

Título: Evaluación del sistema de clasificación automática de contenidos *Resync* en medios de comunicación españoles y mexicanos

Resumen: El objetivo de la investigación es la evaluación de los algoritmos para la clasificación automática de contenidos, diseñados originalmente para la categorización temática de los contenidos y noticias recopiladas con la plataforma *Resync* (BLÁZQUEZ OCHANDO, M. 2012). El proceso de evaluación se lleva a cabo con formularios específicamente diseñados para determinar el grado de relevancia de los contenidos clasificados por los evaluadores. Finalmente se determina que uno de los algoritmos utilizados posee un porcentaje de precisión del 71%. También se determina que las temáticas mejor clasificadas son aquellas relativas a finanzas, derecho y política.

Palabras clave: sistemas de clasificación automática, algoritmos de clasificación, evaluación de algoritmos, recuperación de información

Abstract: The objective of the research is to evaluate algorithms for automatic classification of content, originally designed for thematic categorization of content and news compiled with RESYNC platform (BLÁZQUEZ OCHANDO, M. 2012). The evaluation process is conducted with forms specifically designed to determine the degree of relevance of content classified by the evaluators. Finally, it shows that the algorithm evaluated, has an accuracy of 71%. It was also determined that the highest ranked topics are those related to finance, law and politics.

Keywords: automatic classification systems, classification algorithms, evaluation of algorithms, information retrieval

Reference: BLÁZQUEZ OCHANDO, M. 2013. [Ponencia]. Desarrollo de un sistema de clasificación automática de contenidos en medios de comunicación españoles y mexicanos. En: 10º Seminario Hispano-Mexicano de Bibliotecología y Documentación (Madrid, 22-24 abril).

Introducción

El trabajo expuesto, forma parte de la investigación iniciada en torno al título *Desarrollo de un sistema de clasificación automática de contenidos en medios de comunicación hispano-mexicanos* presentada en el marco del IX Seminario Hispano Mexicano de Biblioteconomía y Documentación, véase (Blázquez Ochando, M. 2012). Inicialmente el objetivo fue el desarrollo de un sistema de clasificación automática para los contenidos recuperados a través de la plataforma experimental de sindicación *Resync*, también presentada en investigaciones anteriores, véase (Blázquez Ochando, M.; Serrano Mascaraque, E. 2011). Ello fue motivado por la gran cantidad de información publicada por los medios de comunicación de España y México y que consecuentemente fue recuperada durante un mes de pruebas. Con un total de 400.000 noticias obtenidas se hizo necesario crear diversos métodos capaces de ordenar los contenidos en función de su categoría temática. El resultado de dicho trabajo conllevó el diseño, desarrollo y prueba de 5 algoritmos de clasificación con los que se obtuvieron distintos porcentajes de clasificación de la colección que variaban desde el 1,81% hasta el 99,15% dependiendo del grado de precisión y exigencia de las condiciones programadas para cada algoritmo. Si bien los datos obtenidos muestran la capacidad para clasificar los contenidos recuperados en los medios de comunicación, no proporcionaban una información cualitativa que determinara los fallos en el proceso de clasificación, dicho de otra forma, cuántos contenidos y noticias fueron correctamente clasificados y con cuántos falló. De esta forma se vislumbró la posibilidad de realizar un proceso de evaluación que identificase el algoritmo que mejor funcionó en cuanto al índice de aciertos.

La evaluación de los sistemas de clasificación automática y por ende de sus algoritmos, empleados comúnmente en la recuperación de información, constituyen uno de los principios básicos por los que es posible perfeccionar, refinar y ajustar la metodología y el cálculo del proceso clasificatorio. Como es bien sabido, la evaluación de los resultados de un sistema de recuperación o clasificación, puede ser medida por medio de tres factores. La exhaustividad, la relevancia del sistema y la relevancia del usuario. Este último factor introduce una variable no controlada, que indica en realidad si los resultados cuantitativos se ajustan a las necesidades informativas del usuario o bien si los resultados fueron correctamente clasificados conforme a una categoría temática determinada. A partir de este punto, la presente investigación retoma el trabajo de evaluar los algoritmos de clasificación utilizados, empleando una serie de instrumentos especialmente diseñados para tal propósito.

Metodología

Hay que diferenciar dos partes en la metodología llevada a cabo para la investigación. En primer lugar la parte de preparación previa que corresponde a la formación de la colección que será clasificada, los algoritmos utilizados para realizar el proceso de clasificación y el método de análisis o instrumentos que hacen posible que los usuarios evalúen los resultados de clasificación del algoritmo. En segundo lugar, la parte relativa al procedimiento de evaluación propiamente dicho, parte en la que se hará mayor hincapié.

Preparación previa

1) Colección de contenidos clasificados

El corpus documental original consta de 431.592 noticias publicadas por los medios españoles y mexicanos. Del corpus total se ha tomado una muestra de 16.627 noticias correspondientes a la totalidad de las categorías temáticas relativas a *vida política, relaciones internacionales, derecho, vida económica, intercambios económicos y comerciales y asuntos financieros* especificadas en el tesoro Eurovoc. (Blázquez Ochando, M. 2012: 5)

2) Algoritmo de clasificación evaluado

Corresponde al algoritmo de clasificación temática por precisión, denominado en la investigación precedente como *algoritmo prueba1*. El algoritmo consulta en lenguaje natural, los textos indexados en la colección de prueba, aplicando la intersección booleana a una consulta conformada automáticamente a partir de la terminología del tesoro Eurovoc. Las condiciones del algoritmo establecen que cada consulta no puede utilizar términos con más de 7 caracteres y que sólo se clasificarán aquellos contenidos y documentos cuyo coeficiente de similitud sea igual o superior a 10. Véase descripción completa del algoritmo en: (Blázquez Ochando, M. 2012: 7)

3) Formularios de evaluación

A partir de los resultados obtenidos con el algoritmo de clasificación *prueba1*, se crean automáticamente una serie de formularios de evaluación que corresponden a la última fase de la investigación. Estos informes tabulados muestran el coeficiente de relevancia para el sistema, la fecha de registro, el título, enlace, resumen y texto completo de cada noticia clasificada bajo cada categoría temática (Blázquez Ochando, M. 2012: 9). Véase *figura 1*

VIDA POLÍTICA >> Marco político

ID	Título	Fecha	Relevancia
74.285580788068	Pedro de Tena: España no es una democracia	2011-07-18 10:22:56 +02:00	80%
60.003179241195	La democracia es una máquina calculadora	2011-06-26 04:09:54 +02:00	80%
56.099986577838	Maneiro (UPyD) pide acabar con el bipartidismo y con el "conservadurismo de los apoltronados"	2011-07-31 17:22:13 +02:00	80%
56.099986577838	Maneiro (UPyD) pide acabar con el bipartidismo y con el "conservadurismo de los apoltronados"	2011-07-31 11:22:15 +02:00	80%
54.871240841042	Democracia y justa indignación	2011-07-24 17:22:28 +02:00	80%
54.624750807571	Los partidos mayoritarios apuestan por "fortalecer" la democracia y los minoritarios critican el bipartidismo	2011-07-12 06:18:25 +02:00	80%
52.45804944379	Ali Lmrabet lamenta que "las revueltas árabes por la democracia no reciben ningún apoyo de Occidente"		

Figura 1. Formulario de evaluación para los contenidos clasificados en la categoría temática de "Marco Político". Obsérvense los controles de evaluación

Procedimiento de evaluación

1) *Sujetos evaluadores*

Las personas encargadas de la evaluación de los contenidos clasificados son los alumnos de las asignaturas de *Evaluación de sistemas de información* (2011-2012) y *Técnicas avanzadas de recuperación de información* (2012-2013) de la Facultad de Ciencias de la Documentación de la Universidad Complutense. A modo de práctica se instó a los alumnos a probar un sistema de evaluación de un motor de búsqueda y clasificación real, que en efecto corresponde al algoritmo *prueba1* citado anteriormente. A los sujetos evaluadores se les dio absoluta libertad para determinar el grado de relevancia de los resultados con respecto a la categoría temática que les fue asignada en su labor. Por otra parte es importante destacar el factor externo de la evaluación. Esto es que al ser formularios en línea disponibles en un servidor abierto, son de libre acceso para cualquier usuario que desee analizarlos, por lo que cabe la posibilidad de que existan más sujetos evaluadores, no controlados. Si bien ello puede suponer un problema en la normalización de los resultados, también es cierto que pueden proporcionar cierto grado de realidad a la prueba, puesto que son los usuarios finales quienes determinan realmente la calidad del proceso clasificatorio.

2) *Asignación de categorías temáticas*

Dado el importante tamaño de la muestra que se pretende analizar, a cada alumno se le asignaron formularios de evaluación correspondientes a una serie de categorías temáticas, variables en función del tamaño. Se pidió que al menos se evaluaran 100 contenidos por cada categoría temática, a fin de obtener una muestra evaluativa suficientemente amplia. En muchos casos, esta cifra fue superada y existen múltiples valoraciones para un mismo contenido. Los índices de asignación y los formularios de evaluación se encuentran disponibles en las siguientes direcciones:

<http://mblazquez.es/testbench/evaluacion/prueba1-es/>

<http://mblazquez.es/testbench/evaluacion/prueba1-mx/>

3) *Mecanismo de evaluación del formulario*

El formulario de evaluación no sólo muestra la información de los contenidos y noticias, sino que proporciona también diversos botones de control para evaluar la relevancia de las mismas con respecto a la temática el que fueron clasificados. El botón “Marcar relevante” permite al evaluador determinar que la relevancia del contenido con respecto a la temática en la que fue clasificado es del 100%. El botón “Marcar irrelevante” determina una relevancia del 0%, por lo tanto el contenido es opuesto a la categoría temática en la que fue clasificado. Y finalmente dado que la relevancia no siempre es binaria, se establecieron una serie de opciones que permiten al evaluador determinar distintos grados de relevancia, del 80%, 60%, 40% y 20%. Cuando el evaluador hace clic en alguno de estos botones, de forma inmediata se remite un mensaje en el que se indica que se registró correctamente la evaluación en la base de datos de evaluaciones creada ex profeso para el caso.

4) *Recopilación de datos*

El inicio de la recopilación de datos se produce a partir del 30 de abril de 2012, finalizando el 29 de marzo de 2013. Durante este tiempo las evaluaciones

efectuadas han sido recogidas por la base de datos, en la que se han contabilizado 10.471 valoraciones sobre los contenidos clasificados por el algoritmo *prueba1*. Véase tabla1 relativa a la distribución del número total de evaluaciones por años. Cabe destacar que el año en el que más valoraciones se realizaron corresponde al 2012, llegando a triplicar la participación de los evaluadores del periodo siguiente.

Año	Nº total de evaluaciones
2012	7.508
2013	2.458
Total	10.471

Tabla 1. Número total de evaluaciones por años

5) *Interpretación de las evaluaciones: cálculo de precisión*

El cálculo se realiza teniendo en cuenta que son en mayor o menor medida relevantes para el usuario, todos aquellos contenidos evaluados desde el 40% hasta el 100% de correspondencia del contenido o noticia con respecto a la temática en la que fue clasificada, tal como se realiza e indica en todos los estudios relacionados como por ejemplo (Cleverdon, C.W.; Keen, E.M. 1966), (Rijsbergen, C.J. 1979), (López Herrera, A.G. 2006: 43-48), (Aparicio Escribano, D. 2009: 31) y (Lin, S.H.; Shihh, C.S.; Chang Chen, M. 2010: 5) . La suma de todas las evaluaciones que cumplen este criterio da lugar al *factor de relevancia global*. Todas las evaluaciones cuyo porcentaje varía del 40% al 0% corresponden a contenidos cuya temática no fue correctamente interpretada por el algoritmo. Por tanto la suma de las evaluaciones que cumplan tal criterio, darán lugar al *factor de irrelevancia global*. Con todos los datos disponibles, se realiza el cálculo del porcentaje de precisión del algoritmo *prueba1* que es el conjunto de noticias o contenidos que fueron clasificados correctamente de forma relevante para el usuario dividido entre el número total de noticias clasificadas por el sistema. De esta forma se obtiene un porcentaje aproximado del éxito o precisión del algoritmo de clasificación, véase *tabla2*.

$Precisión = \frac{tp_i}{tp_i + fp_i}$
<ul style="list-style-type: none"> - tp_i . Es el número total de documentos que han sido correctamente clasificados en la categoría <i>i</i>. Todos aquellos documentos valorados del 40% al 100% de relevancia por el usuario. - fp_i . Es el número total de documentos cuya clasificación fue errónea. Todos aquellos documentos valorados del 40% al 0% de relevancia por el usuario. - $N = tp_i + fp_i$. La suma de los documentos bien y mal clasificados en una determinada categoría temática es igual a <i>N</i> que representa el número total de documentos evaluados en dicho área temática.

Tabla 2. Método de evaluación de la precisión del sistema de clasificación automática

Resultados

La evaluación llevada a cabo sobre una muestra de 16.627 noticias y contenidos publicados en medios de comunicación españoles y mexicanos, tuvo un total de 10.471 evaluaciones lo que supone que la muestra ha sido evaluada en un 62,98%. Las categorías temáticas más evaluadas corresponden al *análisis económico* con un 3.125% puesto que tan sólo existían 8 contenidos clasificados que fueron 250 veces evaluados. Esto ha provocado que la media estadística del 253,76% no represente adecuadamente el total de noticias evaluadas. Por otra parte existen categorías temáticas no evaluadas como *política económica* y *consumo*. Aunque existen porcentajes de evaluación muy variables, la mediana estadística se sitúa en el 48,21% lo que indica, que más de la mitad de todas las noticias, fue evaluada al menos una vez. Véase *tabla2*.

Temática	Categoría	Nº total de noticias	Nº de evaluaciones	Porcentaje de noticias evaluadas
Vida política	Marco político	538	206	38,29%
	Partido político	79	119	150,63%
	Procedimiento electoral y sistema de votación	1.973	1.388	70,35%
	Parlamento	856	849	99,18%
	Trabajos parlamentarios	47	15	31,91%
	Vida política y seguridad pública	618	499	80,74%
	Poder ejecutivo y administración pública	4.677	3.797	81,18%
Relaciones internacionales	Política internacional	144	59	40,97%
	Política de cooperación	67	71	105,97%
	Equilibrio internacional	569	159	27,94%
	Defensa	156	66	42,31%
Derecho	Fuentes y ramas del Derecho	35	613	1.751,43%
	Derecho civil	643	310	48,21%
	Derecho penal	344	257	74,71%
	Justicia	377	11	2,92%
	Organización de la justicia	113	358	316,81%
	Derecho internacional	674	236	35,01%
	Derechos y libertades	535	55	10,28%
Vida económica	Política económica	435	0	0,00%
	Crecimiento económico	57	57	100,00%
	Región y política regional	30	37	123,33%
	Estructura económica	17	82	482,35%
	Contabilidad nacional	72	10	13,89%
	Análisis económico	8	250	3.125,00%
Intercambios económicos y comerciales	Política comercial	724	100	13,81%
	Política arancelaria	94	291	309,57%
	Intercambios económicos	707	28	3,96%
	Comercio internacional	31	97	312,90%

	Consumo	184	0	0,00%
	Comercialización	12	90	750,00%
	Distribución y venta	955	119	12,46%
Asuntos financieros	Relaciones monetarias	329	72	21,88%
	Economía monetaria	128	54	42,19%
	Instituciones financieras y de crédito	334	51	15,27%
	Libre circulación de capitales	55	2	3,64%
	Financiación e inversión	2	7	350,00%
	Seguros	8	56	700,00%
Recuento		16.627	10.471	62,98%

Tabla 3. Porcentaje de noticias evaluadas

En relación a la precisión declarada en las evaluaciones, se puede comprobar que en el año 2012, los evaluadores dieron un 16 % más de relevancia a los contenidos clasificados, que en el año 2013 con un 60% de relevancia, véase *tabla3*.

Combinados los datos de ambos años se obtiene, que la precisión general del algoritmo alcanza el 71% muy cerca de los resultados obtenidos en otros estudios como en el de (Ardo, A. 2007) en el que se asegura que el empleo de todos los términos del corpus clasificatorio y sus derivados, permiten obtener un promedio de en la precisión del 73%.

Años	100%	80%	60%	40%	20%	0%	Precisión
2012	3.688	836	603	608	529	1.244	76,385%
2013	864	246	183	182	149	834	60,008%
Total	4.719	1.095	811	833	709	2.304	71,225%

Tabla 4. Precisión general de la clasificación automática según las evaluaciones efectuadas

Las categorías temáticas mejor clasificadas con un importante nivel de acuerdo entre los evaluadores, corresponden a los contenidos relativos al *poder ejecutivo* y la *administración pública, derecho penal y parlamento*. En todos los casos se supera el 79% de precisión.

Pos.	Temática	Categoría	100%	80%	60%	40%	20%	0%	Precisión
1	Asuntos financieros	Libre circulación de capitales	2	0	0	0	0	0	100%
2	Vida política	Poder ejecutivo y administración pública	2.810	572	130	80	61	144	94,601%
3	Asuntos financieros	Economía monetaria	39	8	2	1	1	3	92,593%
4	Asuntos financieros	Relaciones monetarias	23	12	23	5	2	7	87,5%
5	Derecho	Derecho penal	137	71	8	5	1	35	85,992%
6	Vida política	Parlamento	254	114	132	175	81	93	79,505%
7	Vida económica	Crecimiento económico	15	8	13	7	1	13	75,439%

8	Asuntos financieros	Instituciones financieras y de crédito	35	2	1	0	0	13	74,51%
9	Vida política	Trabajos parlamentarios	11	0	0	0	0	4	73,333%
10	Vida económica	Contabilidad nacional	7	0	0	0	0	3	70%

Tabla 5. Categorías temáticas mejor clasificadas

Los contenidos peor clasificados fueron los correspondientes a las categorías temáticas de *justicia y análisis económico*, con una precisión cercana al 30% y más de 100 evaluaciones realizadas. Aún así, también es relevante observar que sólo 7 categorías temáticas quedan por debajo del 40% de precisión. Esto indica que las restantes 30 categorías de la muestra tienen valores superiores.

Pos.	Temática	Categoría	100%	80%	60%	40%	20%	0%	Precisión
1	Intercambios económicos y comerciales	Intercambios económicos	7	0	0	0	13	8	25%
2	Intercambios económicos y comerciales	Comercialización	14	4	6	1	1	64	27,778%
3	Relaciones internacionales	Política de cooperación	3	2	7	8	1	50	28,169%
4	Derecho	Organización de la justicia	53	15	14	22	62	192	29,05%
5	Intercambios económicos y comerciales	Política comercial	31	0	0	0	0	69	31%
6	Vida económica	Análisis económico	16	26	13	29	57	109	33,6%
7	Intercambios económicos y comerciales	Comercio internacional	27	8	1	2	6	53	39,175%
8	Vida económica	Estructura económica	30	0	1	3	4	44	41,463%
9	Asuntos financieros	Seguros	3	0	0	0	0	4	42,857%
10	Vida política	Región y política regional	11	0	3	3	7	13	45,946%

Tabla 6. Categorías temáticas peor clasificadas

En la *tabla 6* se muestran los resultados del año 2012, en los que cabe destacar que la precisión del algoritmo *prueba1*, se sitúa en un nivel medio alto. Las categorías temáticas con más de un centenar de evaluaciones con resultados entre el 70% y el 80% de precisión son *parlamento, vida política y seguridad pública, derecho civil y fuentes y ramas del derecho*.

Con más del 90% de precisión y más de 3.000 evaluaciones se encuentra la temática *Poder ejecutivo y administración pública*. Con valores medios de precisión entre el 40% y el 50% se encuentran las categorías de *análisis económico, política arancelaria, procedimiento electoral y sistemas de votación*. En conclusión se puede afirmar que aproximadamente dos tercios de los contenidos fueron clasificados correctamente, según el equipo evaluador.

2012								
Temática	Categoría	100%	80%	60%	40%	20%	0%	Precisión
Vida política	Marco político	7	3	0	0	0	0	100,00%
	Procedimiento electoral y sistema de votación	15	46	283	168	248	404	43,99%
	Parlamento	170	94	114	157	53	80	80,09%
	Trabajos parlamentarios	11	0	0	0	0	4	73,33%
	Vida política y seguridad pública	173	15	28	13	0	84	73,16%
	Poder ejecutivo y administración pública	2.666	505	92	40	23	63	97,46%
Relaciones internacionales	Política de cooperación	2	2	6	6	0	26	38,10%
	Equilibrio internacional	9	13	2	4	18	5	54,90%
	Defensa	14	1	4	10	8	8	64,44%
Derecho	Fuentes y ramas del Derecho	186	7	30	156	48	112	70,32%
	Derecho civil	62	18	0	0	0	23	77,67%
	Derecho penal	124	63	0	0	0	8	95,90%
	Justicia	4	0	1	2	0	4	63,64%
	Organización de la justicia	5	15	14	22	62	180	18,79%
	Derecho internacional	45	23	6	2	21	1	77,55%
	Derechos y libertades	10	0	0	0	0	0	100,00%
Vida económica	Crecimiento económico	6	4	11	1	0	0	100,00%
	Región y política regional	5	0	3	2	7	3	50,00%
	Estructura económica	0	0	0	2	3	2	28,57%
	Contabilidad nacional	2	0	0	0	0	0	100,00%
	Análisis económico	10	25	9	21	19	44	50,78%
Intercambios económicos y comerciales	Política comercial	31	0	0	0	0	18	63,27%
	Política arancelaria	99	2	0	0	0	114	46,98%
	Intercambios económicos	7	0	0	0	13	8	25,00%
	Comercio internacional	25	0	0	2	6	53	31,40%

Tabla 7. Evaluaciones y recuentos del año 2012

En la *tabla7*, se muestran los datos correspondientes al año 2013, en los que puede observarse un descenso en el número de evaluaciones. Aunque los valores globales del año dan al algoritmo *prueba1* una precisión del 60%, sólo 10 de las 30 categorías disponibles superan la centena de evaluaciones, de las que 3 obtuvieron precisión por encima del 70% siendo *marco político*, *parlamento*, *poder ejecutivo* y *administración pública*.

Entre el 50 y el 60% se contabilizan 6 categorías a saber; *partido político*, *vida política* y *seguridad pública*, *equilibrio internacional*, *derecho civil*, *derecho internacional*, *distribución* y *venta*. Finalmente queda la categoría de *análisis económico* con menos del 20% de precisión.

2013								
Temática	Categoría	100%	80%	60%	40%	20%	0%	Precisión
Vida política	Marco político	62	15	8	7	5	29	73,02%
	Partido político	43	13	3	19	8	33	65,55%
	Procedimiento electoral y sistema de votación	26	12	8	10	6	9	78,87%
	Parlamento	84	20	18	18	28	13	77,35%
	Vida política y seguridad pública	81	0	13	10	3	79	55,91%
	Poder ejecutivo y administración pública	108	62	20	7	12	29	82,773%
Relaciones internacionales	Política internacional	20	5	3	1	1	29	49,15%
	Política de cooperación	1	0	1	2	1	24	13,79%
	Equilibrio internacional	24	17	3	21	11	32	60,19%
	Defensa	2	0	0	0	0	19	9,52%
Derecho	Fuentes y ramas del Derecho	23	2	7	6	5	31	51,35%
	Derecho civil	69	20	16	11	5	86	56,04%
	Derecho penal	13	8	8	5	1	27	54,84%
	Organización de la justicia	48	0	0	0	0	12	80,00%
	Derecho internacional	50	1	6	15	4	62	52,17%
	Derechos y libertades	3	3	0	0	0	9	40,00%
Vida económica	Crecimiento económico	5	4	2	6	1	8	65,38%
	Región y política regional	5	0	0	1	0	9	40,00%
	Estructura económica	0	0	1	1	1	7	20,00%
	Contabilidad nacional	2	0	0	0	0	0	100,00%
	Análisis económico	6	1	4	8	38	65	15,57%
Intercambios económicos y comerciales	Política comercial	0	0	0	0	0	51	0,00%
	Política arancelaria	32	7	3	5	5	24	61,84%
	Comercio internacional	2	8	1	0	0	0	100,00%
	Comercialización	14	4	6	1	1	64	27,78%
	Distribución y venta	24	7	18	21	10	39	58,82%
Asuntos financieros	Relaciones monetarias	23	12	23	5	2	7	87,50%
	Economía monetaria	39	8	2	1	1	3	92,59%
	Instituciones financieras y de crédito	35	2	1	0	0	13	74,51%
	Libre circulación de capitales	2	0	0	0	0	0	100,00%
	Financiación e inversión	3	0	0	0	0	4	42,86%
	Seguros	15	15	8	1	0	17	69,64%

Tabla 8. Evaluaciones y recuentos del año 2013

Si se comparan los resultados obtenidos en el año 2012 y 2013, puede observarse que aun existiendo diferencias en la precisión obtenida, la línea de tendencia para las temáticas evaluadas sigue un patrón muy similar, véanse figuras 2 y 3.

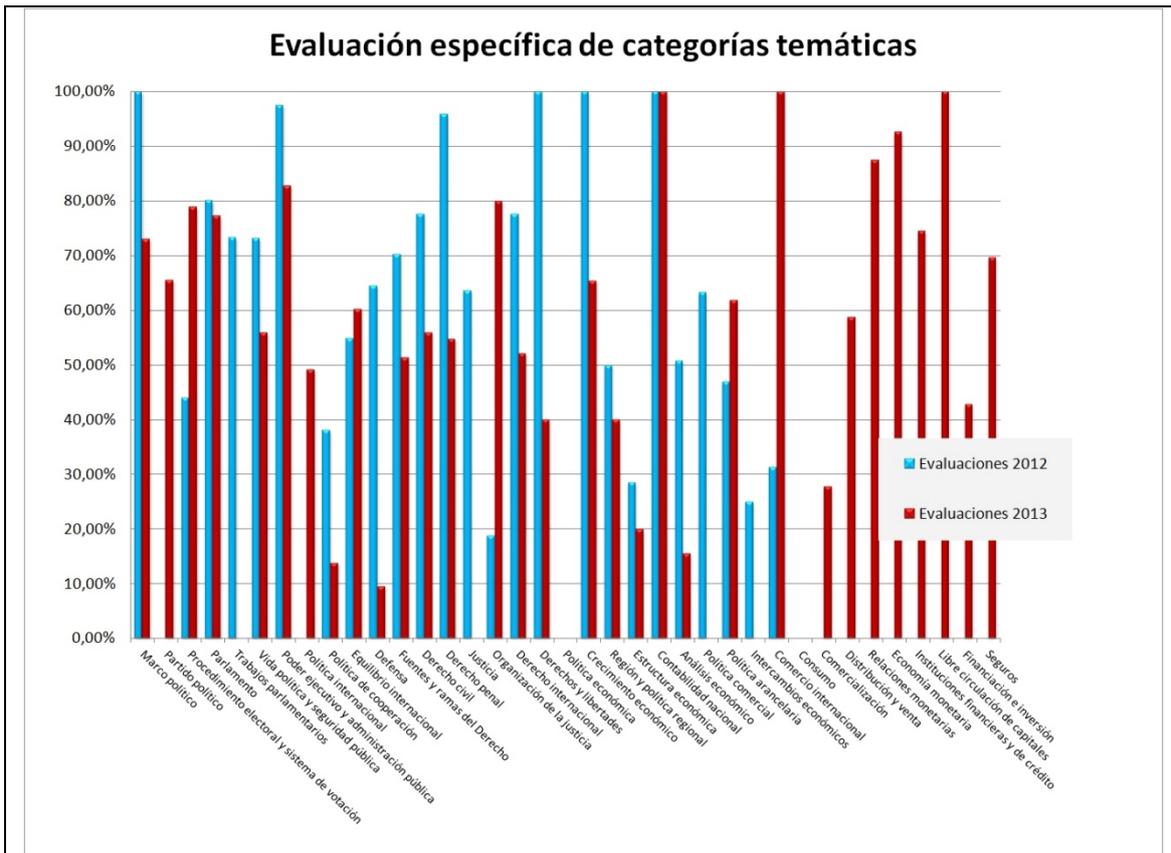


Figura 2. Precisión para cada categoría temática

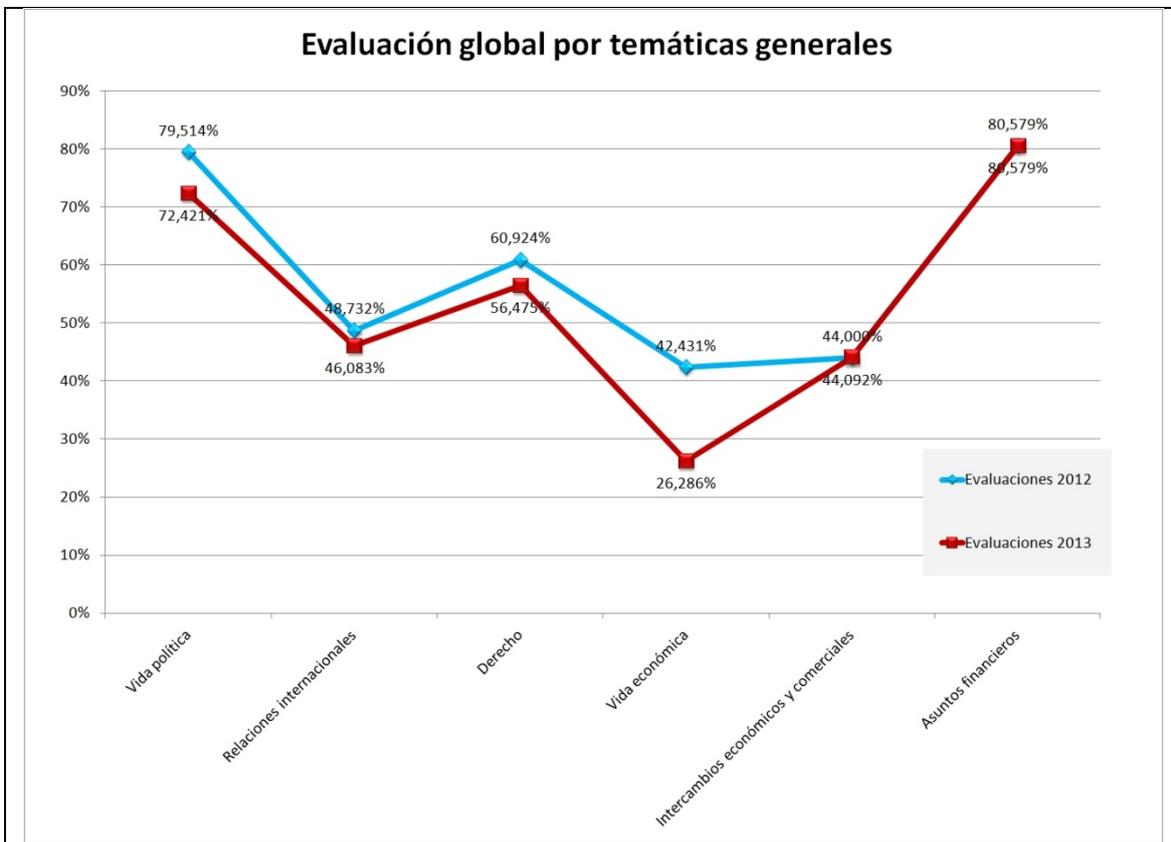


Figura 3. Precisión global por temáticas

Conclusiones

1. A tenor de los resultados obtenidos, el algoritmo *pruebal* empleado para la clasificación de noticias y contenidos sindicados de los medios de comunicación de España y México, tiene una precisión del 71% cuando se emplea el vocabulario del tesoro multilingüe europeo Eurovoc.
2. La categoría temática más evaluada es *poder ejecutivo y administración pública* con 3.546 valoraciones, obtiene el máximo nivel de precisión del 94%, por ende es en la que mejor se clasifican los contenidos.
3. La temática relativa a *vida económica e intercambios económicos* se clasifican peor con precisiones del 30% al 45%, pero no tienen un número de evaluaciones tan alto como el resto de áreas temáticas de la prueba. Por lo que no se puede asegurar todavía un valor definitivo hasta que no se amplíe el número total de valoraciones.
4. Las áreas temáticas mejor clasificadas corresponden a vida política, derecho y asuntos financieros con precisiones superiores al 60%, 70% y 80% en la mayor parte de los casos.
5. Los evaluadores del año 2012 a tribuyeron resultados más relevantes que los obtenidos por los evaluadores del año 2013. A pesar de todo, la gráfica comparativa de precisión global por temáticas, muestra tendencias muy similares. Ello indica un importante nivel de consenso en las valoraciones realizadas.
6. Aunque la muestra original evaluada está constituida por 16.000 contenidos y 37 categorías temáticas, quedan todavía por valorar 86. E llo permitiría obtener resultados más completos para determinar la precisión de clasificación en otras áreas temáticas y ofrecer un dato de precisión global del algoritmo mucho más exacto.

Referencias

- APARICIO ESCRIBANO, D. (2009). Clasificación automática de vídeos. Universidad Carlos III. Disponible en: http://e-archivo.uc3m.es/bitstream/10016/8490/1/PFC_David_Aparicio_Escribano.pdf
- ARDO, A. (2007). Evaluation of automated subject classification. Disponible en: <http://combine.it.lth.se/documentation/DocMain/node7.html>
- BLÁZQUEZ OCHANDO, M.; SERRANO MASCARAQUE, E. (2011). Plataforma para la investigación de contenidos sindicados: desarrollo del sistema ReSync y aplicación a los medios de comunicación españoles y mexicanos. En: 8º Seminario Hispano-Mexicano de Bibliotecología y Documentación (México, 21-25 marzo). Disponible en: <http://eprints.rclis.org/19030/>
- BLÁZQUEZ OCHANDO, M. (2012). Desarrollo de un sistema de clasificación automática de contenidos en medios de comunicación españoles y mexicanos.

En: 9º Seminario Hispano-Mexicano de Bibliotecología y Documentación (México, 7-9 mayo). Disponible en: <http://eprints.rclis.org/19031/>

- CLEVERDON, C.W.; KEEN, E.M. (1966). Factors determining the performance of indexing systems. En: Technical report, College of Aeronautics, Cranfield.
- LANCASTER, F.W. (1979). Information retrieval systems – characteristics, testing and evaluation. En: Criteria by Which Information Retrieval Systems May Be Evaluated. Willey.
- LIN, S.H.; SHIH, C.S.; CHANG CHEN, M. (2010). ACIRD: An Intelligent Internet Information System Based on Data Mining. Disponible en: <http://www09.sigmod.org/disc/disc99/disc/dmkd/acird.pdf>
- LÓPEZ HERRERA, A.G. (2006). Modelos de sistemas de recuperación de información documental basados en información lingüística difusa. Universidad de Granada. Disponible: <http://hera.ugr.es/tesisugr/15894046.pdf>
- SÁNCHEZ JIMÉNEZ, R. (2007). La documentación en el proceso de evaluación de Sistemas de Clasificación Automática. En: Documentación de las Ciencias de la Información. Vol. 30, Disponible en: <http://revistas.ucm.es/index.php/DCIN/article/view/DCIN0707110025A/18959>