

Examen – Sistemas Inteligentes (Pauta)

Martes 29 de noviembre de 2016

Profesor: Alejandro Figueroa

Ayudante: Alexander Espina

- Está prohibido el uso de teléfonos celulares durante el desarrollo de la prueba.
- La prueba debe responderse con un lápiz de tinta indeleble, de lo contrario no hay opción a correcciones.
- Cualquier alumno que sea sorprendido intentando copiar será calificado con una nota 1.
- Está prohibido conversar durante la prueba. Recuerde que su compañero puede estar concentrado y el ruido puede perturbarlo en el desarrollo de su prueba.
- Utilice sólo las hojas entregadas para escribir sus respuestas.
- La nota 4.0 se alcanza con 60 de los 100 puntos que tiene la prueba.

Pregunta 1 (20 puntos)

Indique si cada una de las siguientes aseveraciones es verdadera (V) o falsa (F). No es necesario justificar las falsas. Cada respuesta correcta vale 2 punto, mientras que las incorrectas descuentan 1 puntos. Letras que no sean nítidas serán consideradas malas.

1. ..V.. El crawler necesita una política para poder recolectar páginas web.
2. ..F.. El parser se dedica a analizar el contenido de las páginas web.
3. ..F.. El analizador de vínculos analiza la importancia de cada página de acuerdo a su contenido.
4. ..V.. El ranker establece la relación entre la consulta y los documentos indexados.
5. ..F.. Los modelos de relevancia producen una lista de documentos ordenados de forma independiente de la consulta.
6. ..V.. Los modelos de impostancia analizan la importancia propia de cada página web.
7. ..F.. Los modelos booleanos proveen un score para cada documento.
8. ..V.. Los modelos de espacio vectorial representan los documentos como vectores en un espacio.
9. ..V.. Existen muchas representaciones posibles para un documento en un modelos de espacio vectorial.
10. ...V.. Si el MRR de un sistema es uno, entonces sabemos que su accuracy es uno.

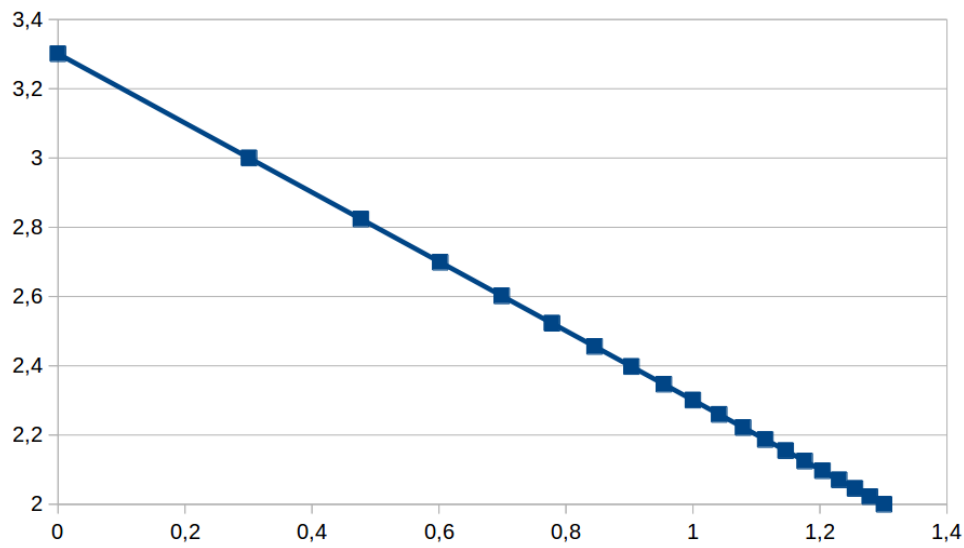
Pregunta 2 (40 puntos)

Considere una colección de documentos sobre construccion que tiene las siguientes veinte palabras más frecuentes (en orden):

a, the, of, chainsaw, trunk, log, tree, fall, cut, leaves, desk, rotten, sky, rain, woods, forest, land, mood, puddle y mud. Además, sabemos que la frecuencia de la última palabra es cien. Estime la frecuencia de cada una de las palabras restantes (5 puntos), y además el número de tokens/palabras que tiene esta colección (10 puntos). Grafique estas palabras de acuerdo a la ley de Zipf en escala logarítmica (25 puntos).

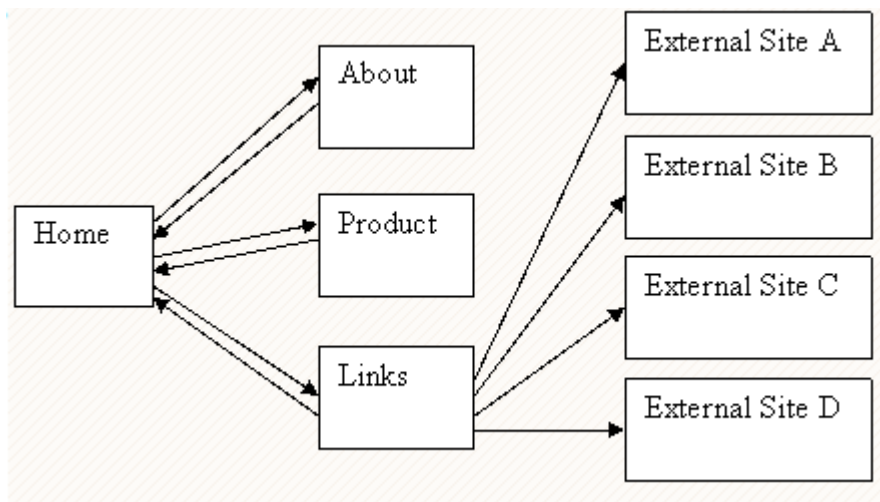
Termino	Frecuencia	position	c	f1	eje X	eje y
a	2000	1	2	1000	0	3,30103
the	1000	2	2	1000	0,30103	3
of	666,666667	3	2	1000	0,47712125	2,82390874
chainsaw	500	4	2	1000	0,60205999	2,69897
trunk	400	5	2	1000	0,69897	2,60205999
log	333,333333	6	2	1000	0,77815125	2,52287875
tree	285,714286	7	2	1000	0,84509804	2,45593196
fall	250	8	2	1000	0,90308999	2,39794001
cut	222,222222	9	2	1000	0,95424251	2,34678749
leaves	200	10	2	1000	1	2,30103
desk	181,818182	11	2	1000	1,04139269	2,25963731
rotten	166,666667	12	2	1000	1,07918125	2,22184875
sky	153,846154	13	2	1000	1,11394335	2,18708664
rain	142,857143	14	2	1000	1,14612804	2,15490196
woods	133,333333	15	2	1000	1,17609126	2,12493874
forest	125	16	2	1000	1,20411998	2,09691001
land	117,647059	17	2	1000	1,23044892	2,07058107
mood	111,111111	18	2	1000	1,25527251	2,04575749
puddle	105,263158	19	2	1000	1,2787536	2,02227639
mud	100	20	2	1000	1,30103	2

Tokens	7195,47931
K	50
B	0,5
M	4241,30856



Pregunta 3 (40 puntos)

Calcule el PageRank de la siguiente estructura de páginas web (10 iteraciones).



	2.02	0.7223333333333334	0.7223333333333334	0.7223333333333334	0.2727966666666667	0.2727966666666667	0.2727966666666667	0.2727966666666667
0	1.5007633333333334	0.5752162777777777	0.5752162777777777	0.5752162777777777	0.2477867672222222	0.2477867672222222	0.2477867672222222	0.2477867672222222
1	1.2256544394444442	0.4972687578425925	0.4972687578425925	0.4972687578425925	0.23453568883324072	0.23453568883324072	0.23453568883324072	0.23453568883324072
2	1.079892577165648	0.45596956353026696	0.45596956353026696	0.45596956353026696	0.22751482580014537	0.22751482580014537	0.22751482580014537	0.22751482580014537
3	1.0026630838015993	0.43408787374378643	0.43408787374378643	0.43408787374378643	0.22379493853644367	0.22379493853644367	0.22379493853644367	0.22379493853644367
4	0.9617443239008806	0.4224942251052495	0.4224942251052495	0.4224942251052495	0.22182401826789241	0.22182401826789241	0.22182401826789241	0.22182401826789241
5	0.9400642009468165	0.416351523601598	0.416351523601598	0.416351523601598	0.22077975901227165	0.22077975901227165	0.22077975901227165	0.22077975901227165
6	0.9285773491349882	0.4130969155882467	0.4130969155882467	0.4130969155882467	0.22022647565000192	0.22022647565000192	0.22022647565000192	0.22022647565000192
7	0.9224912321500214	0.41137251577583944	0.41137251577583944	0.41137251577583944	0.21993332768189272	0.21993332768189272	0.21993332768189272	0.21993332768189272
8	0.9192666045008198	0.4104588712752323	0.4104588712752323	0.4104588712752323	0.2197780081167895	0.2197780081167895	0.2197780081167895	0.2197780081167895
9	0.9175580892846844	0.4099747919639939	0.4099747919639939	0.4099747919639939	0.21969571463387894	0.21969571463387894	0.21969571463387894	0.21969571463387894
10	0.9166579600776596	0.4097102106000777	0.4097102106000777	0.4097102106000777	0.21965711700251695	0.21965711700251695	0.21965711700251695	0.21965711700251695
11								