



Tesis: “Análisis del algoritmo ActiveRank como método de detección automática de contenido dentro de redes de información”

Fernando Luege Mateos

México D.F., Febrero 2010



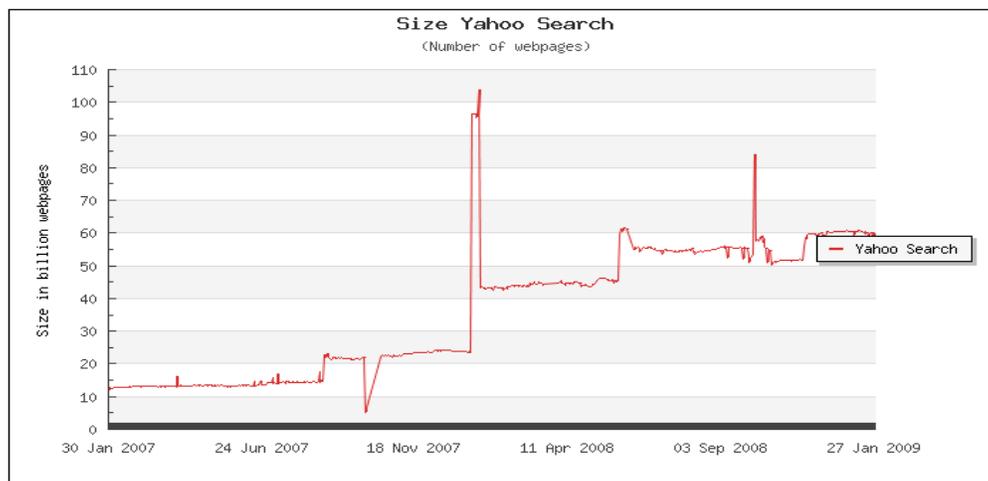
Apéndices



Apéndice A

Dimensiones de la World Wide Web

Las diferentes gráficas que a continuación se muestran han sido obtenidas del sitio web <http://www.worldwidewebsize.com>; el contenido ha sido seleccionado con la intención de introducir únicamente información necesaria para la comprensión de las ideas correspondientes a las dimensiones de la WWW.



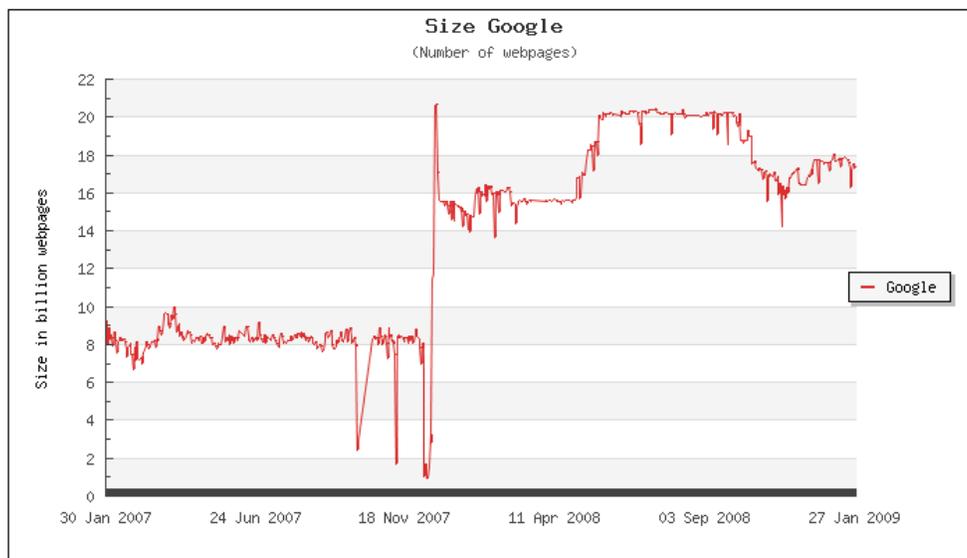
Gráfica A.1. – Número de páginas indexadas por Yahoo a Enero de 2009



Tesis: “Análisis del algoritmo ActiveRank como método de detección automática de contenido dentro de redes de información”

Fernando Luege Mateos

México D.F., Febrero 2010



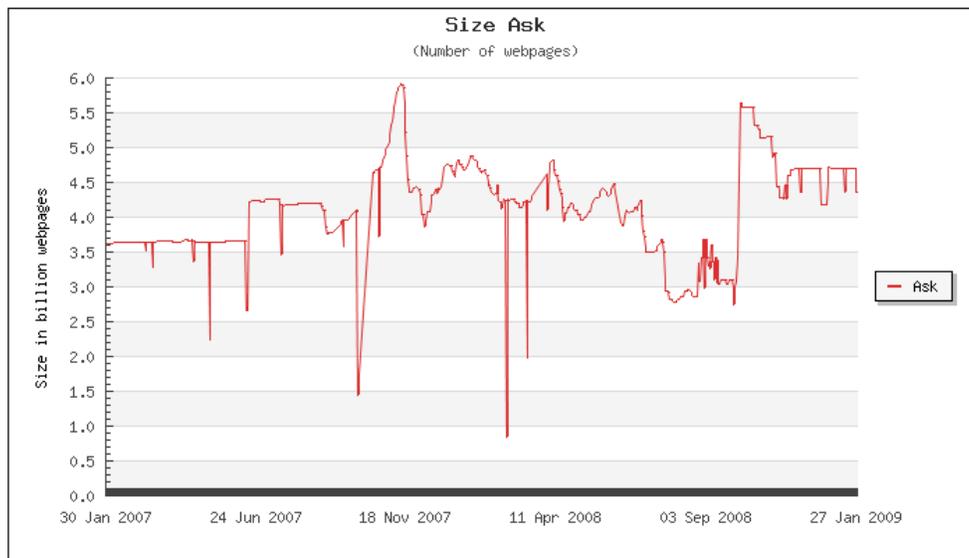
Gráfica A.2. – Número de páginas indexadas por Google a Enero de 2009



Tesis: “Análisis del algoritmo ActiveRank como método de detección automática de contenido dentro de redes de información”

Fernando Luege Mateos

México D.F., Febrero 2010



Gráfica A.3. – Número de páginas indexadas por Ask a Enero de 2009



Apéndice B

Diagrama de flujo de un sistema básico de análisis de la WWW

A continuación se describe de manera general el proceso de análisis de información mediante un diagrama de flujo representado en las figuras B.1. a B.3.; los bloques en azul representan aquellas actividades básicas para poder explorar la WWW, mientras que aquellas en rojo deben ser interpretadas como los procesos llevados a cabo para cumplir un fin específico, en este caso, un análisis semántico de la información textual.

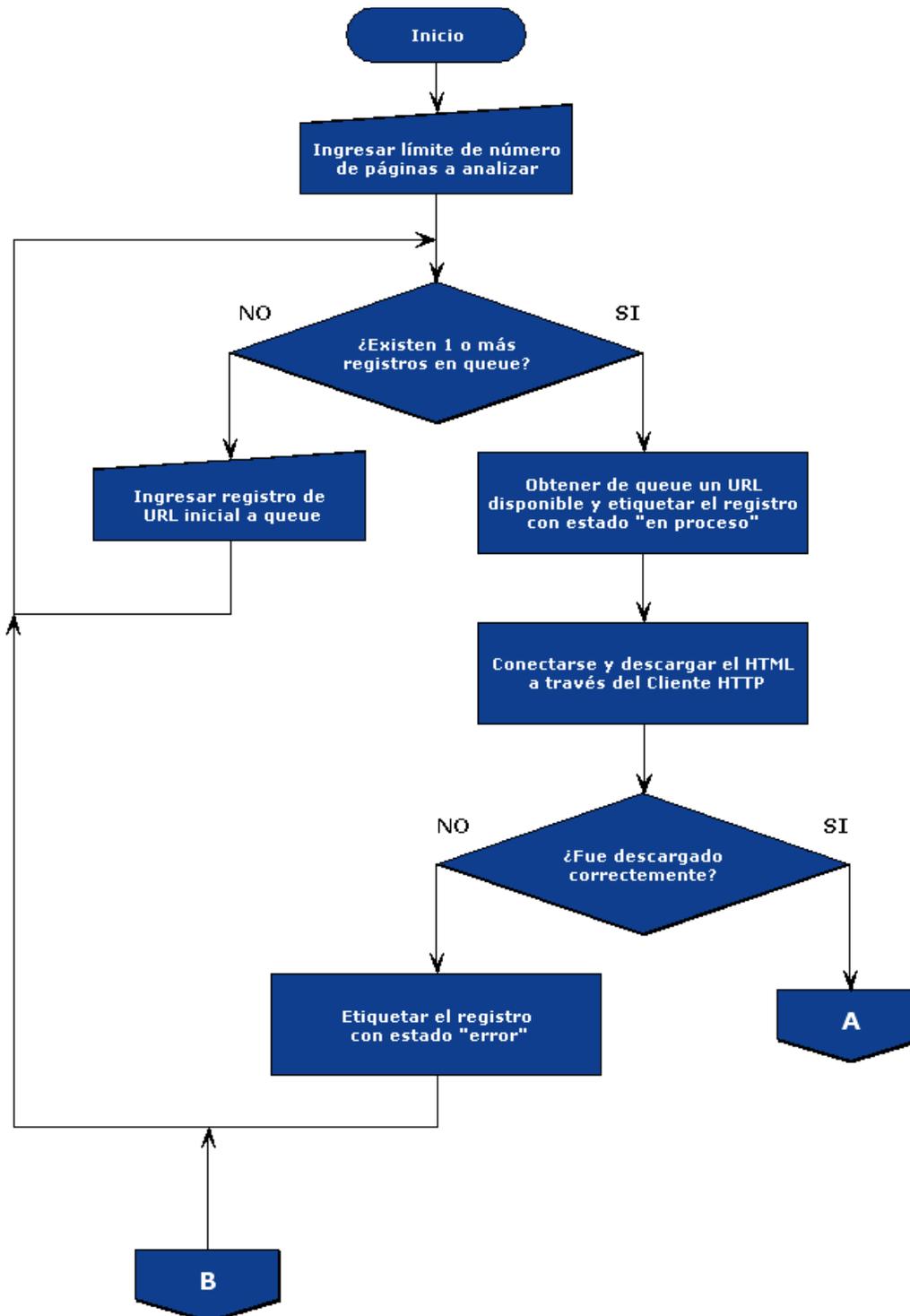


Figura B.1. – Diagrama de flujo de un crawler básico

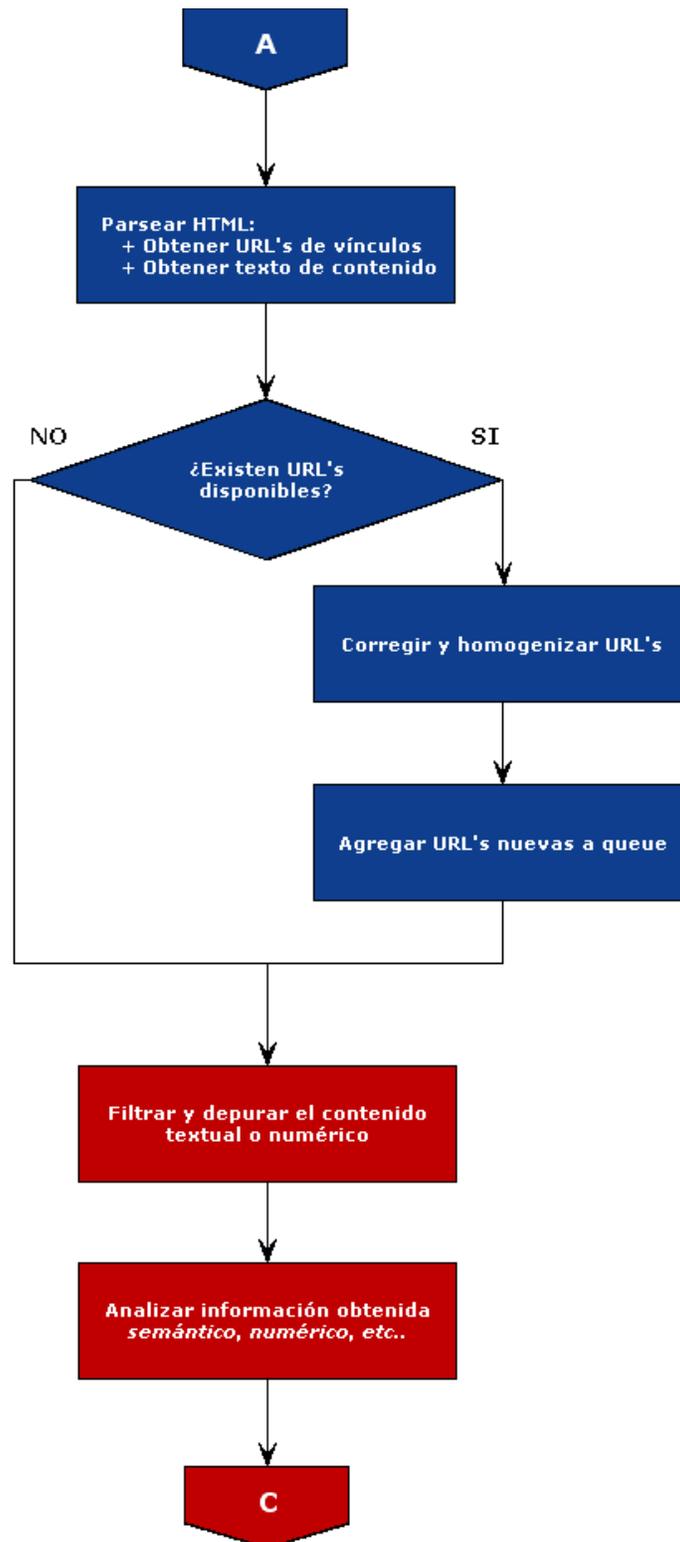


Figura B.2. – Diagrama de flujo de un crawler básico

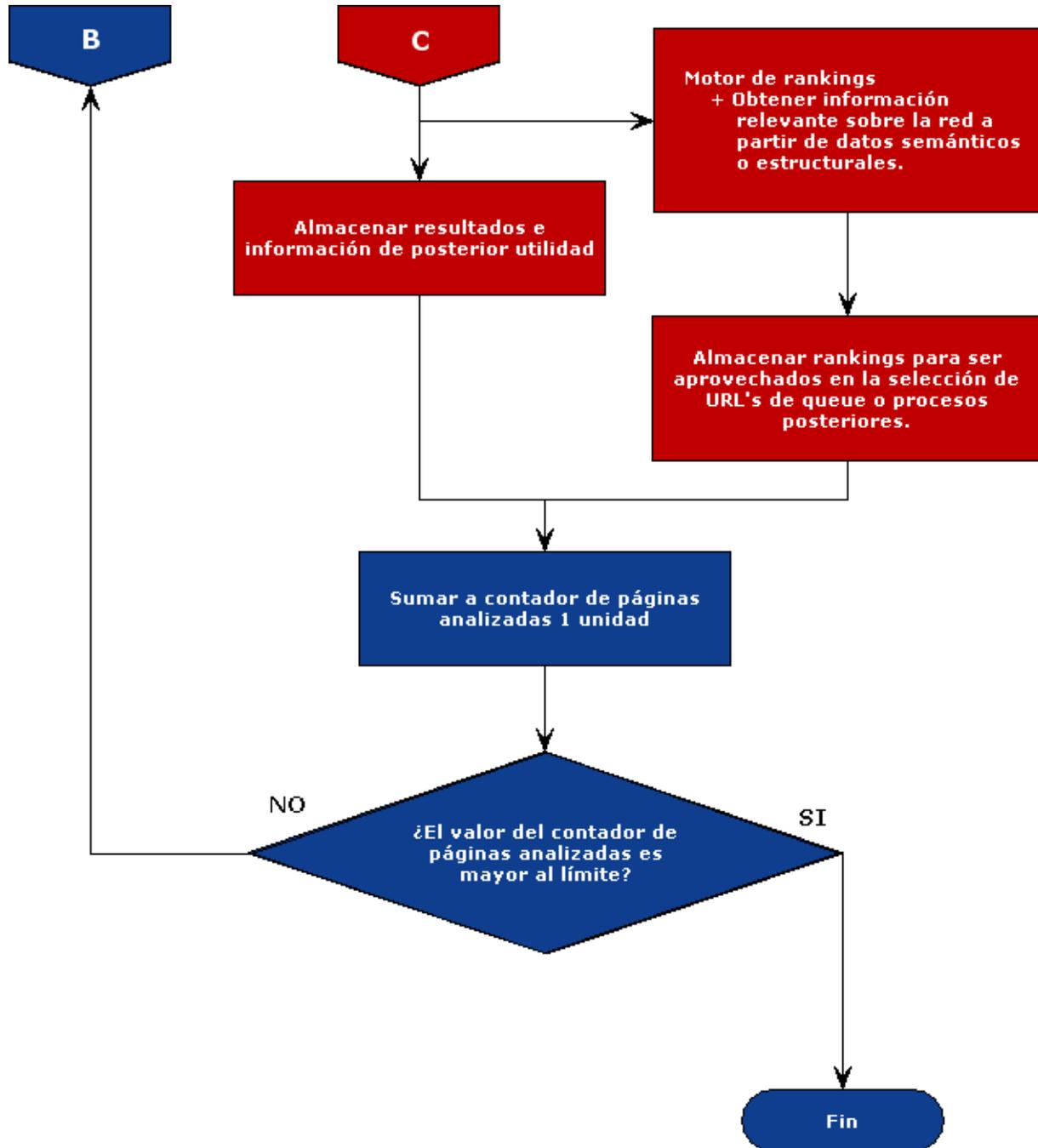


Figura B.3. – Diagrama de flujo de un crawler básico



Apéndice C

Información estadística sobre la pornografía en Internet – Año 2006

Las tablas que a continuación se muestran han sido obtenidas del sitio web <http://internet-filter-review.toptenreviews.com/internet-pornography-statistics.html>; el contenido ha sido traducido y depurado para facilitar la comprensión de los datos, lo anterior sin intención de manipular o sesgar los resultados.

Tabla C.1. - Ingresos mundiales de la industria dedicada a la comercialización de la pornografía en el año 2006.

Ingresos Mundiales por Pornografía – 2006

País	Ingresos [Billones USD]	Notas
China	\$27.40	Incompleto
Corea del Sur	\$25.73	
Japón	\$19.98	
Estados Unidos	\$13.33	
Australia	\$2.00	
Reino Unido	\$1.97	
Italia	\$1.40	
Canadá	\$1.00	
Filipinas	\$1.00	
Taiwan	\$1.00	Incompleto
Alemania	\$.64	Incompleto
Finlandia	\$.60	Incompleto
República Checa	\$.46	Incompleto
Rusia	\$.25	Incompleto
Holanda	\$.20	
Brazil	\$.10	Incompleto
Otros 212 países	-	No Disponible



Tabla C.2. - Distribución demográfica aproximada en la búsqueda de términos para adultos en el año 2006.

Distribución Demográfica Aproximada en Búsqueda de Contenido para Adultos – 2006

Distribución Aproximada	
Menores a 18 años	20.50%
Entre 18 y 24 años	19.50%
Entre 25 y 34 años	18.30%
Entre 35 y 49 años	23.50%
Mayores a 50 años	18.20%

Tabla C.3. - Términos más buscados de contenido para adultos con división por género en el año 2006.

Términos más Buscados de Contenido para Adultos – 2006

Término de Búsqueda	Peticiónes de Búsqueda en 2006	División por Género		Páginas Web que Contienen el Término [Millones]
		Masculino	Femenino	
Sex	75,608,612	50.00%	50.00%	414
Adult Daiting	30,288,325	36.00%	64.00%	1.4
Adult DVD	13,684,718	58.00%	42.00%	1.82
Porn	23,629,211	96.00%	4.00%	88.8
Sex Toys	15,955,566	58.00%	42.00%	2.65
Teen Sex	13,982,729	44.00%	56.00%	2.1
Free Sex	13,484,769	44.00%	56.00%	2.42
Adult Sex	13,362,995	36.00%	64.00%	1.58
Sex Ads	13,230,137	50.00%	50.00%	0.28
Group Sex	12,964,651	50.00%	50.00%	2.07
Free Porn	12,964,651	97.00%	3.00%	2.74
XXX	12,065,000	50.00%	50.00%	181
Sex Chat	11,861,035	50.00%	50.00%	2.21
Anal Sex	9,960,074	67.00%	33.00%	2.95
Cyber Sex	8,502,524	41.00%	59.00%	1.24
XXX Videos	7,411,220	64.00%	37.00%	1.44
Play boy	6,641,209	86.00%	14.00%	43.2
Teen Porn	6,130,065	82.00%	18.00%	1.97
Nude	5,487,925	77.00%	23.00%	71.3
Sexy	4,344,924	50.00%	50.00%	198



Tabla C.4. - Información estadística sobre la pornografía en Internet en el año 2006.

Información Estadística sobre la Pornografía en Internet – 2006

Páginas web pornográficas	420 millones
Búsquedas diarias sobre pornografía en buscadores	68 millones (25% del total de búsquedas en Internet)
Correos electrónicos enviados diariamente con contenido pornográfico o sexual	2.5 billones (8% del total de correos electrónicos)
Usuarios de Internet que accesan a contenido pornográfico	42.70%
Usuarios con exposición a contenido sexual no deseado	34.00%
Correos electrónicos diarios con contenido sexual por usuario	4.5 por usuario de Internet
Descargas (Peer To Peer) mensuales de contenido pornográfico	1.5 billones (35% del total de descargas)

Tabla C.5. - Número de páginas pornográficas por país en el año 2006.

Páginas Pornográficas por País – 2006

País	Páginas Pornográficas [millones]
Estados Unidos	244.66
Alemania	10.03
Reino Unido	8.51
Australia	5.66
Japón	2.7
Holanda	1.88
Rusia	1.08
Polonia	1.05
España	0.85



Tabla C.6. - Información demográfica sobre el acceso y consumo de contenido pornográfico en Internet en el año 2006.

Información Demográfica y Pornografía en Internet – 2006

Adolescentes	
Edad de primera exposición a contenido pornográfico	11 años
Grupo consumidor de pornografía en Internet	35 - 49 años
Grupo de adolescentes entre 15 y 17 años con experiencia sexual	80.00%
Grupo de adolescentes entre 8 y 16 años que han accedido a contenido pornográfico en Internet	90% (la mayoría al realizar tarea escolar)
Grupo de adolescentes entre 7 y 17 años que proporcionan su dirección de casa abiertamente	29.00%
Grupo de adolescentes entre 7 y 17 años que proporcionan correo electrónico abiertamente	14.00%
Mujeres	
Mujeres que mantienen en secreto su acceso a contenido pornográfico	70.00%
Mujeres con adicción a la consulta de contenido pornográfico	17.00%
Proporción de mujeres:hombres en salas de chat virtuales	02:01:00
Relación mujeres:hombres que accesan a sitios pornográficos	01:03:00
Mujeres que accesan a sitios pornográficos	9.4 millones (mensualmente)
Mujeres que admiten accesar a contenido pornográfico en el trabajo	13.00%
Hombres	
Hombres que admiten accesar a contenido pornográfico en el trabajo	20.00%
Adultos estadounidenses que acceden regularmente a sitios pornográficos	40 millones
Hombres con adicción a la consulta de contenido pornográfico	10.00%



Apéndice D

Resultados complementarios

A continuación se muestran algunas gráficas complementarias utilizando procedimientos análogos a los planteados en el desarrollo experimental de esta tesis con los conjuntos de información alternos; así mismo, se puede generar el mismo análisis aplicado en la clasificación de contenido pornográfico y obtener conclusiones particulares para cada caso.



Gráfica D.1. – Gráfica de dispersión de la fila de la matriz de ranking correspondiente al vector de referencia católico contra todos los demás.

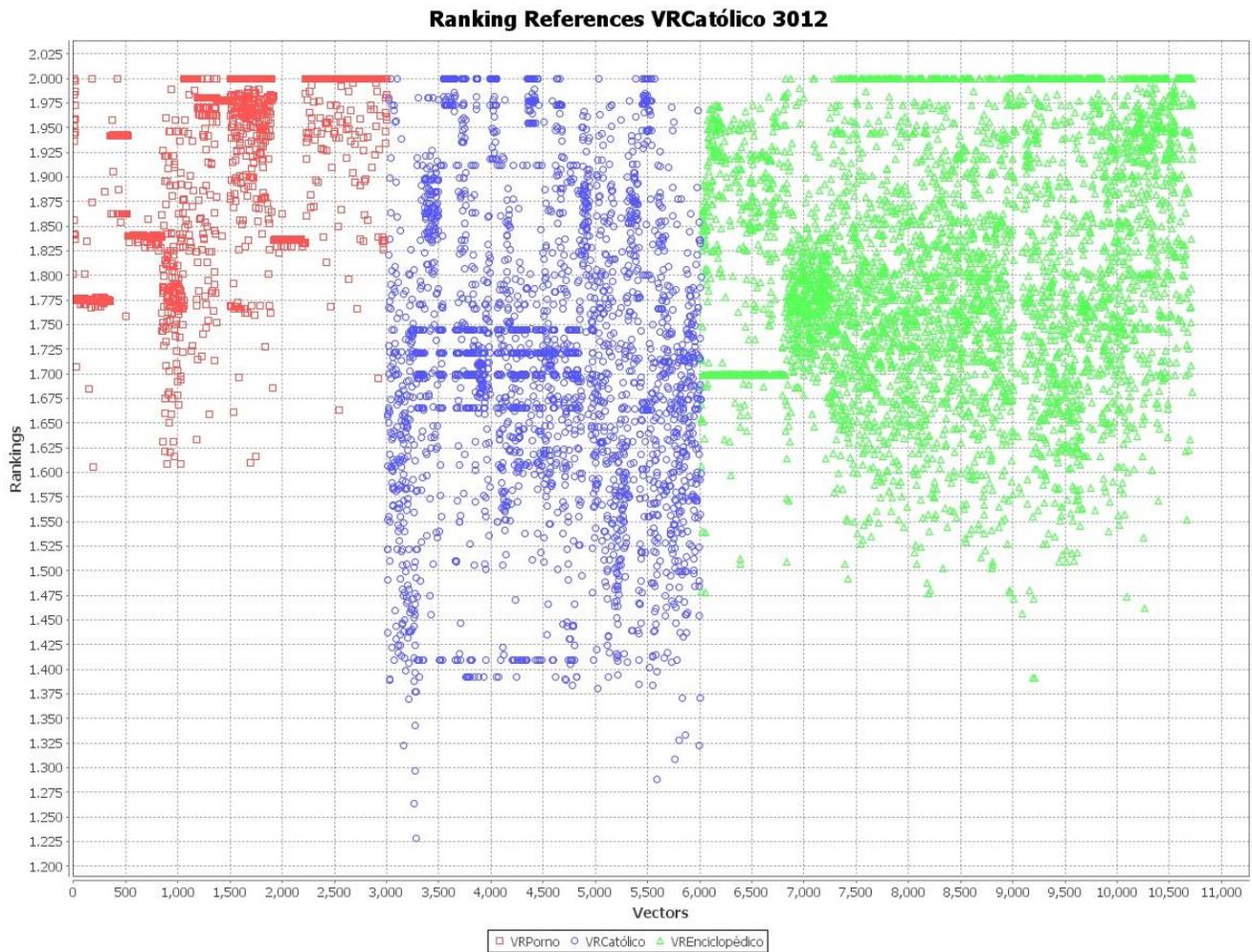


Tabla D.1. – Distribución de valores máximos, mínimos y medios para cada entorno de información con respecto al vector de referencia católico.

	\bar{r}	\bar{r}_{min}	\bar{r}_{max}
Entorno Pornográfico	1.8846	1.5082	2
Entorno Católico	1.5867	1.0885	2
Entorno Enciclopédico	1.7831	1.0885	2



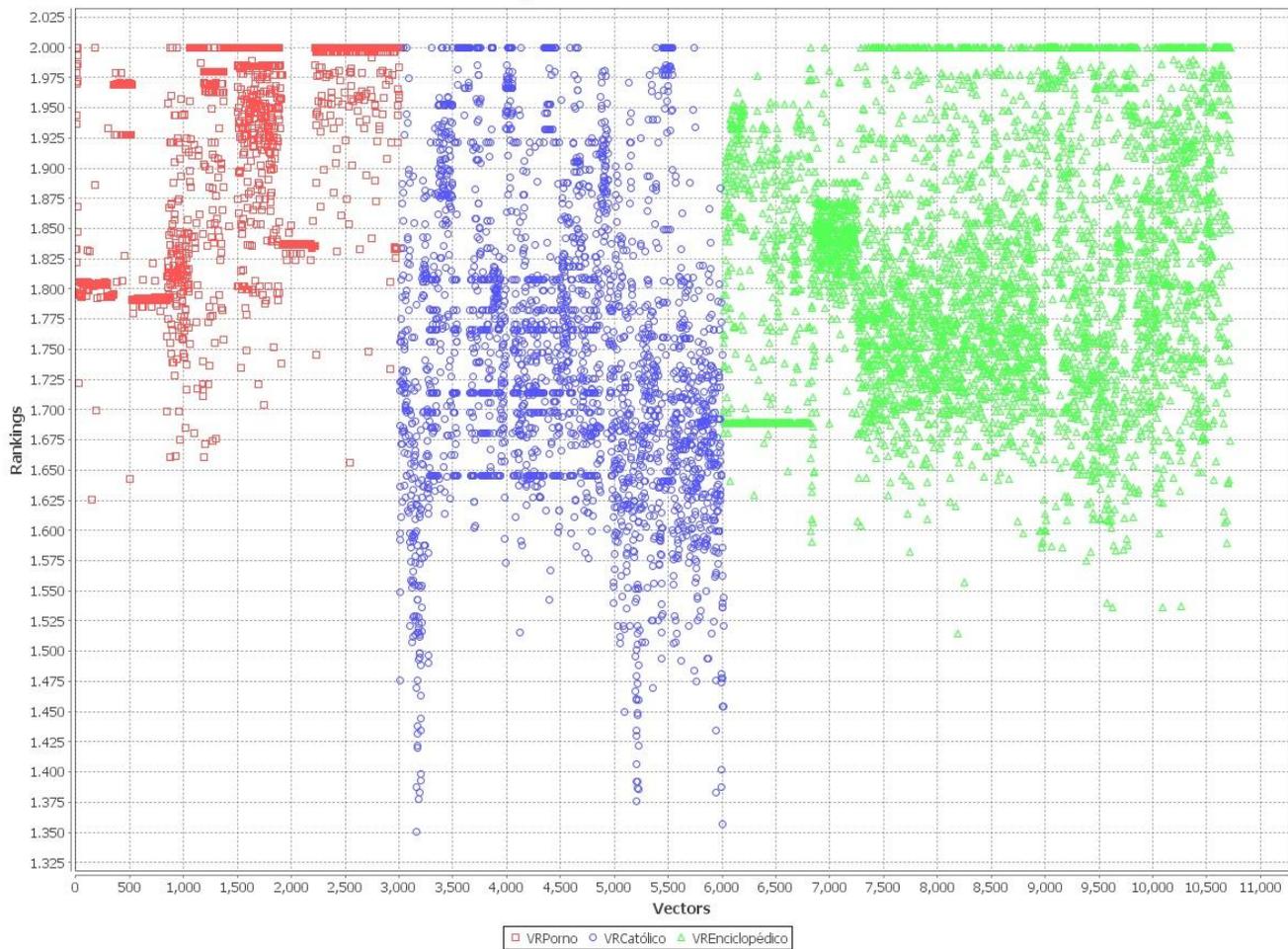
Gráfica D.2. – Gráfica de dispersión de la fila de la matriz de ranking correspondiente al vector 3011 del conjunto católico contra todos los demás.



Gráfica D.3. – Gráfica de dispersión de la fila de la matriz de ranking correspondiente al vector 3012 del conjunto católico contra todos los demás.



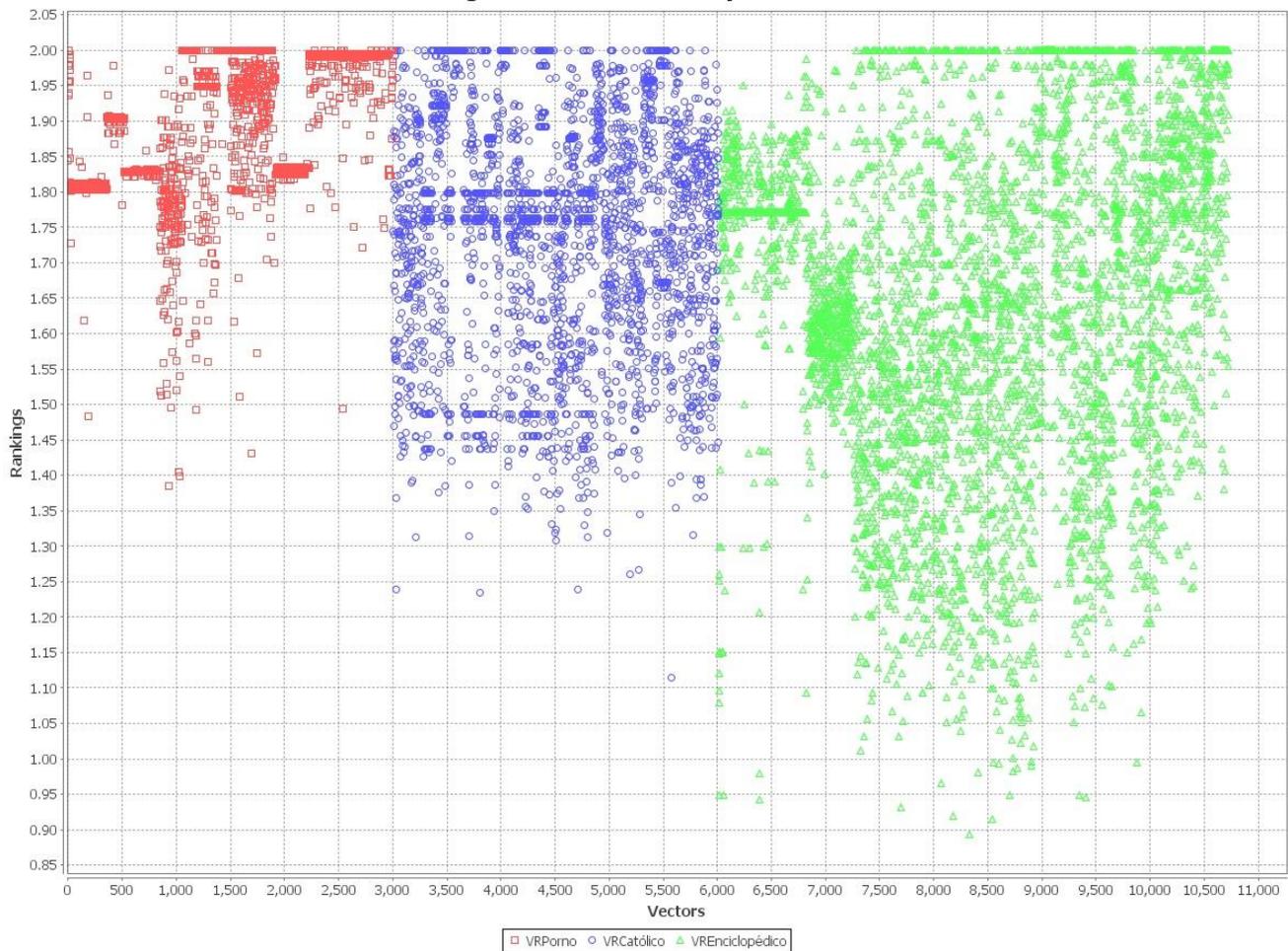
Ranking References VRCatólico 6000



Gráfica D.4. – Gráfica de dispersión de la fila de la matriz de ranking correspondiente al vector 6000 del conjunto católico contra todos los demás.



Ranking References VREnciclopédico 50 nodes



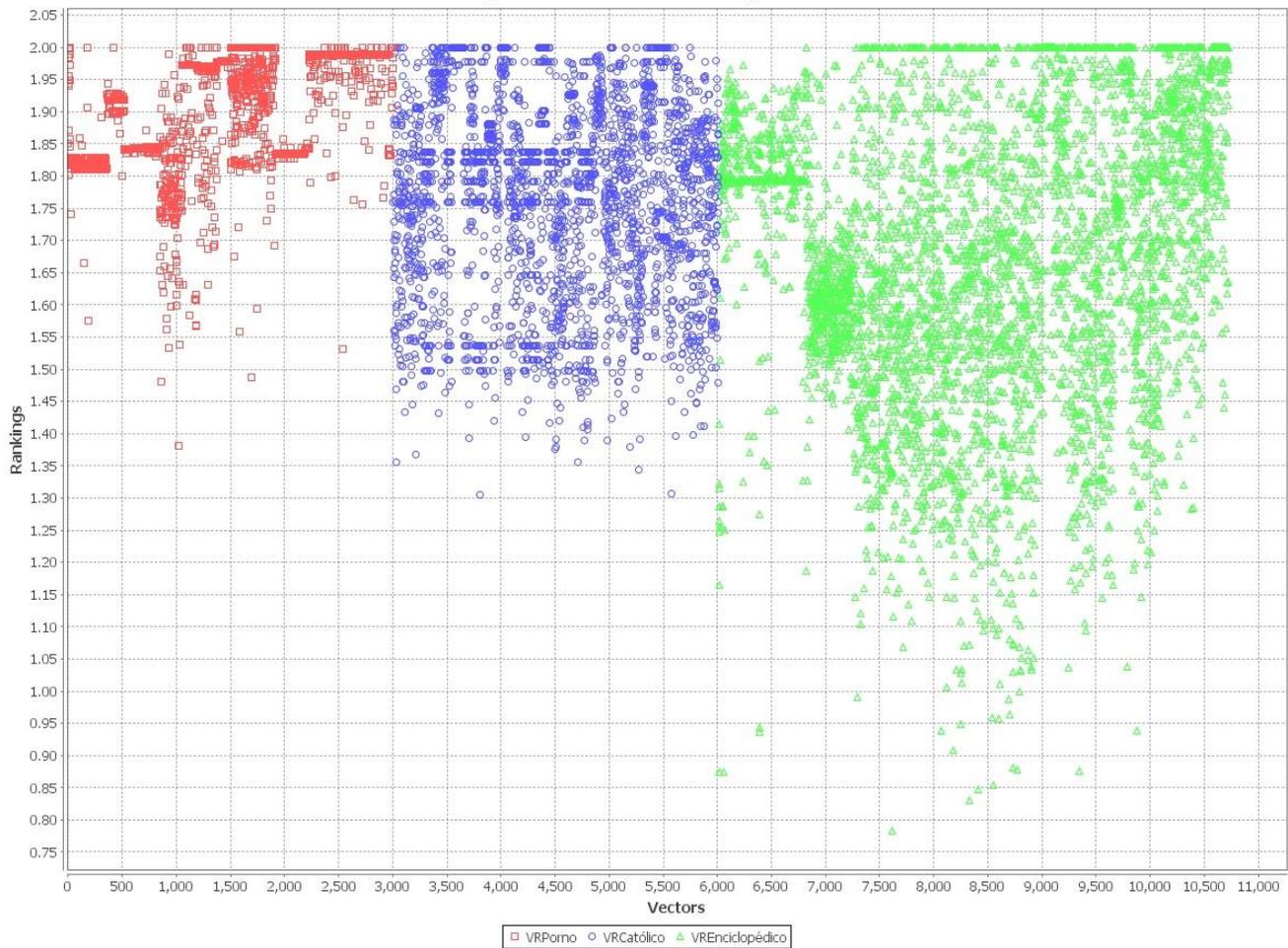
Gráfica D.5. – Gráfica de dispersión de la fila de la matriz de ranking correspondiente al vector de referencia enciclopédico contra todos los demás.

Tabla D.2. – Distribución de valores máximos, mínimos y medios para cada entorno de información con respecto al vector de referencia enciclopédico.

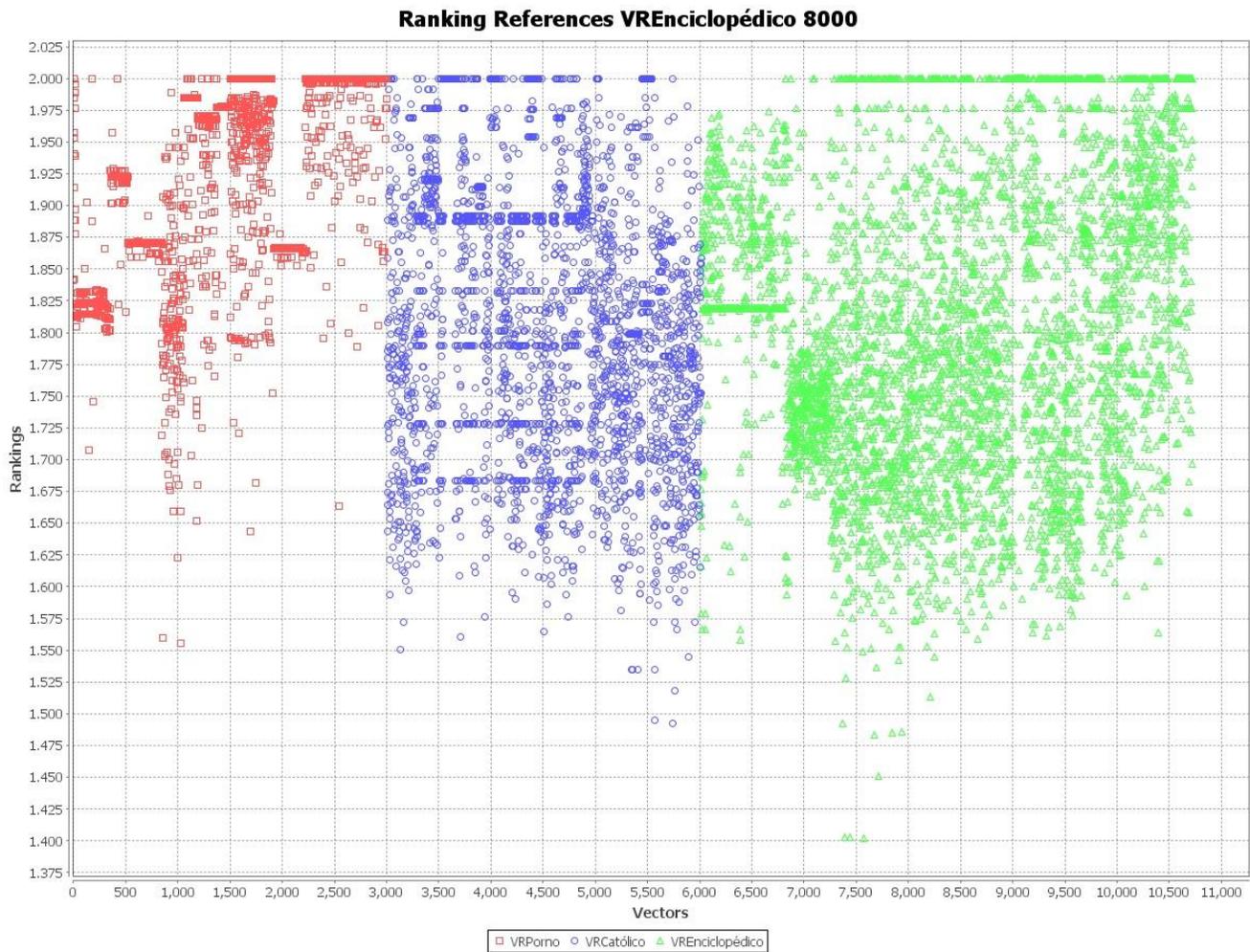
	\bar{r}	\bar{r}_{min}	\bar{r}_{max}
Entorno Pornográfico	1.8998	1.3853	2
Entorno Católico	1.7556	1.1149	2
Entorno Enciclopédico	1.5670	0.8939	2



Ranking References VREnciclopédico 7694



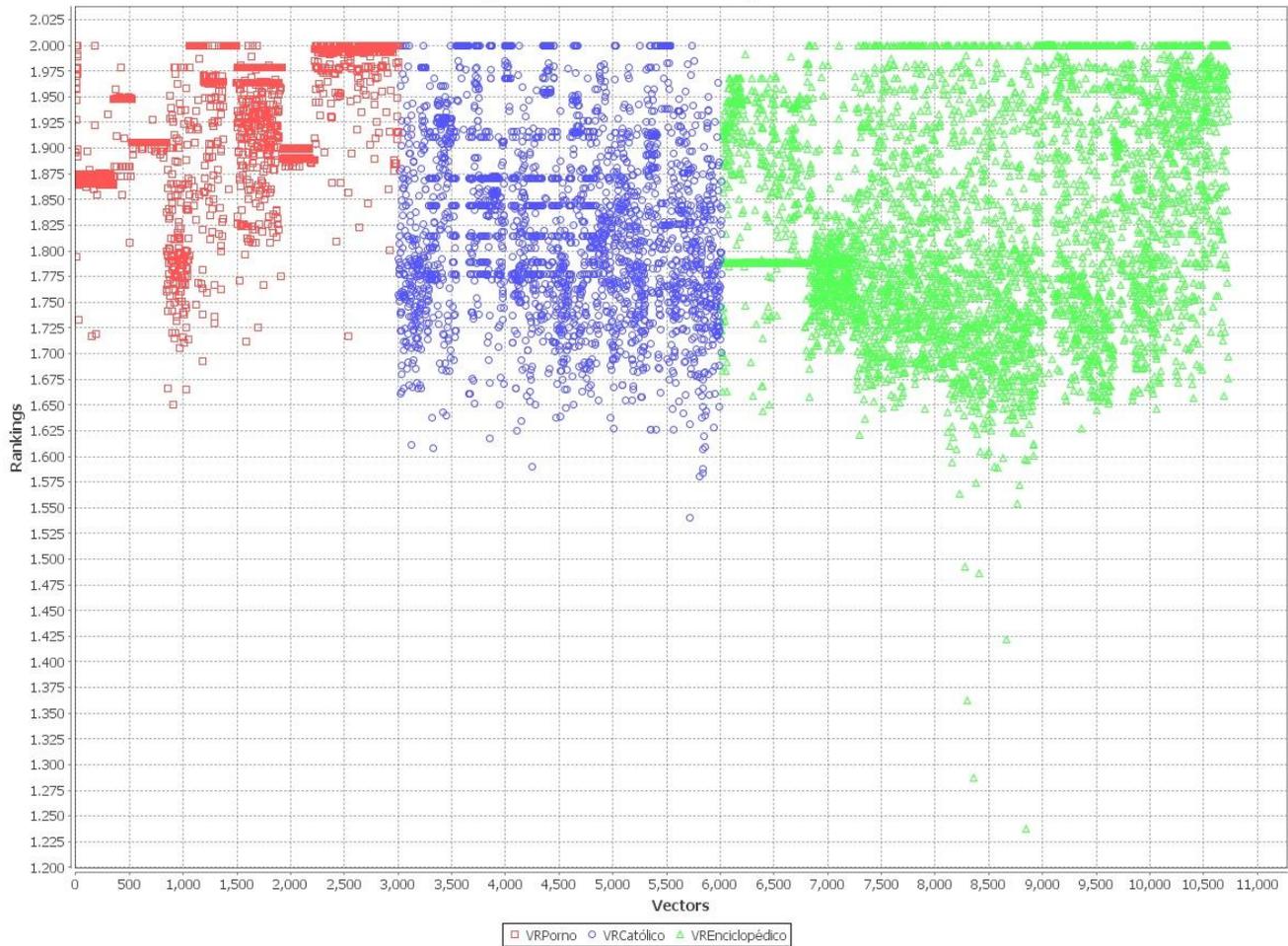
Gráfica D.6. – Gráfica de dispersión de la fila de la matriz de ranking correspondiente al vector 7694 del conjunto enciclopédico contra todos los demás.



Gráfica D.7. – Gráfica de dispersión de la fila de la matriz de ranking correspondiente al vector 8000 del conjunto enciclopédico contra todos los demás.



Ranking References VREnciclopédico 8237



Gráfica D.8. – Gráfica de dispersión de la fila de la matriz de ranking correspondiente al vector 8237 del conjunto enciclopédico contra todos los demás.



Las gráficas D.1. a D.8. fueron obtenidas al calcular los rankings de cada uno de los vectores especificados contra todos los demás elementos inmersos en el sistema; su análisis puede ser desarrollado siguiendo los procedimientos establecidos en el capítulo quinto de la presente tesis. A partir de una rápida interpretación, las observaciones más importantes son las siguientes:

- Mientras más semejantes sean los conjuntos de información entre sí, menor será la separación generada a través del ranking con respecto a un mismo vector de referencia, por lo tanto, mayor será la dificultad para clasificar dicho contenido, o en otras palabras, se tendrá una eficiencia máxima muy por debajo de los resultados obtenidos con escenarios donde los conjuntos de información son básicamente polares.
- Los mejores resultados serán obtenidos al trabajar con el vector de referencia perteneciente al conjunto de información más diferente con respecto al resto presentes y seleccionando la hipótesis más adecuada, según el objetivo del sistema en particular.