone [véase Ilustración 44]:

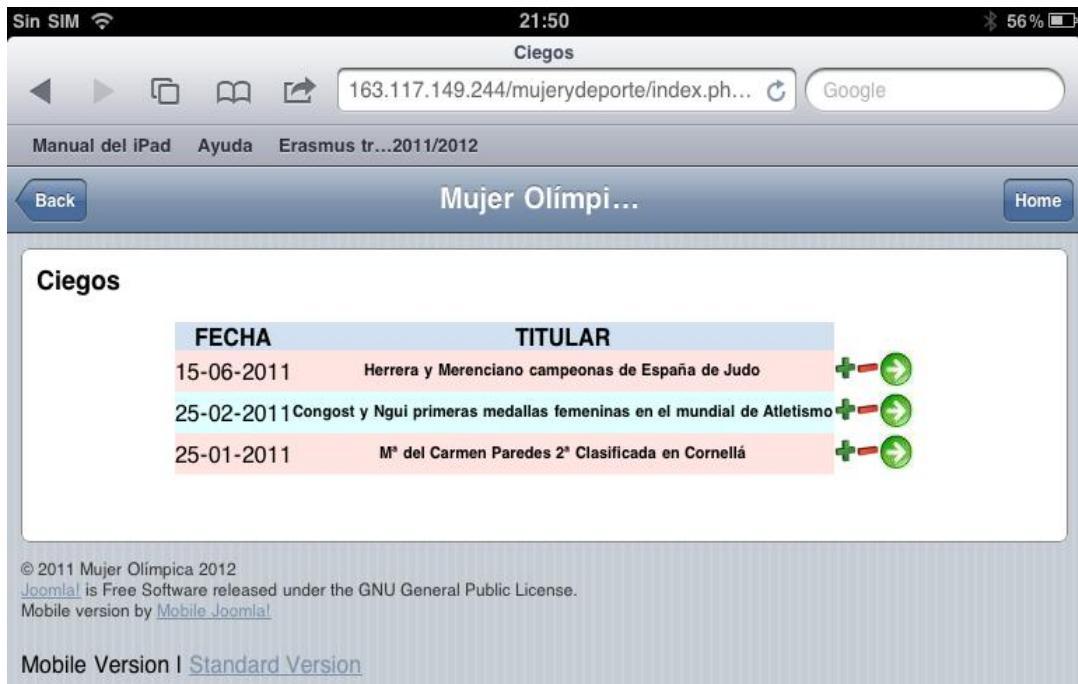


Ilustración 43 Versión para plataformas MAC



Ilustración 44 Versión para plataformas smartphone (no MAC)

### 4.3. SEGURIDAD EN LA PLATAFORMA

Una vez esté estable toda la plataforma llega la hora de dotarle de seguridad buscando los posibles fallos o puntos vulnerables. Debido a la gran popularidad de los CMS en la actualidad es muy frecuente el ataque a este tipo de sitios, sobretodo porque todos guardan la misma estructura y la ingeniería social suele funcionar bastante bien para obtener las contraseñas.

El primer tipo de ataque presumible es la ingeniería social. Joomla! genera un usuario administrador que se llama “*admin*”, por lo que tan solo hace falta obtener la clave. El primer paso y más sencillo es cambiar este usuario y renombrarlo. Además, combinando esta acción con la creación de una contraseña robusta para este usuario con permisos de administración global del sitio se evitaría también un posible ataque de fuerza bruta.

Otro tipo de ataque a evitar es la existencia de *sniffers*. Este tipo de ataques consiste en estar a la escucha de los paquetes que se envían entre cliente y servidor para obtener el usuario y contraseña. Mediante el complemento “*Encryption component*” se cifran estas peticiones de registro en la página web mediante un algoritmo RSA apoyado en una clave de 1024 dígitos que se genera automáticamente cada 180 días. Este algoritmo es añadido mediante la ejecución de un código javascript.

A continuación, con la ayuda del complemento *Firebug*<sup>45</sup> para Mozilla Firefox se muestran las capturas del tráfico de red con [véase Ilustración 46] y sin [véase Ilustración 47] el algoritmo de encriptación.



Inicio Usuario Abrir/Cerrar sesión Acceso

Acceso

Nombre de usuario Alumno1

Contraseña ●●●●●●

Recordarme

Acceso

- [¿Olvido su contraseña?](#)
- [¿Olvidó su nombre de usuario?](#)
- [Regístrese](#)

Ilustración 45 Formulario de login

---

<sup>45</sup> <http://getfirebug.com/>

Limpiar Persistir **Todo** HTML CSS JS XHR Imágenes Flash Multimedia

POST index.php?option=com\_user&view=logi 303 See other 163.117.149.244 42.8 KB

Parámetros Cabeceras **Post** Respuesta Caché

Parámetros application/x-www-form-urlencoded

545840b6f7d6124a5759e1661... 1

Submit Acceso

encrypted\_com-login\_com-f... 1bdfd593ed2623f9f9501bb04f3263320634b1084c812b2e103d73a388580f125e53dc52358c2aa1e477ebbae771cf83e84ea579ed7f5bdf355a4542c9cfc8d

formtoken\_com-login\_com-f... izbtxipyoeprxwmy

option com\_user

passwd rE34f@fffrE34f@fffrE34f@fffrE34f@fff

return aW5kZXgucGhw

task login

username Alumno1

Fuente

Content-Type: application/x-www-form-urlencoded  
Content-Length: 394

formtoken\_com-login\_com-form-login=izbtxipyoeprxwmy&encrypted\_com-login\_com-form-login\_passwd\_passwd=1bdfd593ed2623f9f9501bb04f3263320634b1084c812b2e103d73a388580f125e53dc52358c2aa1e477ebbae771cf83e84ea579ed7f5bdf355a4542c9cfc8d&username=Alumno1&passwd=rE34f@fffrE34f@fffrE34f@fffrE34f@fff&Submit=Acceso&option=com\_user&task=login&return=aW5kZXgucGhw&545840b6f7d6124a5759e16619454486=1

Ilustración 46 Login cifrado

Limpiar Persistir **Todo** HTML CSS JS XHR Imágenes Flash Multimedia

URL	Estado	Dominio	Tamaño	Línea de tiempo
POST index.php?option=com_user&view=logi	303 See other	163.117.149.244	40.9 KB	

Parámetros Cabeceras **Post** Respuesta Caché

Parámetros application/x-www-form-urlencoded

1761e979a2e6660e5da091220... 1

Submit Acceso

option com\_user

passwd alumno1

return aW5kZXgucGhw

task login

username Alumno1

Fuente

Content-Type: application/x-www-form-urlencoded  
Content-Length: 127

username=Alumno1&passwd=alumno1&Submit=Acceso&option=com\_user&task=login&return=aW5kZXgucGhw&1761e979a2e6660e5da09122008bb7b4=1

Ilustración 47 Login en claro

## SEGURIDAD EN LA PLATAFORMA

Ante la posibilidad de registros múltiples realizados por robots se ha implementado un captcha a la hora de dar de alta un nuevo usuario en la base de datos. Esta opción de captcha es facilitada por Google, y solo estará disponible una vez la página se encuentre alojada en un hosting con un dominio definitivo.

A continuación, se muestra la imagen del formulario de registro con el captcha.



The image shows a web browser window with a registration form. At the top, there are navigation buttons: 'Inicio', 'Usuario', 'Abrir/Cerrar sesión', and 'Nuevo'. The main content area is titled 'Regístrese' and contains several input fields, each followed by an asterisk (\*): 'Nombre:', 'Nombre de usuario:', 'E-mail:', 'Contraseña:', and 'Verifique su contraseña:'. Below these fields, there is a CAPTCHA challenge. The challenge consists of a red-bordered box containing the text '355, liberati'. Below this box is a yellow input field with the text 'Type the two words:'. To the right of the input field are three small icons: a refresh icon, a speaker icon, and a question mark icon. Below the CAPTCHA box, there is a 'RECAPTCHA™' logo with the text 'stop spam. read books.' Below the CAPTCHA challenge, there is a note: 'Los campos marcado con un asterísco (\*) son obligatorios'. At the bottom of the form, there is a 'Regístrese' button.

Ilustración 48 Captcha en formulario de registro

Por último, hay que evitar los ataques de ingeniería inversa. En el caso de Joomla! el panel de administración de la plataforma está publicado siempre en la dirección <http://www.tupagina.com/administrator>; de este modo cualquier persona que haya obtenido el usuario y contraseña podrá acceder a él. Para evitarlo, se ha instalado el complemento *JSecure* que oculta de alguna manera este panel. Para ello, el administrador deberá crear un



“token” o palabra clave que sólo él conozca que será añadido a la URL del panel de administración del sitio del siguiente modo: <http://www.tupagina.com/administrator/?token>. Así, si no se introduce el token se redirecciona al front-end del sitio y no está visible el panel de administración.

A continuación, se muestran las capturas de pantalla para ilustrar el uso de este complemento de una manera más gráfica.



**Ilustración 49** Uso de JSecure

Hay que añadir que Joomla! permite la utilización de una conexión segura forzando las conexiones mediante el protocolo SSL. Del mismo modo que la opción del captcha, esta opción no se habilitará hasta que no se aloje en un servidor definitivo y se disponga de un dominio. A continuación se muestra la imagen donde se puede habilitar esta opción en el back-end de la plataforma:

**Configuración del servidor**

Ruta al directorio temporal	<input type="text" value="C:\wamp\www\pfc\tmp"/>
Compresión GZIP de las páginas	<input checked="" type="radio"/> No <input type="radio"/> Sí
Informe de errores	<input type="text" value="Predeterminado del sistema"/> ▼
Forzar SSL	<input type="text" value="Nada"/> ▼

Ilustración 50 Opción SSL

# Capítulo 5: Plan de desarrollo

---

Albert Einstein (14/03/1879 – 18/04/1955). Físico americano.

¿Por qué esta magnífica tecnología científica, que ahorra trabajo y nos hace la vida más fácil, nos aporta tan poca felicidad?

La respuesta es está, simplemente: porque aún no hemos aprendido a usarla con tino.



## 5.1. HERRAMIENTAS UTILIZADAS

### 5.1.1. ECLIPSE

Eclipse es un entorno de desarrollo integrado de código abierto multiplataforma para desarrollar lo que el proyecto llama "Aplicaciones de Cliente Enriquecido", opuesto a las aplicaciones "Cliente-liviano" basadas en navegadores. Esta plataforma, típicamente ha sido usada para desarrollar entornos de desarrollo integrados (del inglés IDE), como el IDE de Java llamado *Java Development Toolkit* (JDT) y el compilador (ECJ) que se entrega como parte de Eclipse (y que son usados también para desarrollar el mismo Eclipse). Sin embargo, también se puede usar para otros tipos de aplicaciones cliente, como BitTorrent y Azureus.

Eclipse es también una comunidad de usuarios, extendiendo constantemente las áreas de aplicación cubiertas. Un ejemplo es la creación del *Eclipse Modeling Project*, cubriendo casi todas las áreas de *Model Driven Engineering*.

Eclipse fue desarrollado originalmente por IBM como el sucesor de su familia de herramientas para VisualAge. A día de hoy, Eclipse es desarrollado por la Fundación Eclipse, una organización independiente sin ánimo de lucro que fomenta una comunidad de código abierto y un conjunto de productos complementarios, capacidades y servicios.

Para el proyecto, Eclipse será el entorno de trabajo de desarrollo en JAVA.

### 5.1.2. JAVA

Java<sup>46</sup> es un lenguaje de programación orientado a objetos desarrollado por Sun Microsystems a principios de los años 90. El lenguaje en sí mismo toma mucha de su sintaxis de C y C++, pero tiene un modelo de objetos más simple y elimina herramientas de bajo nivel, que suelen inducir a muchos errores, como la manipulación directa de punteros o memoria.

Las aplicaciones Java están típicamente compiladas en un *bytecode*, aunque la compilación en código máquina nativo también es posible. En el tiempo de ejecución, el *bytecode* es normalmente interpretado o compilado a código nativo para la ejecución, aunque la ejecución directa por hardware del *bytecode* por un procesador Java también es posible.

La implementación original y de referencia del compilador, la máquina virtual y las bibliotecas de clases de Java fueron desarrolladas por Sun Microsystems en 1995. Desde entonces, Sun ha controlado las especificaciones, el desarrollo y evolución del lenguaje a través del Java Community Process, si bien otros han desarrollado también implementaciones alternativas de estas tecnologías de Sun, algunas incluso bajo licencias de software libre.

#### 5.1.2.1. JDBC

**Java Database Connectivity**, más conocida por sus siglas **JDBC**, es una API que permite la ejecución de operaciones sobre bases de datos desde el lenguaje de programación

---

<sup>46</sup> [ECK+07]

## HERRAMIENTAS UTILIZADAS

Java, independientemente del sistema operativo donde se ejecute o de la base de datos a la cual se accede, utilizando el dialecto SQL del modelo de base de datos que se utilice.

El API JDBC se presenta como una colección de interfaces Java y métodos de gestión de manejadores de conexión hacia cada modelo específico de base de datos. Un manejador de conexiones hacia un modelo de base de datos en particular es un conjunto de clases que implementan las interfaces Java y que utilizan los métodos de registro para declarar los tipos de localizadores a base de datos (URL) que pueden manejar. Para utilizar una base de datos particular, el usuario ejecuta su programa junto con la biblioteca de conexión apropiada al modelo de su base de datos, y accede a ella estableciendo una conexión, para ello provee el localizador a la base de datos y los parámetros de conexión específicos. A partir de allí puede realizar con cualquier tipo de tareas con la base de datos a las que tenga permiso: consulta, actualización, creación, modificación y borrado de tablas, ejecución de procedimientos almacenados en la base de datos, etc.

Para el proyecto, JAVA será el lenguaje usado para crear el buscador de noticias y la API JDBC la solución para conectar el buscador con la base de datos.

### 5.1.3. MYSQL

MySQL es el sistema gestor de base de datos de software más popular del mundo. Es un sistema de gestión de base de datos relacionales y de código abierto. Como sistema de gestión de base de datos está considerado entre los de mejor rendimiento (mayor velocidad), es multiproceso (funciona en subprocesos independientes), es multiusuario y de excelente fiabilidad. Al considerar a MySQL como sistema de gestión de base de datos damos a entender que cumple con la doble función de servidor y gestor de los datos.

Un servidor de base de datos relacionales mantiene la información en tablas independientes siguiendo un modelo de almacenamiento que permite un rápido acceso a los datos mediante el uso de un lenguaje de consulta denominado SQL (Structure Query Language).

MySQL AB fue la creadora de MySQL. Fue fundada en 1995 y es una de las grandes empresas de software libre del mundo. Recientemente, Sun Microsystems ha anunciado un acuerdo para adquirir MySQL AB. Entre los usuarios de MySQL AB se encuentran Alcatel-Lucent, Amazon.com, Google, Digg, Facebook, Nokia, Yahoo y YouTube, entre otros.

A continuación se detallan una serie de características propias de MySQL. En este caso validas para la versión MySQL 5.0.51 que ha sido utilizada para la realización del proyecto.

#### *LICENCIAS DE USO*

En sus inicios MySQL fue software totalmente libre, posteriormente se implanto el criterio de la doble licencia, que es el criterio vigente.

MySQL sigue gozando del estatus de licencia pública general (GPL, General Public License) o GNU que autoriza su uso limitado a cualquier persona con la condición de que ésta no lo redistribuya y si lo hiciese debe hacerlo con el mismo tipo de licencia. La licencia GPL GNU es sin cargo.

La licencia GPL GNU tiene precisas reglas de uso que se pueden consultar en: [www.gnu.org/licenses](http://www.gnu.org/licenses), pero se pueden resumir sus principios básicos en los siguientes:

- Libertad para ejecutar el programa de software libre.
- Libertad para acceder al funcionamiento de código interno.
- Libertad para modificar el código para adaptarlo a nuestros fines.
- Libertad para distribución de copias.

Muchos desarrolladores de software se benefician de que MySQL pueda distribuirse como GPL GNU pero no desean esto para su software. Para ellos se creó la licencia comercial.

Cuando un desarrollador de software no desea que su software sea GPL GNU pero quiere aprovechar la funcionalidad MySQL en sus aplicaciones tiene la opción de adquirir una licencia comercial.

Se debe tener en cuenta que MySQL en su versión GPL no tiene ninguna garantía, pero SUN se responsabiliza por el producto contratado con la licencia comercial.

Para el proyecto, MySQL será la base de datos encargada de almacenar la información encontrada por el buscador de noticias.

#### **5.1.4. PHP**

Es un lenguaje de programación interpretado, diseñado originalmente para la creación de páginas web dinámicas. Es usado principalmente en interpretación del lado del servidor pero actualmente puede ser utilizado desde una interfaz gráfica usando diversas bibliotecas. De forma nativa, incorpora el driver de MySQL para manejar bases de datos.

Originalmente fue creado en 1994 por Rasmus Lerdorf, sin embargo la implementación principal es producida ahora por The PHP Group y sirve como estándar para PHP al no haber una especificación formal. Está publicado bajo la PHP License, que para la Free Software Foundation la considera como licencia de software libre.

Para el proyecto, PHP será el soporte que se usará en la vista de la aplicación para mostrar información de la base de datos desde la vista.

#### **5.1.5. PHPMYADMIN**

Es una herramienta escrita en PHP con la intención de manejar la administración de MySQL a través de páginas web utilizando Internet. Se encuentra vigente desde el año 1998 y se encuentra disponible bajo licencia GPL.

Para el proyecto, phpMyAdmin será el soporte que se usará para administrar la base de datos a modo de sistema gestor de base de datos.

### 5.1.6. JOOMLA!

Joomla! <sup>47</sup> (la traducción “todos juntos”) es un sistema de gestión de contenidos (CMS) cuya principal virtud es la de permitir editar el contenido de un sitio web de manera sencilla. Es una aplicación de código abierto programada mayoritariamente en PHP bajo licencia GPL. Este administrador de contenidos puede trabajar en Internet o intranets y requiere de una base de datos MySQL, así como preferiblemente un servidor HTTP Apache.

Joomla! nace de la escisión de un grupo de programadores de Mambo (otro CMS) en agosto de 2005. Alrededor de este proyecto se ha creado una comunidad de gente que ayuda a evolucionar Joomla! creando una serie de extras como son las plantillas, extensiones, módulos y plugins que añaden funcionalidad a la web.

Joomla! es capaz de crear un portal web de una forma sencilla y rápida mediante una plantilla que se rellena de artículos o de los extras anteriormente mencionados al antojo del administrador mediante una zona de administración muy simple. Además gestiona diversas categorías de roles para los usuarios.

El pasado mes de enero fue liberada la versión Joomla! 1.6 que requiere nuevas versiones de php y MySQL, lo que provoca la actualización de todos los extras existentes en la comunidad. Para el proyecto la versión utilizada ha sido la anterior, Joomla! 1.5.23 más extendida y preparada para todos los extras existentes en la comunidad. Con el paso del tiempo será recomendable evolucionar y actualizar a la versión 1.6 junto con todos los componentes utilizados.

Para el proyecto, Joomla! supondrá ser la vista de la aplicación.

Las extensiones que han sido utilizadas para la mejora del sitio web son:

- *Ajax Shoutbox 1.3*: asistente para la creación de una bandeja de gritos.
- *Akeeba Backup 3.3.b1*: asistente para la creación de respaldos de seguridad del sitio web.
- *Fontsize*: asistente para facilitar al usuario el cambio de tamaño de la fuente.
- *JEvents 1.5*: asistente para la creación y administración de un calendario de eventos.
- *Joomfish 2.04*: asistente para convertir la página en multidioma.
- *Jumi 2.0.6*: asistente para la inyección de código en los artículos y sea ejecutado por el navegador.
- *JoomlaStats 3.0.2*: asistente para la creación de estadísticas del tráfico generado en la plataforma.
- *MobileJoomla 0.9.12*: asistente para crear la visualización en dispositivos móviles.
- *VirtueMart 1.1.9*: asistente para la creación y administración de una tienda virtual.
- *Xmap 1.2.8*: asistente para la creación de un mapa del sitio web.

Cabe destacar el uso de extensiones para mejorar la seguridad de la plataforma, tales como:

---

<sup>47</sup> [MAR+11] y [MAT+08]



- *Encrypt Configuration 1.5.4*: asistente para cifrar la información en los registros.
- *JSecure 1.0.9*: asistente para ocultar el panel de administración de modo que no sea público.
- *Jez ReCaptcha*: asistente para añadir captchas en los formularios, facilitado por Google, en la web.

### 5.1.7. WAMP SERVER

Para la realización de las pruebas, se ha utilizado la versión 2.0 del paquete WAMP (Windows + Apache + MySQL + Php) distribuido bajo licencia libre que permite ejecutar un servidor en Windows para así poder instalar Joomla! 1.5.X. Este paquete incluye un servidor y todas las aplicaciones necesarias para la realización del proyecto descritas anteriormente:

- Apache Server 2.2.14
- MySQL 5.1.41
- PHP 5.3.0
- phpMyAdmin 3.2.0.1

## 5.2. ACOPLAMIENTO DE LAS HERRAMIENTAS

Para entender un poco mejor como se acoplan todas estas herramientas descritas anteriormente para la realización del proyecto, se muestra el siguiente gráfico.

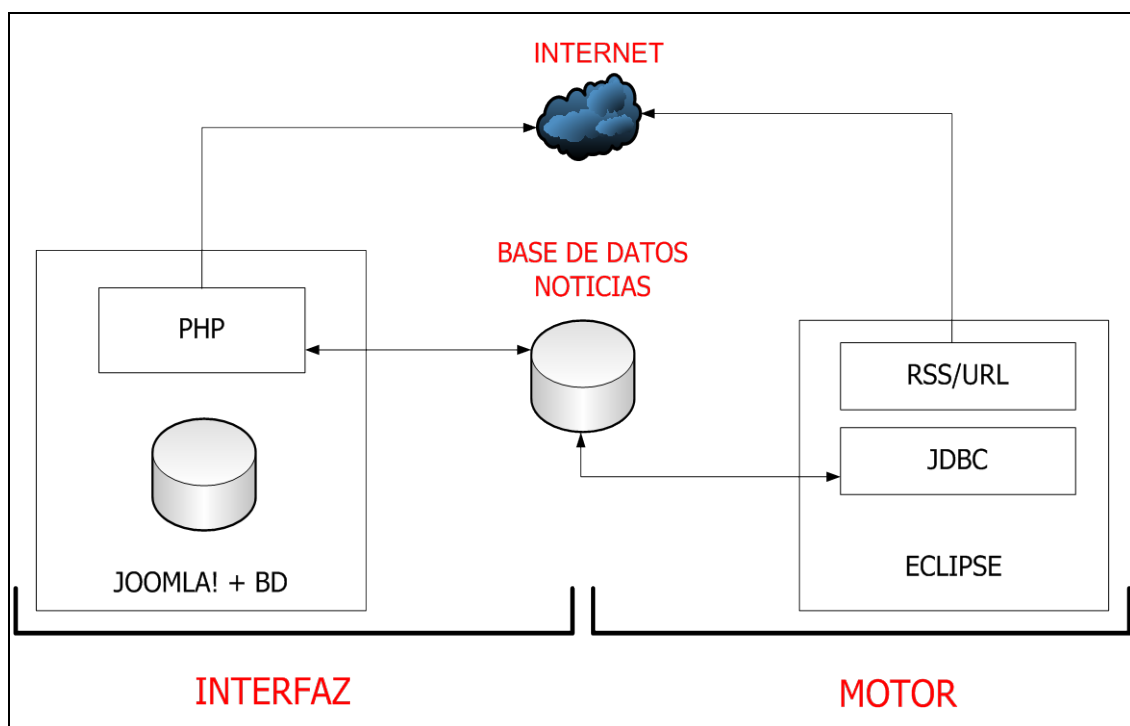
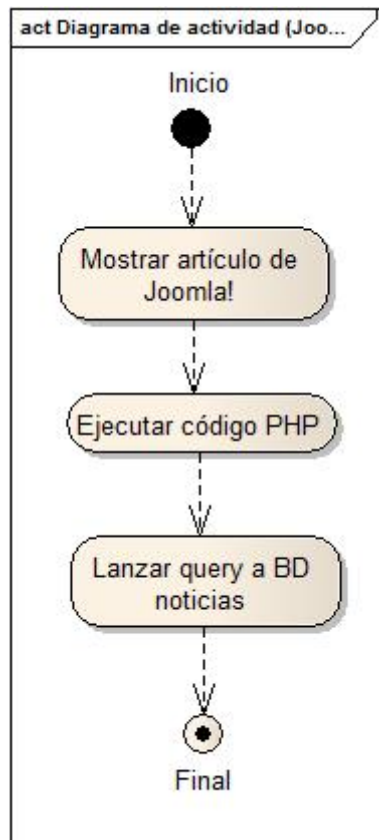


Ilustración 51 Acoplamiento de las herramientas

Se puede observar la existencia de dos partes claramente diferenciadas, la zona del interfaz gráfico con su propia base de datos (base de datos Joomla!) y la zona del motor de búsqueda que interacciona con la base de datos Noticias.

El interfaz gráfico es creado mediante el gestor de contenidos Joomla! que mediante unos scripts en PHP accede a la base de datos Noticias creada en MySQL. Para que este interfaz gráfico sea multidioma, además del uso de un componente para la plataforma Joomla! ha sido necesario la creación de un script en PHP para que los datos sean traducidos a través del API Google Translate y se muestre correctamente en cualquier idioma.

Para intentar esclarecer un poco mejor el complejo sistema de publicación creado en Joomla!, se muestra en el siguiente gráfico un pequeño diagrama de actividad.



**Ilustración 52 Diagrama de actividad de una publicación en Joomla!**

Como se puede observar, se solicita en Joomla! mostrar un artículo existente en la base de datos Joomla!; el contenido de este artículo resulta ser código PHP que ejecuta una consulta a la base de datos Noticias; esta consulta devuelve una tabla con el listado de las noticias que es lo que finalmente el usuario visualiza como artículo de Joomla!.

Por otro lado, está el motor de búsqueda de noticias en Internet que ha sido creado en JAVA mediante el entorno de desarrollo ofrecido por Eclipse, y con la ayuda del JDBC accede a la base de datos Noticias donde almacenará toda la información recopilada.

Para entender un poco más este entorno de trabajo, se presenta de una manera simplificada el diagrama de actividad del motor de búsqueda.

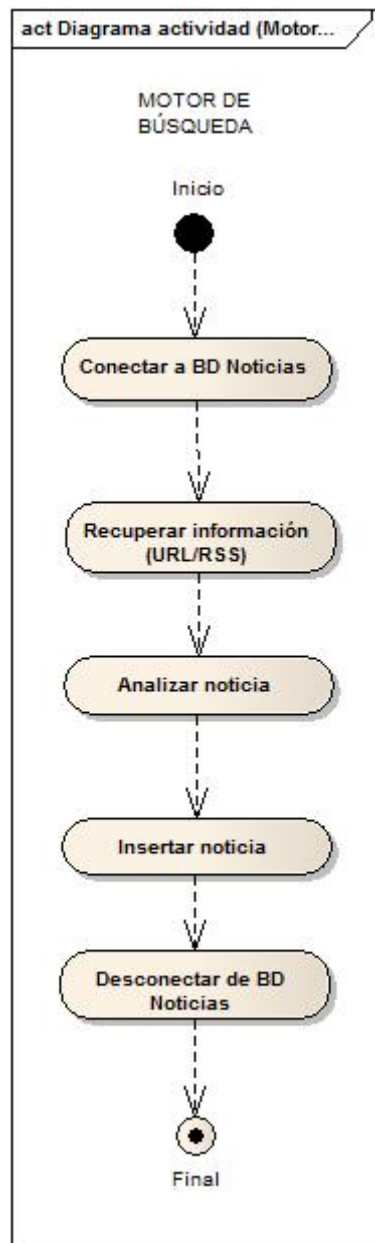


Ilustración 53 Diagrama de actividad del motor de búsqueda

Como se puede apreciar en el diagrama, el motor de búsqueda inicialmente se conecta a la base de datos Noticias para recuperar la URL o RSS que se pretende analizar. A continuación, empieza a analizar las noticias aparecidas en dicha fuente de información aplicando el algoritmo de selección de noticias descrito anteriormente en el capítulo 3 [véase Ilustración 13], y en caso de ser válida se inserta la noticia en la base de datos Noticias.

## 5.2. CICLO DE VIDA DEL PROYECTO

El ciclo de vida utilizado para este proyecto es el ciclo de vida en cascada. Este tiene las siguientes etapas:



Ilustración 54 Ciclo de vida

## 5.3. ESTRUCTURA DE TRABAJO (WBS)

A continuación, se detallan las tareas existentes en cada uno de las etapas del ciclo de vida del proyecto.

### 5.3.1. I. Análisis de Requisitos

#### I.A Inicio del Proyecto.

En esta fase da comienzo el proyecto y en él se elige el ciclo de vida que se va a utilizar y se hace la estimación del dicho proyecto.

#### I.B Plan del Proyecto.

El plan de proyecto es el producto del plan de desarrollo del software que abarca las tareas de gestión del software, tales como la evaluación del proyecto, plan de instalación, plan de documentación, plan de formación, plan de integración y dirección del proyecto.

#### I.C Requisitos de usuario

Han sido definidos por el equipo de trabajo una serie de requisitos de usuario básicos.

### **I.D Análisis y creación de un algoritmo de búsqueda**

Se deberá crear un algoritmo nuevo de búsqueda en base a los requisitos del usuario. Al desarrollar un nuevo algoritmo, esto implica un alto coste de investigación y análisis en como solventar dicho problema.

## **5.3.2. II. Diseño**

### **II.A Definir requisitos Software**

Se deben definir los requisitos de software.

### **II.B Definir el diseño Software**

Se deberá definir el diseño del software, es decir, la arquitectura del software como el diseño de la arquitectura.

## **5.3.3. III. Implementación y pruebas unitarias**

### **III.A Codificación**

Se deberá codificar el proyecto como la creación del código ejecutable.

### **III.B Creación de Datos de Prueba**

Se crearán datos de pruebas para ver el buen funcionamiento de las distintas partes del sistema.

### **III.C Integración**

En esta fase se realiza la integración del algoritmo de búsqueda junto con la plataforma web.

### **III.D Pruebas**

Una vez realizado los datos de prueba y haber hecho la integración del proyecto se deben realizar las pruebas oportunas para la aprobación final del proyecto.

### **III.E Preparación de la Documentación de Usuario**

Se documenta todo lo realizado durante el proyecto, cómo está hecho el proyecto, cómo se utiliza, cómo funciona.

## **5.3.4. IV. Integración y pruebas del sistema**

### **IV.A Instalar Software**

Una vez hecho el software se debe instalar para la del software en el entorno de funcionamiento.

### **IV.B Aceptar Pruebas**

El cliente debe dar el visto bueno a las pruebas realizadas al software.

#### IV.C Formación del usuario

El usuario recibirá formación al respecto del funcionamiento y manejo del software.

#### IV.D Funcionamiento del Software

Se debe dar una explicación detallada de cómo funciona el software realizado.

### 5.3.5. V. Operaciones y Mantenimiento

#### V.A Mantenimiento

El cliente contará con mantenimiento hasta la finalización de los Juegos Olímpicos de Londres 2012<sup>48</sup>.

#### V.B Soporte técnico

El cliente contará con mantenimiento hasta la finalización de los Juegos Olímpicos de Londres 2012.

## 5.4. ESTRUCTURA DE RECURSOS (RBS)

A continuación, se detalla la estructura de recursos utilizados en el proyecto. Como se demuestra en la siguiente ilustración, los recursos utilizados son de dos tipos, recursos materiales y recursos humanos.

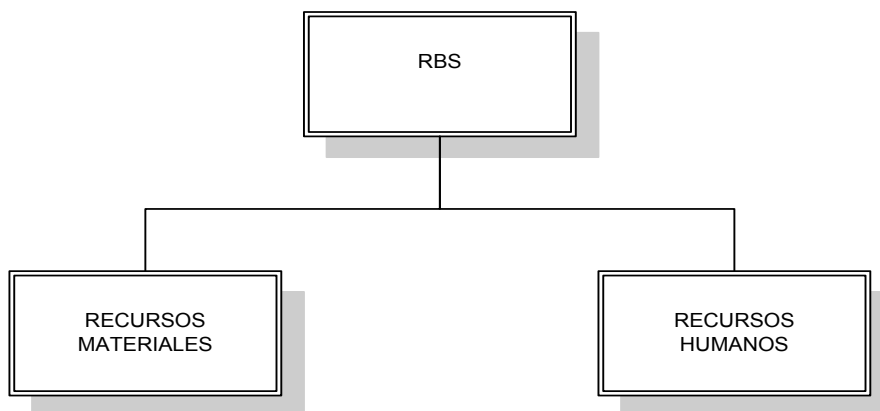


Ilustración 55 RBS

Los recursos materiales es todo aquello relacionado con software, hardware, infraestructuras, material de oficina y gastos generales utilizados durante la realización del proyecto.

---

<sup>48</sup> <http://www.london2012.com>

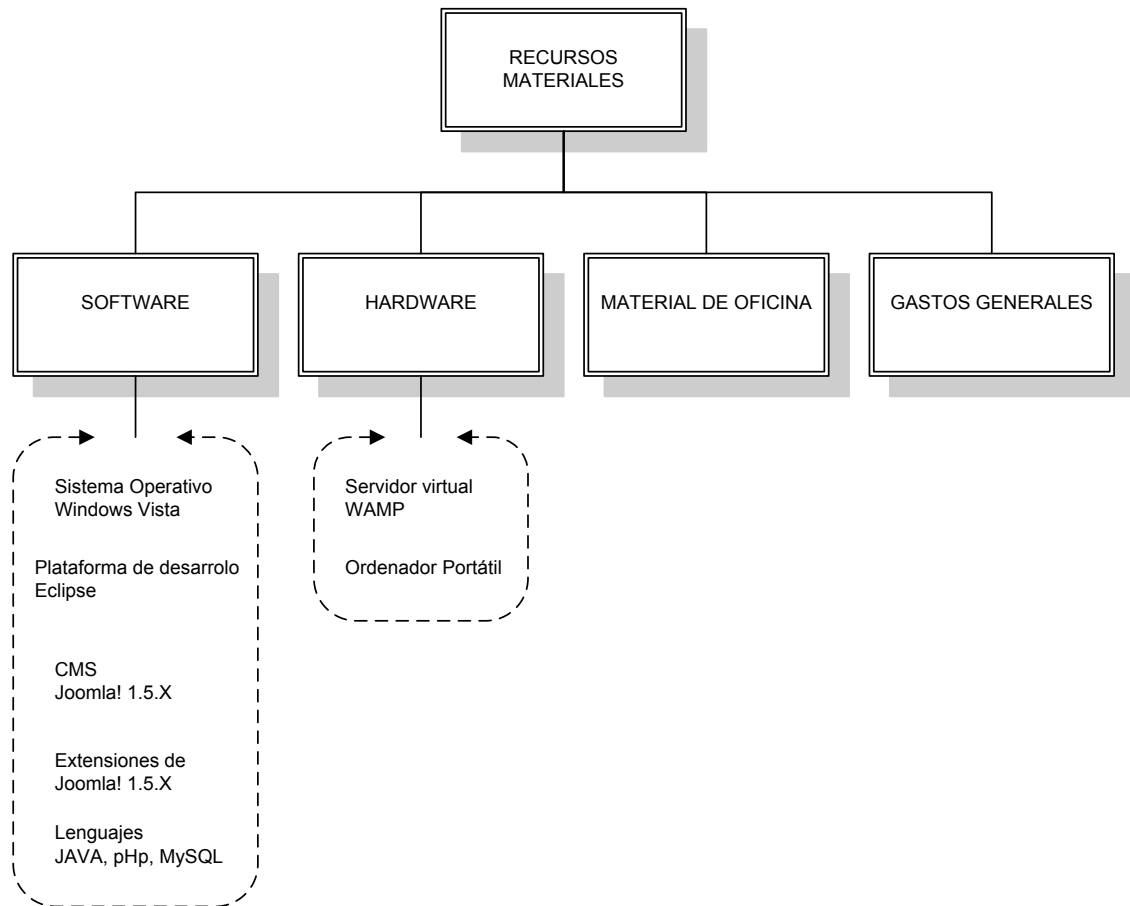


Ilustración 56 Recursos Materiales

Por otra parte, los recursos humanos es toda la plantilla utilizada para la realización del proyecto.

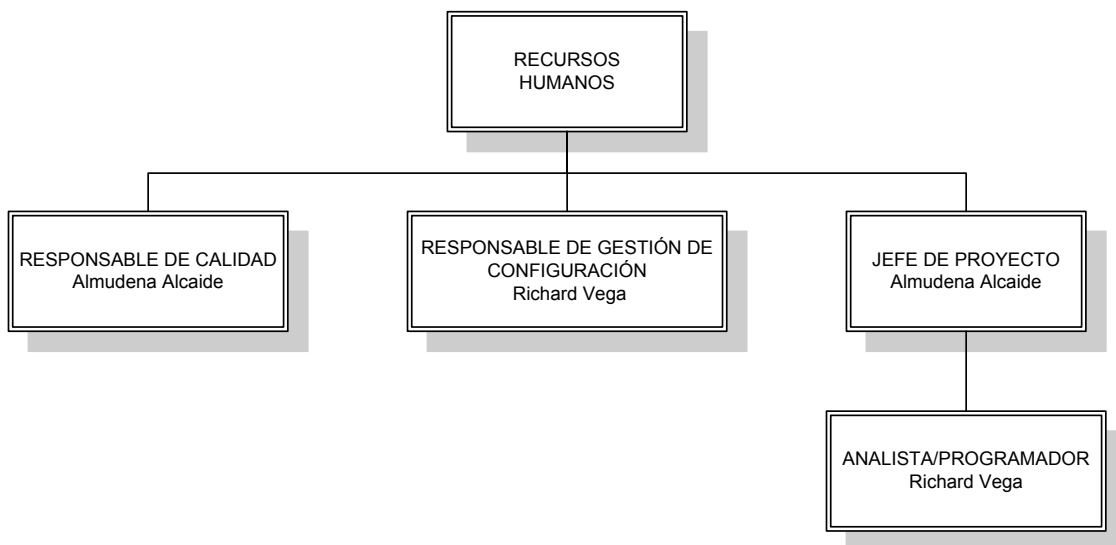


Ilustración 57 Organigrama de RRHH

## 5.5. DIAGRAMA DE GANTT

En la siguiente tabla, se muestra la lista de tareas con sus fechas de realización par el diagrama de Gantt.

Tabla 37 Lista de tareas

Id		Nombre de tarea	Duración	Comienzo	Fin
1		<b>PROYECTO</b>	<b>233 días</b>	<b>jue 15/07/10</b>	<b>lun 06/06/11</b>
2		<b>ANÁLISIS</b>	<b>56 días</b>	<b>jue 15/07/10</b>	<b>jue 30/09/10</b>
3		Inicio del proyecto	1 día	jue 15/07/10	jue 15/07/10
4		Plan del Proyecto	5 días	vie 16/07/10	jue 22/07/10
5		Requisitos de usuario	5 días	vie 23/07/10	jue 29/07/10
6		Análisis y creación del algoritmo	45 días	vie 30/07/10	jue 30/09/10
7		<b>DISEÑO</b>	<b>6 días</b>	<b>vie 01/10/10</b>	<b>vie 08/10/10</b>
8		Definir requisitos softw are	3 días	vie 01/10/10	mar 05/10/10
9		Definir diseño softw are	3 días	mié 06/10/10	vie 08/10/10
10		<b>IMPLEMENTACIÓN Y PRUEBAS</b>	<b>150 días</b>	<b>lun 11/10/10</b>	<b>vie 06/05/11</b>
11		Codificación	75 días	lun 01/11/10	vie 11/02/11
12		Creación de datos de prueba	50 días	lun 11/10/10	vie 17/12/10
13		Integración	10 días	lun 14/02/11	vie 25/02/11
14		Pruebas	20 días	lun 28/02/11	vie 25/03/11
15		Documentación	30 días	lun 28/03/11	vie 06/05/11
16		<b>INTEGRACIÓN</b>	<b>20 días</b>	<b>lun 09/05/11</b>	<b>vie 03/06/11</b>
17		Instalar softw are	5 días	lun 09/05/11	vie 13/05/11
18		Aceptar pruebas	5 días	lun 16/05/11	vie 20/05/11
19		Formación del usuario	5 días	lun 23/05/11	vie 27/05/11
20		Funcionamiento del softw are	5 días	lun 30/05/11	vie 03/06/11
21		<b>OPERACIONES Y MANTENIMIENTO</b>	<b>1 día</b>	<b>lun 06/06/11</b>	<b>lun 06/06/11</b>
22		Mantenimiento	1 día	lun 06/06/11	lun 06/06/11
23		Soporte técnico	1 día	lun 06/06/11	lun 06/06/11

Y a continuación se muestra el correspondiente diagrama de Gantt a la lista de tareas anterior. Para facilitar la lectura del diagrama, se muestra a la izquierda de cada barra que representa una tarea su código identificador con la tabla de la lista de tareas [véase Tabla 36].



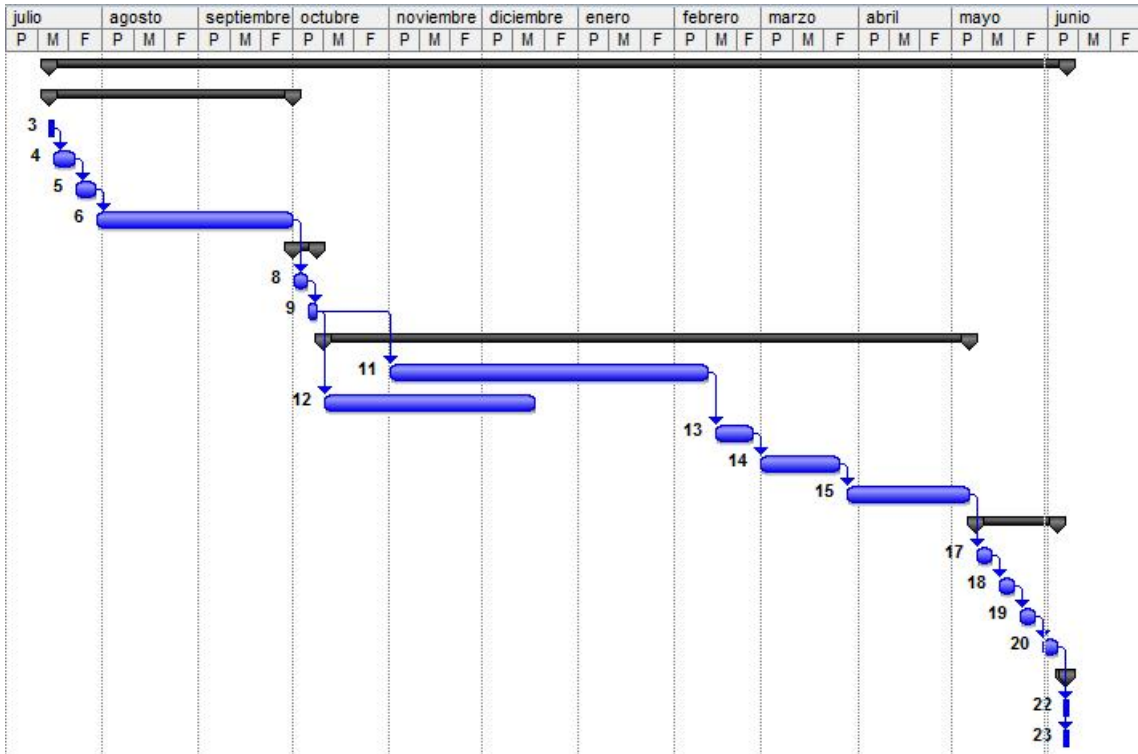



Ilustración 58 Diagrama de Gantt

## 5.6. PRESUPUESTO

 <b>UNIVERSIDAD CARLOS III DE MADRID</b> <b>Escuela Politécnica Superior</b>						
<b>PRESUPUESTO DE PROYECTO</b>						
<b>1.- Autor:</b>						
Richard Vega Caridad						
<b>2.- Departamento:</b>						
Informática						
<b>3.- Descripción del Proyecto:</b>						
- Título	Analizador semántico: Mujer Olímpica 2012					
- Duración (meses)	5					
Tasa de costes Indirectos:	20%					
<b>4.- Presupuesto total del Proyecto (valores en Euros):</b>						
37.211,00 Euros						
<b>5.- Desglose presupuestario (costes directos)</b>						
<b>PERSONAL</b>						
Apellidos y nombre	N.I.F.	Categoría	Dedicación (hombres mes) <sup>a)</sup>	Coste hombre mes	Coste (Euro)	Firma de conformidad
Alcaide Raya, Almudena		Ingeniero Senior	2,22	4.289,54	9.522,78	
Vega Caridad, Richard		Ingeniero	7,93	2.694,39	21.366,51	
					0,00	
					0,00	
<b>Hombres mes 10,15</b>				<b>Total</b>	<b>30.889,29</b>	
				<small>13.677,26 Euros</small>		
<sup>a)</sup> 1 Hombre mes = 131,25 horas. Máximo anual de dedicación de 12 hombres mes (1575 horas) Máximo anual para PDI de la Universidad Carlos III de Madrid de 8,8 hombres mes (1.155 horas)						
<b>EQUIPOS</b>						
Descripción	Coste (Euro)	% Uso dedicado proyecto	Dedicación (meses)	Periodo de depreciación	Coste imputable <sup>a)</sup>	
Ordenador portátil	966,00	100	5	60	80,50	
		100		60	0,00	
		100		60	0,00	
		100		60	0,00	
		100		60	0,00	
					0,00	
				<b>Total</b>	<b>80,50</b>	
<sup>a)</sup> Fórmula de cálculo de la Amortización: $\frac{A}{B} \times C \times D$ A = nº de meses desde la fecha de facturación en que el equipo es utilizado B = periodo de depreciación (60 meses) C = coste del equipo (sin IVA) D = % del uso que se dedica al proyecto (habitualmente 100%)						
<b>SUBCONTRATACIÓN DE TAREAS</b>						
Descripción	Empresa	Coste imputable				
Plantilla CMS	YooTheme	39,00				
Hosting						
Dominio						
	<b>Total</b>	<b>39,00</b>				
<b>OTROS COSTES DIRECTOS DEL PROYECTO<sup>a)</sup></b>						
Descripción	Empresa	Costes imputable				
	<b>Total</b>	<b>0,00</b>				
<sup>a)</sup> Este capítulo de gastos incluye todos los gastos no contemplados en los conceptos anteriores, por ejemplo: fungible, viajes y dietas,						
<b>6.- Resumen de costes</b>						
Presupuesto Costes Totales	<b>Presupuesto Costes Totales</b>					
Personal	30.889					
Amortización	81					
Subcontratación de tareas	39					
Costes de funcionamiento	0					
Costes Indirectos	6.702					
<b>Total</b>	<b>37.211</b>					

# Capítulo 6: Conclusiones

---

John Fitzgerald Kennedy (29/05/1917 - 22/11/1963). Trigésimo quinto presidente de Estados Unidos.

Un hombre inteligente es aquél que sabe ser tan inteligente como para contratar gente más inteligente que él.



## 6.1. CONCLUSIONES

Lo fundamental del proyecto para el alumno desarrollador del proyecto ha sido la adquisición de conocimientos de programación orientada a objetos a nivel práctica en un entorno y lenguaje desconocido como son Eclipse y JAVA, algo que era de lo que era completamente ajeno y necesario para el futuro profesional del alumno.

Además de esta adquisición de conocimiento junto con la de los gestores de contenido, ha servido para realizar una aproximación a la lingüística computacional para poder realizar su propio analizador sintáctico, sin llegar a utilizar un corpus lingüístico como tal se ha implementado un algoritmo capaz de reconocer varios tipos de titulares deportivos en la prensa digital ayudándose de las expresiones regulares.

En cuanto al deporte femenino y su repercusión, una vez estudiado el ámbito del proyecto, se ha observado como hay pequeños brotes de mejoras del deporte femenino por parte de las distintas federaciones intentando realizar jornadas exclusivas para las mujeres y animar así a practicarlo.

Y como se puede apreciar en las noticias, nuestras deportistas consiguen grandes logros, no solo nos encontramos en el *siglo de oro del deporte español* masculino, sino también femenino.

### 6.1.1. RESULTADOS

Para poder analizar los resultados hay que definir previamente dos conceptos:

- ✓ **Falso positivo:** son noticias inválidas que el algoritmo de búsqueda califica como noticias válidas.
- ✓ **Falso negativo:** son noticias válidas que el algoritmo de búsqueda califica como noticias inválidas.

A fecha 10 de julio de 2011, en la base de datos de noticias existen 80 noticias, de las cuáles 9 de ellas no pueden considerarse logros como tal, es decir, falsos positivos. Sin embargo, 4 de estas noticias no son acciones negativas de las deportistas (noticias que sin llegar a ser un logro se consideran acciones positivas), por lo que de las 80 noticias recogidas se observa un margen de error del 5%.

A la hora de analizar los falsos negativos, se observa que el 100% de los que aparecen es debido a la ausencia de los parámetros de búsqueda, es decir, no se tiene en la base de datos la deportista, la acción positiva y/o el logro deportivo.

A estos resultados, hay que añadir la existencia de noticias, que rara vez aparecen, en las que en el titular no se hace referencia a ninguna sujeto femenino y si a sujetos masculinos, y una vez dentro de la noticia se puede apreciar en un párrafo un logro femenino. En este caso, siguiendo las indicaciones de la [GUI+11] dónde se indica que se debe dar prioridad a la mujer en los carteles promocionales, noticias, etc... no se considera como un logro por lo que el algoritmo lo considera correctamente como una noticia inválida.

## 6.2. LÍNEAS FUTURAS

Cabe destacar que este proyecto tiene un carácter de prototipo con vistas a futuros desarrollos.

El primer paso que requiere el proyecto antes de ser publicado es obtener el consentimiento por parte de todas las fuentes de información del uso de de toda la información que se vaya a publicar en la plataforma web debido a la existencia de derechos de autor por parte de las distintas fuentes. En caso de no obtener el consentimiento, se deberá optar por dejar los titulares y los enlaces a las páginas y no mostrar ninguna información adicional de las noticias.

A nivel de interfaz gráfico, una vez el proyecto se encuentre asentado, habría que comprobar la viabilidad de migrar la plataforma a la nueva versión Joomla! 1.6.X con todos sus módulos. Debido a que durante la realización del proyecto apareció una nueva versión del gestor de contenidos usado no sé creyó conveniente la actualización debido a la cantidad de componentes externos utilizados y no todos se encontraban actualizados o soportaban esta nueva versión. Con el paso del tiempo, los distintos componentes han ido actualizando la versión y soportando la última versión de Joomla! 1.6.5.

Según han ido avanzando el desarrollo del proyecto se ha ido observando cómo los sujetos de las noticias pueden ser localizados con distintas opciones:

- Nombre.
- Nombre y apellido.
- Apellido.
- Apodo.

Para resolver este problema se optó por añadir una nueva entrada en la base de datos provocando una duplicidad de entradas para referirse al mismo sujeto por lo que habrá que plantear una nueva estructura para la tabla SUJETO y así lograr una correcta optimización del espacio. Esta estructura podría ser la siguiente:

```
SUJETO (nombre, apellido1, apellido2, alias)
```

Siendo la clave principal la composición de los campos: nombre + apellido1 + apellido2 para identificar al sujeto. Cabe destacar que este cambio provocará una modificación en el algoritmo de búsqueda.

Con el paso del tiempo el número de noticias va incrementándose, surgiendo así el problema del espacio. La primera forma de optimizar de una manera rápida la base de datos NOTICIA sería creando una tabla FECHA de modo que el campo fecha de la tabla NOTICIA sea una referencia a la tabla FECHA.

Además, este aumento del número de noticias provoca la necesidad de un mejor filtrado en la visualización para el usuario. A día de hoy, la plataforma dispone de un filtrado por federación, por día actual, por día anterior y el resto de noticias provocando en esta última la aparición de más del 95 % de las noticias. Una de las formas de mejorar esto, sería realizando una pantalla de búsqueda dónde el usuario pueda filtrar las noticias por las distintas opciones a su gusto fecha, federación y deportista.

Para la mejora del motor de búsqueda, existe la posibilidad de mejorar el algoritmo con la introducción de un corpus lingüístico genérico de la lengua castellana para clasificar correctamente cualquier tipo de palabra en la noticia. Cabe destacar, que esto sólo sería posible para las tablas VERBOP y VERBON para así realizar un análisis morfológico y sintáctico más exhaustivo de los verbos.

La actual configuración del algoritmo, permite crear distintos criterios de búsqueda, en distintos ámbitos periodísticos, definiendo las estructuras de los titulares en dos tipos de fuentes para reconocer las noticias válidas. De este modo, el usuario del algoritmo, podrá modificar en cualquier momento dichos criterios, fuentes de búsqueda y/o sujetos de búsquedas.

Debido a la posibilidad de la existencia de administradores inexpertos que quieran modificar los criterios de búsqueda, sería conveniente crear una serie de opciones de mantenimiento de la base de datos para que éstos puedan modificar desde la propia plataforma web vez los criterios de búsquedas, de modo que puedan insertar, modificar o eliminar fuentes de información, deportistas, logros, etc...

Una vez implantado el sistema en un dominio se procederá a las opciones de SEO para su posicionamiento en Internet. Estas opciones principalmente son dos, construir direcciones amigables y facilitar a los grandes buscadores de Internet el mapa web de la página.

Debido al gran auge de los *"Smartphone"*, la telefonía móvil e Internet móvil con todas sus aplicaciones en los distintos sistemas operativos (Symbian, Mac OS, Android,...) este proyecto deberá servir de base para lanzar una versión en este tipo de plataformas. En la actualidad, la plataforma web cuenta con una versión para una correcta visualización en plataformas con sistema Mac OS y smartphone.





# Capítulo 7: Bibliografía

---

Julius R. Oppenheimer (22/04/1904 – 18/02/1967). Físico estadounidense.

Un científico debe tomarse a libertad de plantear cualquier cuestión, de dudar de cualquier afirmación y de corregir errores.



## 7.1. BIBLIOGRAFIA

- [ADO] Asociación de Deportistas Olímpicos españoles [Internet] <<http://www.ado.es>> [26 de julio de 2011]
- [AHO+08] Aho, Alfred V. *Compiladores: principios, técnicas y herramientas*. Pearson Educación. Edición 2ª. ISBN: 97-897-0261133-2.
- [ARS+04] Sánchez Pérez, Arsenio. *Redacción avanzada*. Thomson. Edición 2ª. ISBN: 97-897-0686412-3.
- [BAR] Barrero, David, David Camacho, María D. R-Moreno. *Automatic Web Data Extractio base don Genetic Algotithms and Regular Expressions*.
- [BER+96] Berners-Lee, T., R. Fielding, UC Irvine y H. Frystyk. *Hypertext Transfer Protocol -- HTTP/1.0*. Internet Engineering Task Force. Request for Comments: 1945. Mayo 1996. Disponible [Internet]: <<http://www.ietf.org/rfc/rfc1945.txt>> [26 de julio de 2011]
- [BOE] *Boletín Oficial del Estado*. Ley Orgánica 3/2007, 22 de marzo 2002. Disponible [Internet]: <<http://www.boe.es>> [26 de julio de 2011]
- [CET] Cetinkaya, Ahmet. *Regular Expression Generation through Grammar Evolution*.
- [COE] Comisión Mujer y Deporte del Comité Olímpico Español [Internet] <<http://www.mujerydeporte.org>> [26 de julio de 2011]
- [CSD] Consejo Superior de Deportes [Internet] <<http://www.csd.gob.es>> [26 de julio de 2011]
- [DPD06] Real Academia de la Lengua Española. *Diccionario Panhispánico de dudas*. Edición 2ª. 2006. ISBN: 97-884-2940623-8. Disponible [Internet]: <<http://www.rae.es>> [26 de julio de 2011]
- [DRAE01] Real Academia de la Lengua Española. *Diccionario de la Lengua Española*. Edición 22ª. 2001. ISBN: 97-884-2396813-8. Disponible [Internet]: <<http://www.rae.es>> [26 de julio de 2011]
- [ECK+07] Eckel, Bruce. *Piensa en JAVA*. McGraw Hill 1ª Edición. 2007. ISBN: 97-884-8966034-2.
- [FIE+99] Fielding, R., J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach y T. Berners-Lee. *Hypertext Transfer Protocol -- HTTP/1.1*. Internet Engineering Task Force. Request for Comments: 2616. Junio 1999. Disponible [Internet]: <<http://www.ietf.org/rfc/rfc2616.txt>> [26 de julio de 2011]
- [FIE+06] Friedl, Jeffrey E.F. *Mastering regular expressions*. O' Reilly. Edición 3ª. ISBN: 97-805-9625812-6.

## BIBLIOGRAFÍA

- [GON+09] González Bodega, Patricia. *Técnicas de etiquetado y desambiguación morfológica del castellano con reducida información contextual*. Proyecto final de carrera: Ingeniería Técnica en Informática de Gestión [Internet] <<http://hdl.handle.net/10016/6687>> [26 de julio de 2011]
- [GOO] Google Inc. Información corporativa. Disponible [Internet]: <<http://www.google.com/intl/es/corporate/tech.html>> [26 de julio de 2011]
- [GUI+11] Guía para la incorporación de la perspectiva de género a la gestión deportiva local. ISBN: 78-84-92494-25-5.
- [IM] Instituto de la Mujer [Internet] <<http://www.inmujer.es>> [26 de julio de 2011]
- [MAR+11] Marriot, Jennifer y Waring, Elin. *El libro oficial de Joomla!*. Anaya Multimedia. 1ª Edición. ISBN: 97-884-4152959-5.
- [MAT+08] Mateos Barrado, Juan Félix. *Joomla! 1.5.X (Guía práctica)*. Anaya Multimedia. 1ª Edición. ISBN: 97-884-4152456-9.
- [MEN+02] Mencher, Melvin. *News reporting and writing*. McGraw Hill. Edición 1ª. ISBN: 97-800-7256497-6.
- [MI] Ministerio de Sanidad, Política Social e Igualdad: Área de Igualdad. [Internet] <<http://www.migualdad.es>> [26 de julio de 2011]
- [NIE+09] Nieto Vidaurrázaga, Ignacio. *Memetracker – TDT: Detección y seguimiento de temas en noticias y blogs*. Proyecto final de carrera: Ingeniería Técnica en Informática de Gestión. [Internet] <<http://hdl.handle.net/10016/7580>> [26 de julio de 2011]
- [RAE09] Real Academia de la Lengua Española y Asociación de Academias Americanas. *Nueva Gramática de la Lengua Española*. Edición 1ª. 2010. ISBN: 97-884-6703207-9.
- [RSS] *RSS Advisory Board*. Disponible [Internet]: < <http://www.rssboard.org/rss-specification> > [26 de julio de 2011]
- [SAN+88] Sanchís Llorca, F.J. *Compiladores: teoría y construcción*. Paraninfo 2ª Edición. ISBN: 84-428-31469-1.
- [W3C] *World Wide Web Consortium*. Disponible [Internet]: <<http://www.w3c.es>> [26 de julio de 2011]
- [XHE] Xhemali, Daniela, Christophjer J.Hinde, Roger G.Stone. *Genetic Evolution of Regular Expressions for Automated Extraction of Course Name from the Web*.

# Apéndices

---

Gregorio Marañón (19/05/1887 - 27/03/1960). Médico español.

Toda la historia del progreso humano se puede reducir a la lucha de la ciencia contra la superstición.



## APÉNDICE A: CONCLUSIONES DEL SEMINARIO PERMANENTE MUJER Y DEPORTE

En el siguiente apéndice, se adjuntan las conclusiones del seminario permanente “*Mujer y Deporte*” celebrado en la facultad de Ciencias de la Actividad Física y del Deporte (INEF) de la Universidad Politécnica de Madrid en las jornadas sobre “*El deporte femenino en los medios de comunicación*” celebrado en la semana del 5 al 11 de noviembre de 2009.

### Presentación

El Seminario Permanente Mujer y Deporte de la Facultad de Ciencias de la Actividad Física y del Deporte-INEF de la Universidad Politécnica de Madrid, organiza las Jornadas “El deporte femenino en los medios de comunicación”, con la colaboración del Instituto de la Mujer y del Consejo Superior de Deportes, haciéndose eco de la problemática sobre el tratamiento que los medios de comunicación hacen del deporte femenino y las repercusiones que este hecho tienen en la práctica deportiva de las deportistas y de las mujeres en general; y por considerar que la información sobre el deporte femenino y la imagen de las deportistas resultan prioritarios para el conocimiento, difusión y creación de modelos femeninos en el deporte.

No es la primera vez que el Seminario Mujer y Deporte se interesa por este tema y lo ha tratado monográficamente en la Convocatorio de 1996 y, posteriormente, en la del año 2000, a través de una ponencia con ocasión del tratamiento del deporte de alta competición.

En esta ocasión, pretendemos difundir los estudios que se han venido haciendo estos últimos años, así como profundizar en aquellas cuestiones que consideremos tienen una mayor incidencia en el desarrollo y promoción del deporte femenino. También se abordan temas relacionados con la formación de los y las profesionales de la comunicación y con el impacto que en la práctica deportiva de las mujeres tiene la falta de modelos deportivos femeninos.

Es objeto, además, de estas Jornadas analizar la información deportiva desde la perspectiva del lenguaje escrito e icónico y la presencia de estereotipos sociales de género en dicha información.

Con la organización de estas Jornadas específicas el Seminario Mujer y Deporte pretende, además de estudiar en profundidad las causas y consecuencias de la situación existente en la actualidad, analizar las posibles estrategias que deberían implementarse para llegar a un tratamiento equitativo de la información deportiva en los medios de comunicación. Entre estas estrategias parece fundamental incluir formación inicial para los/as periodistas e informadores/as, por lo que se ha contado con la colaboración especial de las Facultades de Ciencias de la Información y de profesionales de los diferentes medios de comunicación.

### Objetivos

1. Estudiar desde la perspectiva de género el tratamiento de los medios de comunicación (TV, Radio y Prensa escrita) dan al deporte.
2. Valorar el interés que suscita la información sobre el deporte de las mujeres.

3. Destacar la influencia que los medios de comunicación tienen en la visibilidad de las mujeres en el deporte y en la creación de un modelo de mujer asociado a la actividad física y el deporte.
4. Analizar la repercusión que tiene en la promoción de las actividades físico-deportivas femeninas la escasa atención que los medios de comunicación dan al deporte de las mujeres.
5. Valorar la imagen de la mujer deportista transmitida por los medios de comunicación.
6. Definir estrategias para promover el interés de los medios de comunicación y de los espectadores por el deporte de las mujeres.

## Conclusiones

1. Se confirma que la información deportiva femenina es muy escasa en los diferentes medios de comunicación, fundamentalmente en televisión y prensa escrita. Solamente cuando hay éxitos muy importantes se incluye a las mujeres que los obtienen, aunque su tratamiento suele ocupar menor tiempo y valor que cuando se tratan los éxitos de los hombres.
2. Los motivos que apuntan los y las profesionales de los medios de comunicación en relación con el tratamiento que dan al deporte femenino son: el desinterés de las audiencias, menores éxitos del deporte femenino, las políticas internas de quienes dirigen, la comercialización de la información, la menor presencia de mujeres en los puestos de dirección e, incluso, la inercia en el quehacer informativo.  
**No obstante, se ha demostrado que en nuestro país el éxito de las mujeres en algunos deportes es igual o superior al de los hombres y, a pesar de ellos, las mujeres siguen sin tener apenas cabida en los medios.**
3. La repercusión mediática de la información deportiva tiene influencias importantes en el desarrollo deportivo, profesional y personal de cada deportista y en la expansión política y económica de cada deporte. En consecuencia, se considera que una menor atención de los medios a las deportistas redundará en una desigualdad de oportunidad para ellas.
4. Se denuncia el efecto fagocitador del fútbol en los medios de comunicación y se reclaman más espacios informativos para otros deportes minoritarios y para el deporte femenino.
5. Las consecuencias que tiene la poca presencia de las deportistas y del deporte femenino en los medios de comunicación están relacionadas con la ausencia de modelos femeninos en el deporte, que repercute en la falta de identificación de las niñas y de las jóvenes con la práctica deportiva y que tiene como consecuencia menores niveles de práctica de actividades físicas por parte de la población femenina.
6. Se observa que el tratamiento que los medios de comunicación dan al deporte femenino suele estar impregnado de estereotipos de género y la información tiende a resaltar más a la mujer que a la deportista. Se denuncia una creciente tendencia a presentar a las mujeres deportistas como iconos eróticos más que como ídolos deportivos en sentido estricto.
7. Se detecta la ausencia de formación con perspectiva de género en los y las profesionales de la comunicación. En los estudios de Ciencias de la Información no



- existen materias específicas que sensibilicen sobre estos temas y que proporcionen conocimientos y estrategias al respecto.
8. Los medios de comunicación utilizan un lenguaje sexista en la información deportiva que se caracteriza por la invisibilidad y el estereotipo, una identificación femenina nula o distorsionada, la aplicación de una mirada androcéntrica a los temas y su tratamiento, una subordinación temática, sintáctica y simbólica de las mujeres y una banalización u ocultación de sus actuaciones, logros y declaraciones.
  9. Entre las estrategias que se apuntan para la búsqueda de un tratamiento equitativo en la información deportiva señalan las siguientes:
    - Sensibilizar y formar a los y las profesionales de la comunicación.
    - Implementar normas que aseguren la equidad en la información.
    - Elaborar “guías de buenas prácticas” para el ámbito profesional de la información y la comunicación.
    - Establecer el uso de un lenguaje no sexista.
    - Eliminar la influencia de los estereotipos de género en la información deportiva.
    - Inducir a los gabinetes de prensa de los organismos y administraciones deportivas a que visibilicen a las mujeres deportistas y difundan los éxitos de las mujeres en la misma medida que los de los hombres.



## APÉNDICE B: HEROÍNAS FUERA DE FOCO (08.01.2011)

El siguiente apéndice es una noticia publicada en el diario ABC por el periodista Raúl Cosín en el que describe la situación del baloncesto femenino durante la celebración de la Copa de la Reina'11 en Valencia justo tres meses después de conseguir la medalla de bronce en el Mundobasket'10 de la República Checa.

### Heroínas fuera de foco

Poco después de que la selección española de baloncesto femenino lograra la medalla de bronce en el Mundial de la República Checa en octubre de 2010 —la mejor clasificación en la historia del combinado nacional—, los focos se apagaron. Pasado el gran partido y el triunfo, las chicas volvieron a la rutina de la competición doméstica, que queda en la sombra y en un «tercer» plano de la ventana de los medios. Los halagos, las alabanzas mundialistas a las féminas de la selección, se quedaron en un efecto efervescente. No hay luces. No hay un seguimiento destacado. Ni un merecido reconocimiento para el día a día no sólo de las doce jugadoras que componen la mejor generación del basket femenino español, sino para la Liga profesional, las participaciones de los clubes con presencia en competiciones europeas o un torneo como la Copa de la Reina, que se celebra este fin de semana en Valencia.

Las doce jugadoras que ganaron el bronce disputarán el trofeo de la Reina en las filas de cuatro de los clubes referencia de España: el Ciudad Ros Casares Valencia —vigente campeón copero y que busca su quinto título consecutivo—, el Perfumerías Avenida de Salamanca, el Rivas Ecópolis y el Mann Filter Zaragoza. Nadie mejor que las propias baloncestistas para poner palabras a las pocas luces y las muchas sombras en las que se mueve el baloncesto femenino, aunque la historia se repite en otras muchas disciplinas deportivas en categoría femenina.

La pívot internacional española Cindy Lima, de padre congoleño y madre cubana, es un referente para el deporte baloncestístico. La barcelonesa, hoy en las filas del Ros Casares, dejó las canchas a nivel profesional durante unos años para trabajar como modelo en París. No le fue mal, pero anhelaba la competición, el trabajo sobre la pista, pese al poco reconocimiento que tiene el basket femenino. «En líneas generales, el tratamiento que se le da en España al deporte femenino es paupérrimo, casi inexistente. Tienes que ganar el campeonato “interestelar” para tener cinco minutos de presencia en los medios, sobre todo a nivel nacional», explica la pívot.

La jugadora del Ros Casares también lanza un mensaje hacia el estamento político en su discurso y lamenta que «es una pena. Había un Ministerio de Igualdad y otras tantas cosas, pero a la hora de la verdad queda mucho camino que recorrer. Creo que todavía estamos abriendo brecha nosotras para las que vengan por detrás, porque incluso a nivel selección, que subimos al podio casi siempre, no hay mucha cobertura. De hecho hay gente que no sabe siquiera que hay una liga profesional».

### Ni en Europa

Sólo las participaciones en mundiales o europeos con la Selección española abren la ventana mediática. Pero esto es más costoso, ya no sólo para las jugadoras internacionales, sino también a los clubes y a la propia existencia de una competición nacional profesionalizada.

Laia Palau, capitana del club anfitrión de la Copa de la Reina, advierte que hay gente que piensa que sólo jugamos con la Selección. La verdad es que la camiseta de la roja vende mucho, pero a nivel de clubes nada». La base catalana explica un caso palmario de falta de seguimiento a este deporte: Recuerdo el año pasado que llegamos a la “Final Four” en Euroliga y prácticamente no tuvimos repercusión, jugando una final de la Copa de Europa, en la que estaban muchas de las mejores jugadoras del mundo».

No obstante, el color rojo de la camiseta española también tiene un proceso efervescente. Para la base Nuria Martínez, con la Selección nos dan cobertura, pero tampoco es una cosa destacada. Hicimos bronce en un Mundial, que es lo más grande que ha conseguido el baloncesto femenino español y nos dieron un punto de cobertura y se evaporó enseguida. En la misma línea opina la hermana del jugador de la NBA Rudy Fernández, Marta. La escolta es concisa y asegura que cuando hemos jugado con la selección se nos ve en el momento, pero luego se olvida rápido.

Son momentos puntuales. Destellos de interés. «Cuando comienza un campeonato, no tiene mucha repercusión. La gente valora su importancia cuando alcanzas los cuartos, pasas a semifinales y se cae en la cuenta de que se puede hacer historia, entonces se llega a un punto de que se encuentra apoyo por todas partes», concluye Lucila Pascua, pívot del Mann Filter Zaragoza.

Todas reconocen que se van dando pasos, pero la realidad es que todavía queda mucho camino por recorrer. Aunque hoy todo es un poco mejor que ayer. «Hasta hace poco, había gente que nos quería seguir y tampoco podía. Mis familiares me preguntaban muchas veces: “¿Dónde dan el partido? Pero yo les decía que ya se lo contaría yo a la vuelta. Va habiendo iniciativas, pero va muy poco a poco», asegura Cindy Lima.

### **Mucho trabajo por hacer**

Las «chicas de bronce» llamaron por unos días la atención de la actualidad deportiva española, pero reconocen que no es fácil competir con deportes como el fútbol, el baloncesto masculino, la Fórmula 1 o el tenis, ni siquiera con cualquier deporte en su categoría masculina. Al fin y a la postre, actualmente, en todas estas disciplinas España está a la cabeza mundial «Nosotras vamos dando pasitos de hormiga, poco a poco, y tenemos que seguir así. Estamos haciendo mucho para el baloncesto femenino y recogiendo apoyos y reconocimiento», explica animada la capitana de la Selección española, Amaya Valdemoro, que recuerda de su paso por Rusia que la cadena de televisión pública, que tenía los derechos de todas las competiciones, seguía mucho el basket.

Valdemoro, referente del basket español, recuerda que, aunque los equipos punteros «tienen a sus jugadoras bien pagadas en tiempo y forma; los problemas son mayores en clubes de mitad de tabla. El Rivas nos paga muy bien, no hay retrasos y estamos contentas», explica la capitana de la roja, quien recuerda que «algunos clubes tienen serios problemas».

Se pudiera pensar que la situación para el baloncesto femenino en España es la que es y que está al nivel de otros países europeos. Sin embargo, se dan situaciones para todos los gustos. Cindy Lima, antes de fichar por Ros Casares, tuvo experiencias en el TTT Riga (Letonia) y en el

KV Imperial AEL (Chipre). «En Letonia, por ejemplo, las chicas de la selección, que nunca han ganado nada, iban a todas partes y todo el mundo sabía quiénes eran, estaban muy protegidas por la federación, tienen mucha presencia,... yo aluciné», explica la pívot.

Por su parte, Nuria Martínez, que retornó a España procedente del Bracco Geas Sesto San Giovanni, opina que «en Italia es parecido a España. En medios locales y ciudades pequeñas es más fácil que el deporte femenino salga a la luz. Pero donde todo es más grande, en un lugar como Moscú —donde también ha jugado la española—, la repercusión era nula.

La experiencia de Marta Fernández en el Wisla de Cracovia sí que le dejó un buen sabor de boca respecto al respaldo que encontraban las jugadoras: «En Polonia sí que había mucha repercusión. Estuve en Cracovia, donde no había equipo masculino de primer nivel, y la ciudad estaba volcada con el equipo femenino. Esto sí que se ve en Salamanca. Pero nada es comparable con lo que se puede ver en Estados Unidos con la WNBA: «Allí es otra cosa. Es un formato NBA femenina. Es diferente. También los meses en los que se juega son diferentes. Es todo mucho más grande, más espectacular», indica la escolta de la Selección.

Con todo, las jugadoras entienden que dada las circunstancias su deporte tiene un mejor y mayor reconocimiento que otras disciplinas deportivas en féminas. «El basket femenino es dentro del deporte femenino el que está mejor tratado», a lo que Cindy Lima apostilla que «lo mejor tratado es la natación sincronizada», para los que reconocen la carrera de Gemma Mengual y apuntan: «Mira lo que ha tenido que hacer: ganarlo todo».

### **Las 12 se reúnen en Valencia**

Las doce jugadoras que lograron para España la medalla de bronce en el Mundial de la República Checa se miden este fin de semana por la Copa de la Reina, que se celebra en Valencia. Elisa Aguilar, Anna Cruz, Laura Nicholls y Amaya Valdemoro, del Rivas Ecópolis; Lucila Pascua, del Mann Filter Zaragoza; Sancho Lyttle, Anna Montañana y Alba Torrens, del Perfumerías Avenida de Salamanca; y Marta Fernández, Cindy Lima, Nuria Martínez y Laia Palau, por el Ros Casares, se juegan la Copa. Cuarenta medios —en la Copa del Rey, la de los chicos, hay hasta la fecha 150 medios acreditados— seguirán el torneo, que abre hoy con el Ros Casares-Zaragoza, con el Salamanca-Rivas en la otra semifinal.



## APÉNDICE C: GRAMÁTICA ESPAÑOLA

En el siguiente apéndice se va a tratar la gramática española, estudiándola de una forma inductiva desde lo más pequeño, la palabra, a lo más genérico la oración, dando unas nociones básicas sobre las palabras y sus categorías, la oración o proposición y su estructura para ayudar a comprender el algoritmo de selección de noticias.

### C.1. LA PALABRA

En la gramática tradicional, una palabra es cada uno de los segmentos limitados por pausas o espacios en la cadena hablada o escrita, que puede aparecer en otras posiciones, y que está dotado de una función. Existen varias formas de clasificar las palabras:

- Morfosintácticamente:
  - Según su categoría gramatical.
  - Según su estructura interna: simple, derivada, compuesta o parasintética.
- Fonológicamente:
  - Según su acentuación: aguda, llana, esdrújula o sobreesdrújula.
  - Según el número de sílabas: monosílaba, bisílaba, trisílaba,...

Para este proyecto interesa la clasificación según su categoría gramatical, por lo que a continuación se profundiza en dicha clasificación.

#### C.1.1. El sustantivo

Son las clases de palabras mediante las que se designan objetos y seres pertenecientes a la realidad extralingüística.

Además, no solo designa entes físicos, sino cualidades, propiedades, acciones, procesos o relaciones si son pensados como conceptos independientes. Tipos:

- *Comunes*: designan clases de seres, es decir, agrupan bajo una misma denominación conjuntos de objetos o seres característicos comunes, por tanto son clasificadores. El significado de un nombre común contiene los rasgos semánticos que definen el objeto. Ej: *Árbol, casa, ventana...*
- *Propios*: designan seres únicos. Señalan y distinguen un individuo particular entre los restantes de su clase. Por tanto, se señalan e individualizan para que no se confundan unos con otros. Se distinguen:
  - *Antropónimos*: nombres de persona. Ej: *Almudena, Richard, Begoña...*
  - *Topónimos*: nombres de lugar. Ej: *España, Madrid, Leganés...*
- *Concretos*: designan objetos o realidades físicas, perceptibles por los sentidos y que se pueden representar mentalmente por imágenes. Se tienen:
  - *Genéricos*: limitados en forma y extensión.
    - *Individuales*. Ej: *Cerdo...*
    - *Colectivos*: designan a un conjunto de objetos con las mismas características. Ej: *Piara...*
  - *De materia*: designan sustancias continuas.

- *Abstractos*: designan objetos mentales solo captables por la inteligencia y representados mediante objetos. Sus referentes no tienen existencia independiente de la realidad.
  - *De cualidad*: si derivan de un objeto.
  - *De fenómeno*: si derivan de un verbo.
  - *Cuantitativos*:
    - *Precisos*: cualidades precisas.
    - *Imprecisos*: cualidades imprecisas.

### C.1.2. El pronombre

Son morfemas gramáticos libres y susceptibles que desempeñan cualquier función idéntica a la del sustantivo.

Pueden actuar como pronombre sustantivo, es decir, si funciona como núcleo del sintagma nominal o pronombre adjetivo, si funciona como determinante, bien sea como actualizadores o cuantificadores (excepto pronombres personales).

Corresponden a esta categoría de pronombres aquellas palabras que remiten a un nombre como sustitutivos del mismo. Carecen de función específica y ocupan la del sustantivo al que hacen referencia. Su significado es ocasional y depende del contexto o situación.

Todos tienen una función deíctica ya que hacen referencia o señalan a una realidad extralingüística. La deixis suele ser absoluta cuando hace referencia al hablante. El pronombre carece de determinación y solo puede ser determinado por algunos términos que redundan el significado apartado por él.

Si se quiere agregar alguna determinación al pronombre, se reúne esencialmente a aposiciones que actúan como términos adyacentes y cuya misión es completar el significado al que alude el pronombre.

Tipos:

- *Personales*: aquellos que designan la persona gramatical que realiza la acción expresada por el verbo. Ej: *me, te, se, nos, os, yo, tú, él...*
- *Posesivos*: indican relación con los pronombres gramáticos, cuya identidad queda definida en la 1ª, 2ª y 3ª persona. Tienen función deíctica en el discurso y pueden adquirir valores anafóricos. Ej: *mío, tuyo, suyo, nuestro,...*
- *Demostrativos*: son deícticos porque sitúan el espacio a las personas gramaticales. Son actualizadores y suelen tener valores catafóricos. Ej: *éste, ese, aquél...*
- *Indefinidos y cuantitativos*: se dividen en:
  - *Cuantitativos*: si modifican al sustantivo expresando cantidad no precisa. Ej: *varios, muchos, poco, bastante...*
  - *Identificativos*: implican identidad o semejanza. Ej: *mismo, misma, otro, propio, tal...*
  - *Existenciales*: indican existencia de lo designado. Ej: *alguien, alguno, nadie, cierto, quienquiera...*



- *Numerales*: son cuantitativos que expresan cantidad precisas. No guardan relación con las personas gramaticales y tienen significación permanente salvo los ordinales que tienen significado ocasional.
  - *Cardinales*. Ej: uno, dos, tres...
  - *Ordinales*. Ej: primero, segundo, tercero...
  - *Múltiplos*. Ej: doble, triple, cuádruple...
  - *Partitivos*. Ej: medio, mitad, tercio...
- *Interrogativos, exclamativos, relativos*.
  - *Interrogativos*: son sustitutos de un elemento desconocido que el oyente debe aportar. Ej: qué, cuál, cuánto, quién...
  - *Exclamativos*: adquieren un significativo ponderativo y puede funcionar como pronombre adjetivo y pronombre sustantivo. Ej: Qué, cuánto...
  - *Relativos*: son sustitutos relacionantes. El elemento reemplazado se denomina antecedente del relativo. Ej: qué, el cuál, quién, cuyo...

### C.1.3. El determinante

Palabra que tiene como función introducir el nombre en la oración y precisar su extensión significativa, señalando cuáles o cuántas de las entidades designadas por el nombre se refiere el que habla, o bien si estas son o no conocidas o consabidas por los interlocutores.

Tipos:

- *Artículos*: determina al sustantivo, lo restringe y precisa. Ej: el, la, los las.
- *Demostrativos*: sitúa el nombre en un entorno espacial o temporal. Ej: este, ese, aquel...
- *Posesivos*: establecen una relación deíctica respecto de las personas del discurso. Ej: mi, tu, su...
- *Cuantificadores*: indican la cantidad de elementos del sintagma nominal. Pueden ser indefinidos si es aproxima, o numerales si es concreta. Ej: alguno, mucho, poco, dos tres, primero...

### C.1.4. El adjetivo

Desde el punto de vista formal, el adjetivo está constituido por un lexema y una serie de morfemas inherentes de género y número.

Desde el punto de vista semántico, el adjetivo designa las cualidades de la realidad aludida al sustantivo que le acompaña.

Desde el punto de vista funcional, el adjetivo es un término adyacente al núcleo de un sintagma nominal.

### C.1.5. El adverbio

Es la parte invariable de la oración, que sirve para calificar o determinar la significación, de un adjetivo o de un verbo. Funciona como núcleo del sintagma nominal. Tipos:

- *De lugar*:
  - *Demostrativos*: aquí, ahí, allí, acá, allá...

- *De situación locativa*: cerca, lejos, arriba, debajo, dentro, fuera, encima...
- *Relativos*: donde.
- *De tiempo*:
  - *De identificación u orientación temporal*: hoy, mañana, ayer, anoche...
  - *De matiz temporal*: nunca, ya, aún, todavía...
  - *De situación temporal*: pronto, tarde, temprano...
  - *De orden temporal*: antes, después...
  - *Relativos*: cuando.
- *De modo*:
  - *De modo*: así, bien, mal, regular, peor, mejor...
  - *Relativos*: como.
- *De cantidad*: más, menos, poco, bastante, mucho, muy, demasiado...
- *De afirmación/negación*: sí, también, no, nunca, jamás...
- *De duda*: acaso, quizás, tal vez...

### C.1.6. El verbo

Es la palabra que denota acción, estado o proceso, capaz de funcionar como núcleo del predicado y cuyas desinencias expresan modo, tiempo, número y persona. Tipos:

- *Auxiliar*: El que sirve para formar los tiempos compuestos de los verbos, la pasiva perifrástica y las perífrasis verbales: HE terminado; FUISTEIS premiados; HEMOS de trabajar; TIENE que comer...
- *Causativo*: los que tienen sentido o valor causativo cuando la entidad designada por el sujeto no realiza por sí misma la acción de la que se habla, sino que la ordena o la encarga a otros. Ej: cortarse el pelo...
- *Copulativo*: El que, prácticamente vacío de significado léxico, sirve de unión entre un sujeto y un atributo y admite que este sea sustituido por el pronombre neutro *lo*. Ej: *ser, estar y parecer*.
- *De afección psíquica*: El que denota procesos que afectan al ánimo o producen acciones o reacciones emotivas. Ej: aburrir, divertir, admirar, temer...
- *De estado*: el que implica estado o situación. Ej: permanecer, sentarse...
- *Defectivo*: el que no se conjuga en todos los tiempos y personas. Ej: soler, preterir...
- *De influencia*: el que expresa una acción que tiene como objetivo influir en alguien para que haga o deje de hacer algo. Ej: aconsejar, exhortar, obligar, ordenar...
- *Impersonal*: el que carece de sujeto, sea tácito o expreso.
- *Intransitivo*: el que no puede llevar complemento directo. Ej: ir, nacer...
- *Irregular*: el que, al conjugarse, sufre cambios en la raíz en algunas de sus formas o toma desinencias distintas de las del modelo regular que le corresponde por su terminación. Ej: nevar, conducir...

- *Persona*: el que lleva sujeto, sea tácito o expreso.
- *Pronominal*: el que se construye en todas sus formas con un pronombre átono que concuerda con el sujeto y que no desempeña ninguna función sintáctica oracional. Algunos verbos son exclusivamente pronominales, como *arrepentirse* o *vanagloriarse*, y otros adoptan determinados matices significativos o expresivos en la construcción pronominal, como *caerse* o *morirse*, frente a *caer* o *morir*.
- *Regular*: el que se ajusta en todas sus formas al modelo fijado como regular que le corresponde por su terminación.
- *Semicopulativo*: los verbos que sirven de enlace entre un sujeto y un atributo, pero no admiten la sustitución de este por el pronombre *lo*. Ej: *andar*, *seguir*, *volver*...
- *Terciopersonal*: verbo de cuyas formas personales se usan solo las de tercera persona del singular y del plural. Ej: *ocurrir*, *atañer*...
- *Transitivo*: el que se construye con complemento directo, Ej: *tener*, *decir*...

### C.1.7. Las proposiciones

Son palabras invariables que introducen un complemento obligatorio y establecen una relación de subordinación o dependencia de ese término y otro elemento de la oración. Ej: *a*, *ante*, *bajo*, *cabe*, *con*, *contra*, *de*, *desde*, *hasta*, *para*, *por*, *según*, *sin*, *so*, *sobre*, *tras*.

### C.1.8. Las conjunciones

Son palabras invariables que relacionan unidades sintácticas. Tipos:

- *Coordinantes*: unen dos o más unidades del mismo nivel jerárquico.
  - *Causal*: *porque*, *pues*.
  - *Copulativas*: *y*, *e*, *ni*.
  - *Disyuntivas*: *o*, *u*, *o bien*.
  - *Adversativas*: *pero*, *más*, *sino*, *aunque*.
  - *Distributivas*: *ya... ya*, *bien... bien*.
- *Subordinante*: introducen una oración en verbo personal que convierten en dependiente de otra oración o de alguna palabra de otra oración. Ej: *que*, *cuando*, *donde*, *como*, *porque*, *si*...

### C.1.9. Las interjecciones

Son palabras invariables, con autonomía sintáctica, con la que el hablante expresa sentimientos o sensaciones, o induce a la acción al interlocutor. Ej: ¡*ay!*, ¡*oh!*, ¡*cielos!*, ¡*hola!*...

## C.2. LA PROPOSICIÓN

En castellano, la unión de palabras propicia la construcción de oraciones o proposiciones, es la unidad lingüística de estructura oracional. Una proposición consta de dos partes, sintagma nominal y un sintagma verbal. Cada uno de ellos cumple una función estratégica en la proposición. Por extensión, las palabras que conforman los sintagmas reciben también el nombre de estas funciones, que son:

### C.2.1. Sujeto

Función sintáctica ejercida por la palabra o grupo de palabras de cuyo referente se predica algo. El sujeto está compuesto por un sintagma nominal.

### C.2.2. Predicado

Función sintáctica ejercida por la palabra o grupo de palabras que predicen algo del sujeto; cuyo núcleo, generalmente un verbo, que concuerda con el núcleo del sujeto

De este modo, la estructura de una oración se podría representar gráficamente mediante la siguiente estructura árbol.

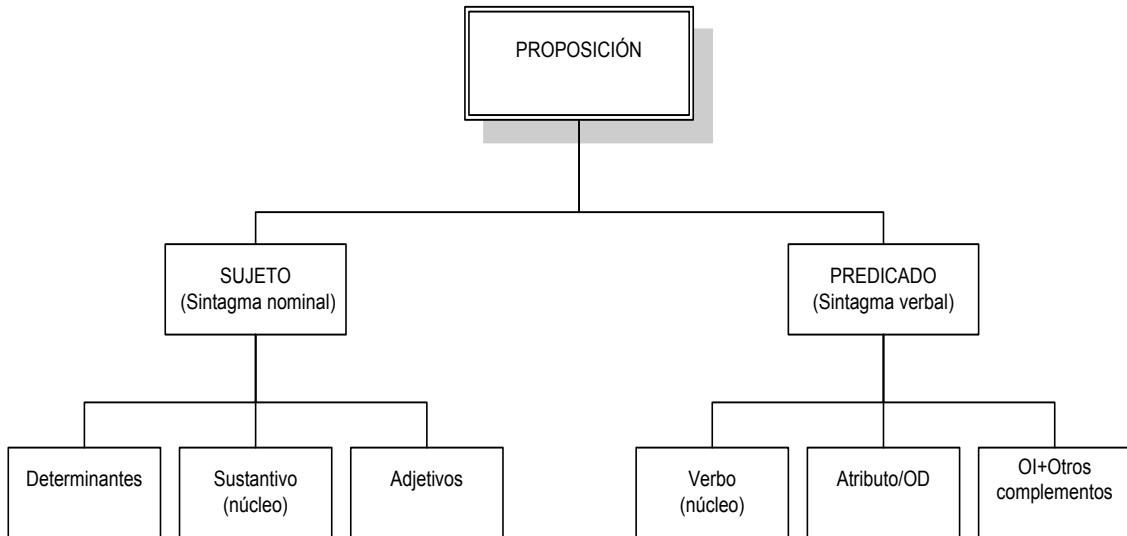


Ilustración 59 Estructura de una oración

Siendo:

- *Atributo*: Si el verbo es copulativo o semicopulativo.
- *OD*: Objeto Directo. Si el verbo es transitivo.
- *OI*: Objeto Indirecto. Si el verbo es intransitivo
- *Otros complementos*: Sintagmas nominales, preposicionales, adjetivales o adverbiales que añaden información sobre la acción realizada por el sujeto, como lugar, tiempo, etc....

## APÉNDICE D: BREVE HISTORIA DE LOS BUSCADORES DE INTERNET

Con la aparición de Internet y su progresión geométrica en cuanto a usuarios y cantidad de información aparece un cuello de botella, la recuperación de datos y la clasificación de dicha información provocando una serie de innovaciones con el paso de los años.

El 30 junio de 1993 aparece el índice “Wandex” realizado por la World Wide Web Wanderer (W4), un robot desarrollado en Perl por Matthew Gray en el MIT. Este índice fue motivado para comprobar el crecimiento de Internet, estuvo en funcionamiento hasta junio de 1996. A día de hoy Matthew Gray trabaja para Google.

Un año después, en 1994, aparece el primer motor de búsqueda de texto completo, WebCrawler en la Universidad de Washington por Brian Pinkerton, que a diferencia del anterior permitía a los usuarios una búsqueda por palabras en cualquier página web, es decir, utilizar el estándar de exclusión de Robots o Spiders, convirtiéndolo en ejemplo para los posteriores. Hoy en día, pertenece a Excite! empresa que nació 1994 con el objetivo de obtener la mejor forma de buscar y recuperar información en la Universidad de Stanford.

En junio del mismo año, Lycos lanza su primer motor de búsqueda creado en la Universidad Carnegie Mellon; en 1998 incorporó a su sistema un software de creación de directorios creando una base de datos con una cantidad inmensa de información.

En 1995 llega AltaVista que incorpora el multilinguaje en los motores de búsqueda utilizando los servicios del primer traductor on-line. Otra de sus novedades fue la búsqueda de contenidos multimedia. Este software fue el primero que rastreaba Internet, indizando y mostrando información recolectada, hoy en día este sistema pertenece a Yahoo!. Por el contrario, Yahoo! es un portal de búsqueda en Internet, lo que implica que principalmente es un directorio Web y no un motor de búsqueda, por lo que los usuarios necesitan estar en la base de datos de este directorio. Desde el 2003, Yahoo! usa su propio motor de búsqueda basado en AltaVista y en Inktomi, otro motor de búsqueda desarrollado por la Universidad Berkeley que aparte de tener un potente motor de búsqueda usa la tecnología “Concept Induction TM” que automáticamente analiza y categoriza millones de documentos confeccionando directorios utilizando las páginas obtenidas por su Spider.

En el año 1997 aparece el gigante Google y no es hasta el año 2000 que incorpora el denominado PageRank siendo usado en el proyecto Open Directory. El PageRank es un algoritmo que asigna un valor numérico en función de enlaces de otras páginas que la referencia, la importancia de las páginas que la referencia y otra serie de criterios no desvelados.

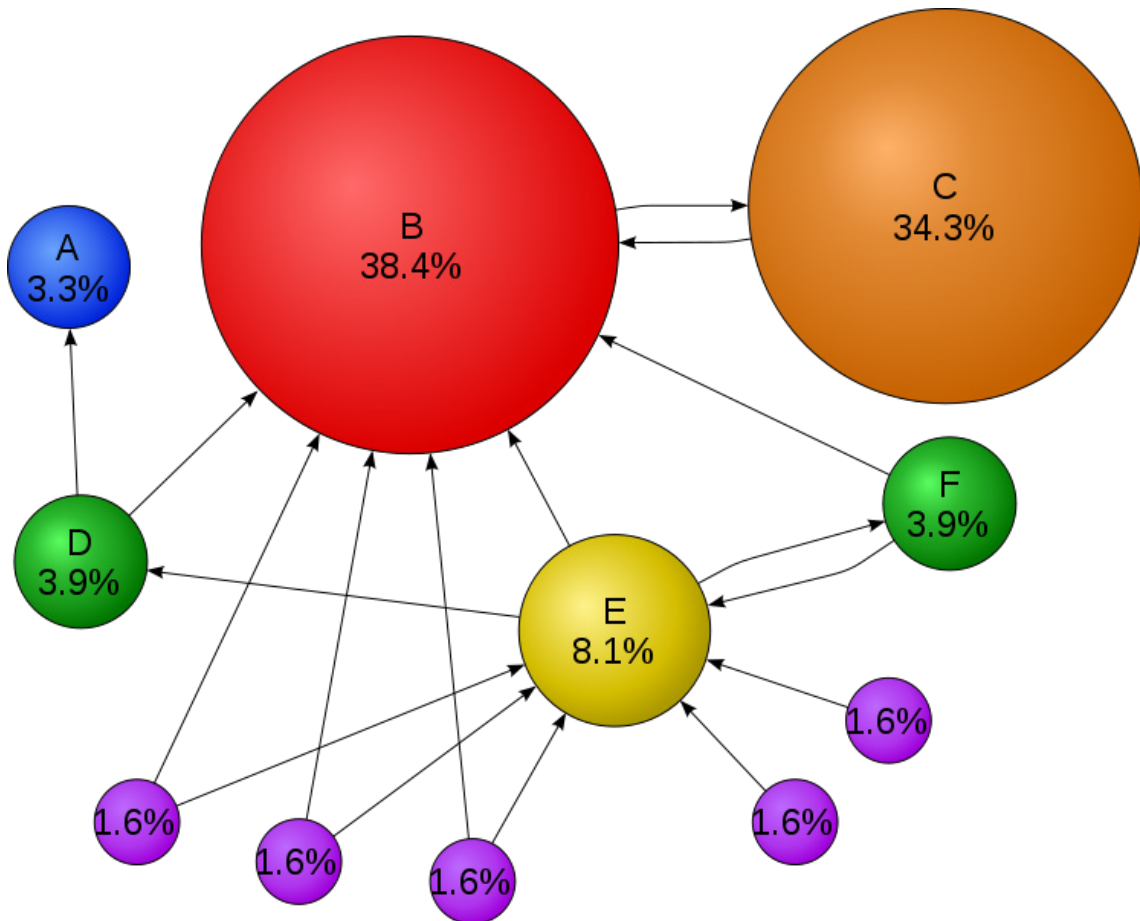


Ilustración 60 Ejemplo de PageRank

El algoritmo inicial del PageRank era:

$$PR(A) = (1 - d) + d * \sum_{i=1}^n \frac{PR(i)}{C(i)}$$

Dónde:

- **PR(A)** es el PageRank de la página **A**.
- **d** es un factor de amortiguación que tiene un valor entre 0 y 1.
- **PR(i)** son los valores de PageRank que tienen cada una de las páginas **i** que enlazan a **A**.
- **C(i)** es el número total de enlaces salientes de la página **i** (sean o no hacia **A**).

Actualmente dicho algoritmo tiene más de 500 millones de variables y 2 millones de términos.

El ODP es un proyecto colaborativo (administrado por Netscape) en el que se categorizan y listan miles de páginas web que dio a luz en 1998. Este gran directorio contiene información de Netscape Search, AOL Search, Google, Lycos, HotBot, DirectyHit y muchos otros.

## APÉNDICE E: Fuentes Web o RSS

Las fuentes web o RSS (Really Simple Syndication) fue creado en 1999 con la idea de compartir información de manera sencilla. Poco a poco fue haciéndose popular hasta que se definió un estándar según la *World Wide Web Consortium (W3C)* en un formato de archivo en lenguaje *XML*.

El lenguaje *XML (eXtensible Market Language)* es un formato de texto derivado del *SGML (Standar Generalized Marked Language, ISO 8879)* que se puede utilizar para el intercambio rápido y sencillo de información mediante el uso de etiquetas.

Actualmente, la versión estable del RSS es la 2.0.11 que sigue la siguiente especificación:

- **xml version:** versión del XML.
- **rss versión:** versión del RSS.
- **Channel:** que está subordinado al campo rss con los metadatos del canal y del contenido de éste.

### *Obligatorios:*

- **Title:** es el nombre del canal, en caso de pertenecer a una página web deberá corresponderse con el nombre de dicha página.
- **Link:** es la URL de la página web al que corresponde el canal.
- **Description:** pequeña descripción sobre el canal.

### *Opcionales*

- **Language:** idioma en el que está escrito el canal. Los valores están definidos por la W3C.
- **Copyright:** derechos de autor del contenido.
- **ManagingEditor:** correo electrónico del responsable de contenidos o del editor.
- **WebMaster:** correo electrónico del responsable de la página web.
- **PubDate:** fecha de publicación siguiendo el estándar RFC 882.
- **LastBuildDate:** fecha de última modificación.
- **Category:** tipo de categoría del canal.
- **Generator:** nombre del programa usado para generar el canal.
- **Docs:** direcciones web con la documentación del formato usado en el RSS.
- **Cloud:** registro para ser notificado de las actualizaciones del canal.
- **Ttl:** número de minutos en los que se encuentra en memoria caché el RSS.
- **Image:** imágenes en formatos GIF, JPEG o PNG que aparezcan en el canal.
- **Rating:**
- **TextInput:** cuadro de diálogo para mostrarse con el canal.
- **SkipHours:** indicación para los lectores automáticos de las horas a evitar.
- **SkipDays:** indicación para los lectores automáticos de los días a evitar.
- **Item:** representa cada uno de los elementos o noticias de un canal. Contiene los siguientes campos:

### *Obligatorios:*

## APÉNDICE E: Fuentes Web o RSS

- *Title*: es el título del ítem.
- *Link*: es la URL del ítem.
- *Description*: contenido de la noticia, o pequeña sinopsis de la noticia.

### *Opcionales:*

- *Author*: correo electrónico del autor del ítem.
- *Category*: categoría del ítem.
- *Comments*: URL con comentarios sobre el ítem.
- *Enclousure*: descripción del contenido multimedia adjunto al ítem.
- *Guid*: identificador del ítem.
- *Pubdate*: fecha de publicación del ítem.
- *Source*: fuente de donde viene el ítem.



## APÉNDICE F: Noticia F.C. Barcelona (22.02.2011)

El siguiente apéndice es una noticia publicada por el gabinete de prensa del F.C. Barcelona en su página web, <[www.fcbarcelona.es](http://www.fcbarcelona.es)> el pasado 22 de febrero.

### Palco de Barça-Zaragoza estará integrada por mujeres

En el próximo partido en el Camp Nou ante el Real Zaragoza, el 5 de marzo, se visualizará la voluntad de la Junta de reivindicar el papel histórico de la mujer barcelonista.

Las acciones para reivindicar el papel que ha tenido la mujer a lo largo de la historia del FC Barcelona y para potenciar su participación en la vida ordinaria del Club se comenzarán a visualizar en el próximo partido de Liga en el Camp Nou. **El próximo sábado 5 de marzo**, con motivo del partido que enfrentará al Barça y el Real Zaragoza, el Palco Presidencial del Camp Nou estará ocupado, por primera vez en la historia, íntegramente por mujeres.

#### Gestiones en marcha

Ocuparán los lugares más preeminentes del Palco Presidencial las directivas del FC Barcelona **Susana Monje** y **Pilar Guinovart**, que estarán acompañadas de las autoridades invitadas, que ese día también serán mujeres, como la representación del Real Zaragoza. El departamento de protocolo del Club también está contactando con los patrocinadores que tienen espacio fijo en Palco para pedirles que ese día deleguen su representación en mujeres de sus respectivas empresas. Asimismo, habrá una amplia presencia de mujeres del ámbito económico, social y cultural de la ciudad. El presidente **Sandro Rosell** presidirá el partido, junto con el presidente del Real Zaragoza, en una ubicación aún por determinar, desde donde presenciarán el partido acompañados del resto de componentes de las directivas de los dos clubes.

#### La primera iniciativa

Esta iniciativa ha sido la primera de las propuestas del grupo de trabajo nombrado por la Junta y que forman las directivas Susana Monje y Pilar Guinovart, las patronas de la Fundación Lola Bou, Rosa María Lleal y Marta Rodríguez de Llauder, y la directora de comunicación del Club, Ketty Calatayud. Este grupo de trabajo ha escogido el día 5 de marzo, tres días antes del Día de la Mujer Trabajadora, como pistoletazo de salida para visualizar esta nueva voluntad.

#### Las mujeres, una cuarta parte de la masa social

La presencia de la mujer ha dotado de una especial singularidad al Club, ya que no hay muchos clubes en el mundo donde la mujer tenga un papel tan destacado en su masa social. Hay que recordar que en estos momentos **el Club cuenta con 42.000 socias, lo que representa casi un 24% de esta masa social.**

#### Precedentes simbólicos

El Club quiere recuperar la memoria de mujeres como **Edelmira Calvetó**, que fue admitida como primera socia en enero de 1913 después de varias peticiones, ya que no estaba previsto que las mujeres pudieran afiliarse al Club, dado que los Estatutos de 1911 sólo admitían socios hombres, o la de **Anna Maria Martínez Sagi**, primera directiva de la historia del Club, en 1934, bajo la presidencia de **Esteban Sala**.

**Monje: "El Barça siempre ha sido un club pionero"**

Para Susana Monje, directiva y una de las impulsoras de este grupo de trabajo, el Barça debe seguir siendo un reflejo y una referencia para la sociedad. "Queremos potenciar el papel de la mujer dentro del Club. El Barça siempre ha sido un Club pionero, que incorporó a la mujer hacia el 1913. El Barça siempre ha sido un reflejo de la sociedad, un Club activo y muy pionero y líder. La Junta quiere hacer iniciativas para que continúe siendo así y se incorpore a la mujer dentro todos los ámbitos del Club. El palco del partido del Zaragoza será la primera acción y servirá para que haya una visualización de cómo la mujer participa en la vida del Club".

## APÉNDICE G: Código

En el siguiente apéndice se van a mostrar el código relevante tanto del algoritmo como de la página web.

### G.1. ALGORITMO DE BÚSQUEDA

En esta primera función, se analiza la estructura del titular de la noticia, localizando cada uno de los elementos que conforman una proposición: sujeto, verbo y complementos.

```

/** Comprueba que es una noticia valida
 * @param noticia: titular de la noticia
 *         sujetos: lista de palabras sujeto
 *         verbosP: lista de palabras verbos positivos
 *         verbosN: lista de palabras verbos negativos
 *         ods    : lista de palabras objeto directo
 * @return exito : true si la noticia es valida, false en caso
contrario
 */
public boolean comprobarNoticia(DatoBE dato, List<String> sujetos,
List<String> verbosP, List<String> verbosN, List<String> ods){

    boolean exito = true;
    String sujeto = "";
    List<Object> result;
    String noticia = dato.getTitular();

    // Busca la existencia de un SUJETO
    Iterator<String> iterador = sujetos.iterator();
    result =buscarPatrones(iterador,noticia,null,1);
    if (((Boolean) result.get(0)).equals(true)){
        // Busca la existencia de un SUJETO + VERBO POSITIVO
        iterador = verbosP.iterator();
        sujeto = (String) result.get(1);
        dato.setSujeto(sujeto);
        result =buscarPatrones(iterador,noticia,sujeto,2);
        if (((Boolean) result.get(0)).equals(true))
            exito=(Boolean) result.get(0);
        else { // Busca la existencia de un VERBO NEGATIVO
            iterador = verbosN.iterator();
            result =buscarPatrones(iterador,noticia,null,3);
            if ((Boolean) result.get(0))
                // Se ha encontrado un verbo NEGATIVO y NO POSITIVO
                exito=false;
            else { // Busca la existencia de un SUJETO+OD
                iterador = ods.iterator();
                result =buscarPatrones(iterador,noticia,sujeto,4);
                if (((Boolean) result.get(0)).equals(true))
                    exito=(Boolean) result.get(0);
                else{ // Busca la existencia de un OD + SUJETO
                    iterador = ods.iterator();
                    result =buscarPatrones(iterador,noticia,sujeto,5);
                    exito=(Boolean) result.get(0);
                }
            }
        }
    }
    }else
        exito = false;
    return exito;
}

```

## APÉNDICE G: Código

Esta segunda función es llamada desde la anterior y va comprobando la existencia de los distintos patrones o estructuras posibles, mediante el uso de expresiones regulares, que se han descrito en el análisis del proyecto.

```
/** Realiza una búsqueda de patrones
 * @param iterador: lista de patrones a buscar
 * @param noticia: cadena donde se buscan los patrones
 * @param sujeto: cadena con el sujeto del titular
 * @param tipo: entero para indicar el tipo de búsqueda a realizar
 * @return lista: posicion 1-> resultado de la búsqueda, posicion 2->
patron encontrado
 */
public List<Object> buscarPatrones(Iterator<String> iterador,String
noticia,String sujeto,int tipo){

    boolean exito=false;
    String patron="";
    Pattern p;
    Matcher m;
    List<Object> result = new ArrayList<Object>();

    while(iterador.hasNext() && exito==false){
        patron=iterador.next();
        switch (tipo) {
            case 1: // Esta buscando el sujeto
                p =
                    Pattern.compile("(.*?)"+patron+"(.*?)",Pattern.CASE_INSENSITIV
E | Pattern.UNICODE_CASE);
                break;
            case 2: // Esta buscando sujeto + verbo
                p =
                    Pattern.compile("(.*?)"+sujeto+"(.*?)"+patron+"(.*?)",Pattern
.CASE_INSENSITIVE | Pattern.UNICODE_CASE);
                break;
            case 3: // Esta buscando verbo negativo
                p =
                    Pattern.compile("(.*?)"+patron+"(.*?)",Pattern.CASE_INSENSITIV
E | Pattern.UNICODE_CASE);
                break;
            case 4: // Esta buscando sujeto + od
                p =
                    Pattern.compile("(.*?)"+sujeto+"(.*?)"+patron+"(.*?)",Pattern
.CASE_INSENSITIVE | Pattern.UNICODE_CASE);
                break;
            case 5: // Esta buscando od + sujeto
                p =
                    Pattern.compile("(.*?)"+patron+"(\\s.*?)"+sujeto+"(.*?)",Patt
ern.CASE_INSENSITIVE | Pattern.UNICODE_CASE);
                break;
            case 6: // Esta buscando sujetos no maliciosos
                p =
                    Pattern.compile("(.*?)"+patron+"(.*?)",Pattern.CASE_INSENSITIV
E | Pattern.UNICODE_CASE);
                break;
            default:
                return null;
        }
        // Evalua el patron
        m = p.matcher(noticia);
        exito = m.matches();
    }
}
```

```

    if (tipo==6){
        if (comprobarSujetosMaliciosos(patron))
            exito=false;
    }
}
result.add(exito);
if (tipo==1)
    result.add(patron);
else
    result.add("");

return result;
}

```

## G.2. PÁGINA WEB

En el siguiente apartado se van a mostrar todos los scripts realizados en pHp y javascript para la correcta visualización de las noticias. Para poder usar los script de pHp ha sido necesario instalar en Joomla! el plugin Jumi.

### G.2.1. Código para la visualización de las noticias

En este apartado muestra el código utilizado para la visualización de las noticias en una tabla dentro de un artículo de Joomla!. La forma de uso es introduciendo en el contenido del artículo la siguiente llamada:

```
{jumi [myscript.php]}
```

Esto produce una llamada al fichero *myscript.php* y ejecutar su código. En este caso, el código a ejecutar sería el siguiente:

```

<!--Carga las funciones javascript para abrir y cerrar las noticias-->
<SCRIPT TYPE="text/javascript" LANGUAGE="javascript"
SRC="fichero.js"></SCRIPT>

<?php
require_once("definicion.php"); // carga las constantes
require_once("consultas.php"); // carga las consultas
if ($link=mysql_connect(ENTORNO,USUARIO,CONTRA)) {
    // selecciona la base de datos
    $db_selected = mysql_select_db(BD, $link);
    if ($db_selected) {
        $condicion='B';
        // selecciona la consulta
        $query=sprintf(ATLETISMO,mysql_real_escape_String($condicion));
        // ejecuta la consulta
        $result=mysql_query($query);
        if ($result) {
            $i=0;
            $lColor=true; // variable para el color de fondo de la tabla
            de noticias
            require_once("traducir.php");
            // comprueba que devuelve noticias
            if (mysql_num_rows($result)>0){

```

## APÉNDICE G: Código

```
// pinta la cabecera de la tabla
printf("<TABLE BORDER=0 CELLSPACING=3 CELLPADDING=3
rules='none' style='margin: 0 auto; border-collapse:
separated; border-spacing=2px;'>");

printf("<TR><TH bgcolor='#D1E1F1'>FECHA</TH><TH
bgcolor='#D1E1F1'>TITULAR</TH></TR>");
// pinta las noticias
while($row = mysql_fetch_array($result)) {
    $row['deporte']=caracteres("ES",$row['deporte']);
    if ($lColor) {
        printf("<tr><td bgcolor='mistyrose'>%s</td><td
bgcolor='mistyrose' align='center'><h6>%s</h6><div
align='justify' id='noticia".$i."'
style='display:none'>%s</div></td><td><img
src='./images/+.png' alt='mas' title='abrir'
onclick='abrir($(".$i.")'></td><td><img src='./images/-
.png' alt='menos' title='cerrar'
onclick='cerrar($(".$i.")'></td><td><a href='%s'
target='_blank'><img src='./images/url.png' title='ir
a'></a></td></tr>",date('d-m-
Y',strtotime($row['fecha'])), $row['titular'], $row['texto'
], $row['referencia']);

        $lColor=false;
    } else {
        printf("<tr><td bgcolor='lightcyan'>%s</td><td
bgcolor='lightcyan' align='center'><h6>%s</h6><div
align='justify' id='noticia".$i."'
style='display:none'>%s</div></td><td><img
src='./images/+.png' alt='mas' title='abrir'
onclick='abrir($(".$i.")'></td><td><img src='./images/-
.png' alt='menos' title='cerrar'
onclick='cerrar($(".$i.")'></td><td><a href='%s'
target='_blank'><img src='./images/url.png' title='ir
a'></a></td></tr>",date('d-m-
Y',strtotime($row['fecha'])), $row['titular'], $row['texto'
], $row['referencia']);

        $lColor=true;
    }
    $i++;
}
printf('</TABLE></div>');
} else
printf("No existen resultados disponibles para Atletismo.");
// libera la memoria de la consulta
mysql_free_result($result);
} else {
    // Error al ejecutar la consulta
    $message=TEXT01ES.mysql_error()."<\br>";
    $message.=TEXT02ES.$query;
    echo $message;
```

```

    }
} else {
    //echo 'Error seleccionando la base de datos';
}
mysql_close($link); //cierra la conexion
} else {
    //echo 'Error conectando a la base de datos';
}
?>

```

Como se puede observar, se dibuja una tabla con una zona oculta en la etiqueta `<div>` con el contenido de la noticia; y los iconos incluyen la llamada a funciones de una librería en javascript [véase apéndice G.2.4] para visualizar u ocultar el contenido de su noticia y la posibilidad de visualizar la noticia en la página de referencia de dicha noticia.

Se puede observar en el código la existencia de dos librerías:

- ❖ **definicion.php**: esta librería contiene todas las constantes utilizadas para la conexión a la base de datos, es decir: entorno, usuario, contraseña, base de datos...
- ❖ **consultas.php**: en esta librería se define como constantes de PHP las consultas en MySQL que se van a lanzar cuando se visualiza las noticias.

## G.2.2. Código para la traducción de noticias on-line

En el siguiente apartado se muestra el código en lenguaje PHP utilizado para la traducción on-line de las noticias al idioma seleccionado por parte del usuario.

```

function traduccion ($idioma,$cadena){
    switch ($idioma){
        case "CA":
            /* se define la URL de la API de Google Translate y se
            asigna en la variable el texto a traducir */
            $url=
            'http://ajax.googleapis.com/ajax/services/language/translat
            e?v=1.0&q='.urlencode($cadena).'&langpair=es|ca';
            break;

        case "EN":
            /* se define la URL de la API de Google Translate y se
            asigna en la variable el texto a traducir */
            $url=
            'http://ajax.googleapis.com/ajax/services/language/translat
            e?v=1.0&q='.urlencode($cadena).'&langpair=es|en';
            break;
    }

    /* En caso de existir un proxy
    $loginpassw = 'login:password'; // login y password del proxy
    $proxy_ip = '127.0.0.1'; // IP del proxy
    $proxy_port = 8080; // Puerto del proxy
    */

    // Se inicia y configura curl_init();

```

## APÉNDICE G: Código

```
$curl_handle = curl_init();
curl_setopt($curl_handle,CURLOPT_URL, $url);
curl_setopt($curl_handle,CURLOPT_CONNECTTIMEOUT,2);
curl_setopt($curl_handle,CURLOPT_RETURNTRANSFER,1);

/* En caso de existir un proxy
curl_setopt($ch, CURLOPT_PROXYPORT, $proxy_port);
curl_setopt($ch, CURLOPT_PROXYTYPE, 'HTTP');
curl_setopt($ch, CURLOPT_PROXY, $proxy_ip);
curl_setopt($ch, CURLOPT_PROXYUSERPWD, $loginpasswd);
*/

$code = curl_exec($curl_handle);
curl_close($curl_handle);
// la api devuelve los resultados en forma de objeto stdClass
$json = json_decode($code);
$traduc = $json->responseData;
$cadena = $traduc->translatedText;
return $cadena;
}
```

Mientras se va dibujando la tabla con las noticias según el apéndice anterior, se llama a esta función para traducir el titular y el contenido en modo on-line. El único requisito es conocer a qué idioma es necesaria la traducción.

En la actualidad el API de Google Translate permite la traducción a más de 50 idiomas distintos.

Como se puede apreciar en el código, se ofrece la posibilidad de configurar la función ante la existencia de un proxy.

Para optimizar el uso de esta función y todo lo relacionado con la traducción de idiomas se ha creado la librería **traducir.php** que es invocada a la hora de traducir.

### G.2.3. Javascript

Según se ha visto en el apéndice G.2.1 de la visualización de las noticias, el contenido de la noticia se muestra en un <div> oculto y para visualizarlo es necesario una pequeña librería en javascript, con una función para mostrar ese contenido y otra función para volver a ocultarlo.

```
function abrir(i) {
    document.getElementById('noticia'+i).style.display="";
}

function cerrar(i) {
    document.getElementById('noticia'+i).style.display="none";
}
```

Para llamar a estas funciones el usuario deberá pulsar en las imágenes con los símbolos de sumar y restar existentes al lado de los titulares de las noticias.





**Ilustración 61 Símbolos de abrir y cerrar noticias**

Recordar que esta funcionalidad, no se encuentra operativa al 100% debido a los derechos de autor de las distintas federaciones.



## APÉNDICE H: Listado de palabras

### H.1. VERBOS

#### H.1.1. Verbos positivos

En la siguiente tabla se pueden apreciar la lista de verbos positivos.

Tabla 38 Tabla de verbos positivos

VERBOS POSITIVOS
arrolla
clasifica
conquista
consigue
derrota a
doblega
domino
dominó
gana
logra
mejora
obtiene
ocupa
pasa
premiada
premiados
proclama
repite
se lleva
suma
supera

---

**vence**

---

### H.1.2. Verbos negativos

En la siguiente tabla se pueden apreciar la lista de verbos negativos.

Tabla 39 Tabla de verbos negativos

<b>VERBOS NEGATIVOS</b>
<b>acaba con</b>
<b>caen</b>
<b>dejan</b>
<b>derrota</b>
<b>es derrotada</b>
<b>es eliminada</b>
<b>fuera</b>
<b>no logra</b>
<b>no se clasifica</b>
<b>perdió</b>
<b>perdieron</b>
<b>pierde</b>
<b>son derrotadas</b>
<b>son eliminadas</b>

### H.2. LOGROS

En la siguiente tabla se pueden apreciar la lista de posibles logros de las deportistas, los objetos directos en la semántica de los titulares.

Tabla 40 Tabla de logros

<b>LOGROS</b>	
<b>4 Naciones</b>	<b>bronce</b>
<b>campeona</b>	<b>campeonato de Europa</b>

campeonato de liga	campeonato europeo
campeonato internacional	campeones de
copa de Europa	copa del mundo
cuartos de final	europeo
final	Internacional de España
ITF femenino	líder
Ladies British	liga europea de naciones
Londres 2012	medalla
medallista	mejor jugadora
mejora marca	mejor posición
mejor puesto	mejores jugadores
octavos de final	oro
plata	premio
récord	record
semifinal	Shield of the Town Cheb
subcampeona	título
victoria	