

¿Personas o máquinas? Servicios de preguntas y respuestas vs. buscadores Web

RICARDO BAEZA-YATES
ricardo.baeza@upf.edu

LILIANA CALDERÓN-BENAVIDES
liliana.calderon@upf.edu

CRISTINA GONZÁLEZ-CARO
cristina.gonzalez@upf.edu

MARI-CARMEN MARCOS
mccarmen.marcos@upf.edu

Universitat Pompeu Fabra
La Rambla, 30 - 08002 Barcelona
Tel. 93 542 24 96/1433

RESUMEN

Los servicios de preguntas y respuestas que ofrecen algunos sitios web son, al igual que los buscadores, una forma de obtener información en la Web. A diferencia de los buscadores, las preguntas a estos servicios se realizan en lenguaje natural y son respondidas por otros usuarios. Partiendo de que las preguntas que las personas nos hacemos se pueden clasificar en función del tipo de respuestas que esperamos (objetivas con una sola respuesta válida posible, objetivas con más de una res-

puesta válida posible o subjetivas), el estudio muestra cómo los buscadores ofrecen resultados más relevantes para las preguntas que esperan respuestas con una sola posible respuesta correcta. No se corrobora la hipótesis de que los servicios de preguntas/respuestas sean mejores para dar respuestas que se esperan que sean subjetivas, de hecho, este tipo de preguntas no quedan bien respondidas ni por los buscadores ni por los servicios de preguntas/respuestas.

PALABRAS CLAVE: Recuperación de información; Servicios de preguntas y respuestas; Buscadores; Web social; Evaluación; Satisfacción.

1. Introducción

Uno de los fenómenos más importantes de la web en los últimos tiempos es la proliferación de servicios basados en la participación de los usuarios como blogs y wikis, y servicios para compartir información, muchas veces creada por ellos mismos, por ejemplo música, fotos, vídeos o enlaces favoritos. Es lo que se denomina Web Social, Web 2.0 o Social Media (Baeza-Yates; Marcos 2007; Marcos 2008).

Recientemente se han puesto en marcha diversas comunidades con servicios de preguntas y respuestas (P/R) en los que los usuarios plantean preguntas de cualquier temática y otros usuarios les responden. Los más populares en este momento son Yahoo! Answers (<http://answers.yahoo.com>) y Live QnA (<http://qna.live.com>), el primero a la cabeza por ser pionero, contar con más usuarios y estar disponible para comunidades de varios países, entre ellos España. También las comunidades profesionales están comenzando a incorporar secciones donde sus miembros plantean preguntas y otros miembros responden. Este servicio, que sería de alguna forma una continuación de los foros, lo ofrece entre otros LinkedIn (<http://www.linkedin.com>).

Si bien es verdad que el uso de estos servicios de P/R va en aumento, los buscadores continúan siendo a día de hoy las herramientas más utilizadas para la recuperación de información en la Web. Las diferencias entre buscadores y servicios de P/R como forma de obtener información para una determinada necesidad son notorias en todo el proceso de recuperación, y ambos presentan ventajas e inconvenientes consecuencia de la interacción con máquinas o con personas. Entre las diferencias que más afectan a la recuperación de información se encuentran las siguientes:

- Lenguaje natural vs. lenguaje de interrogación. Los servicios de P/R utilizan el lenguaje natural para introducir las preguntas, mientras que los buscadores presentan mejores resultados si se plantean consultas (*queries*) menos elaboradas, es decir, sólo con palabras clave. Como consecuencia de esta diferencia, y para evitar confusiones, en este trabajo nos referiremos a «preguntas» como situaciones en las que una persona necesita cierta información. Utilizaremos la palabra «respuestas» para la información que se obtiene de servicios de P/R, y diremos «resultados» o «URLs» para indicar que la información procede de buscadores.
- Cantidad de información recuperada. Habitualmente los buscadores ofrecen más información que los servicios de P/R, si bien esto no significa que necesariamente sea más precisa ni de mejor calidad.
- Inmediatez de las respuestas y resultados. Mientras que los buscadores dan los resultados de forma inmediata, los servicios de P/R dependen de los usuarios que estén conectados y quieran responder a las preguntas.
- Capacidad de interacción con las respuestas y resultados. En el caso de los buscadores, la forma posible de interacción es el replanteamiento de la consulta para afinar los resultados. En cambio, en los servicios de P/R se puede llegar a generar un foro en el que la persona que formuló la pregunta y otros miembros de la comunidad dialogan y opinan acerca de las respuestas que se van dando.

Con este estudio queremos saber si los buscadores y los servicios de P/R pueden jugar un papel importante en la satisfacción de necesidades de información para distintos tipos de preguntas. A priori suponemos que los buscadores serán mejores para resolver las preguntas que esperan resultados objetivos, y en cambio los servicios de P/R serán mejores para recibir respuestas subjetivas. Por medio de un experimento vamos a comprobar si existe una correlación entre los tipos de preguntas y los tipos de servicios: buscadores y P/R.

2. Intencionalidad de las preguntas y expectativas en los resultados y las respuestas

Para este estudio contemplaremos distintos tipos de preguntas en función de las expectativas de resultados y respuestas que tienen quienes las plantean (Baeza-Yates; González-Caro; Calderón-Benavides, 2006; Jansen; Spink; Saracevic, 2000; Lee; Liu; Cho, 2005; Rose; Levinson 2004). La clasificación propuesta es la siguiente:

1. Preguntas que esperan una respuesta con información objetiva:

- Unirespuesta. Casos en los que sólo una respuesta es correcta, como ocurre en el caso de datos exactos, por ejemplo *en qué año terminó la guerra de secesión*
- Multirespuesta. Casos en los que varias respuestas diferentes pueden ser correctas, por ejemplo *cuál es la receta del tiramisú*

2. Preguntas que esperan una respuesta subjetiva, por ejemplo *qué hotel hay céntrico en Barcelona con buena relación calidad/precio*

3. Planteamiento del experimento

3.1. El perfil de los usuarios

Hay que decir que la batería de consultas a partir de la que desarrollamos el experimento está relacionada con la informática, por lo tanto el perfil de los usuarios que realicen las pruebas (su formación y su experiencia en tecnologías) será una variable importante para considerar. En primer lugar, todos ellos deberán poseer conocimientos de informática y de inglés suficientes para comprender las preguntas y poder evaluar las respuestas y los resultados. A partir de esta premisa, y pensando en la continuidad de este trabajo, contaremos en las distintas fases con tres perfiles que creemos que pueden ser más. Por un lado —y concretamente para los resultados presentados en este trabajo— los profesionales de las Ciencias de la Documentación con conocimientos técnicos medios pero con gran experiencia en el manejo de internet y en especial de los buscadores. Por otro lado, y para experimentos futuros, los profesionales de las Ciencias Sociales, que si bien manejan los buscadores con asiduidad, no tienen una formación específica en el uso de éstos, y cuyos conocimientos en informática son también medios. Y en tercer y último lugar los profesionales de la Informática, expertos en las materias planteadas y también conocedores del uso de los buscadores

Un aspecto relacionado con los experimentos con sistemas de recuperación de información en los que participan usuarios es su conocimiento previo del tema sobre el

cual se trabaja. En este experimento en concreto nos hemos encontrado con usuarios que plantearon las consultas a buscadores usando información que ellos tenían pero que no les había sido proporcionada, por ejemplo ante la pregunta de «quién es el autor del libro de arquitectura de la información que tiene un oso polar en la portada», un usuario agregó el nombre de la editorial —O'Reilly— a la consulta.

Como adelantábamos, en esta primera fase que presentamos hemos contado con la colaboración de un grupo de profesionales de las Ciencias de la Documentación, concretamente con 20 usuarios para realizar la parte de generación de consultas y con 8 para la parte de evaluación de respuestas y resultados. En futuras pruebas se realizará este experimento con otros perfiles y podremos estudiar cómo la formación previa en la materia influye en la forma de plantear las preguntas y en la forma de evaluar los resultados y respuestas obtenidas (Cox; Fisher, 2004).

3.2. Las preguntas

Para estudiar los resultados de los buscadores y las respuestas de los servicios de P/R se han preparado 3 bloques de preguntas en lenguaje natural con 4 preguntas en cada uno, de manera que en total hay un conjunto de 12 preguntas.

— Preguntas que esperan respuestas «unirespuesta» (una sola respuesta es correcta):

1. *Who was the Google creator and in what year created it?*
2. *What is the tag used to start and finish a paragraph in html?*
3. *How many KiloBytes are in a TeraByte?*
4. *Who is the author of the information architecture book which is famous for its polar bear on the cover?*

— Preguntas que esperan respuestas «multirespuesta» (son posibles varias respuestas):

5. *I would like to know which are the main Linux distributions and their advantages and disadvantages*
6. *What do I need to do to make PHP and MySQL work together?*
7. *My antispam filter blocks emails that aren't spam and I don't know how to solve this problem. What can I do?*
8. *What is the difference between, say, a software that is «Open Access» and one that is «Open Source»?*

— Preguntas que esperan respuestas «subjetivas» (implican opiniones):

9. *I'm a beginner on programming with Java Server Pages, could you recommend some web sites to me?*
10. *I would like to do a master on computer sciences, which one can you recommend to me?*
11. *I would like to buy a laptop and I have been thinking about a Sony Vaio, what is your opinion about them?*
12. *Which programming language is more robust C# or Java?*

A partir de estas 12 preguntas se ha pedido a los usuarios que formulen las consultas para los buscadores y se han recogido los resultados; se han enviado las preguntas a los servicios de P/R; y se ha pedido a 8 usuarios que evalúen los resultados y respuestas sin decirles de qué servicios provenían. A continuación detallamos cómo se han llevado a cabo estas acciones.

3.3. Formulación de las consultas a los buscadores

En primer lugar, se ha propuesto a los 20 usuarios que indiquen de qué forma buscarían información en un buscador sobre cada una de las 12 situaciones planteadas. De esta manera hemos obtenido un conjunto de 20×12 *queries* (240 consultas), que hemos lanzado a los 3 buscadores (Yahoo! Search, Google y Live) y de los que hemos recogido los 10 primeros resultados que ha dado cada uno de ellos. En caso de obtener resultados anidados, la herramienta utilizada para la recogida de URLs los ha considerado como una página web distinta. De esta forma se han obtenido un total de 6.653 URLs, sin tener en cuenta la repetición de algunas de ellas, que ha sido del 52%. Hay que decir que no se aprecian diferencias sustanciales en los porcentajes de URLs únicas y repetidas en los diferentes buscadores pues los valores son muy cercanos: 44% de URLs únicas en Google, 47% en Yahoo! y 53% en Live.

Dado el contexto de la investigación, en el que para la misma situación fueron planteadas 20 preguntas (una por cada usuario), y teniendo en cuenta que para contabilizar las URLs únicas y repetidas sólo se han tenido en cuenta los 10 primeros resultados (los supuestamente más relevantes), era de esperar que la concentración de URLs hubiera sido mayor. Sin embargo, esto es un reflejo de que la forma de plantear la consulta afecta los resultados a obtener. Veamos un ejemplo de cómo distintos usuarios plantearon la consulta para la siguiente pregunta: *I would like to know which are the main Linux distributions and their advantages and disadvantages.*

- *intitle: «Linux distributions» advantages and disadvantages*
- *Linux «advantages and disadvantages»*
- *Linux distributions*
- *linux distributors comparative*
- *all about linux distributions*

A partir de la colección de URLs se han tomado las 5 primeras obtenidas en cada *query* (consulta), que son las 5 primeras URLs resultado de lanzar las 12 *queries* de los 20 usuarios, es decir 1.200, y se han ordenado teniendo en cuenta la frecuencia con la que aparecían entre los 5 primeros resultados, así como la posición en la que aparecían (1, 2, 3, 4 o 5). De esta forma hemos obtenido un listado de resultados ordenado por relevancia, del que hemos tomado los 3 resultados más relevantes para ser evaluados por el segundo grupo de usuarios.

En esta fase del experimento somos conscientes de las limitaciones que tienen los estudios de laboratorio en los que intervienen usuarios, originadas por tratarse de un entorno ficticio en el que los usuarios deben plantear consultas para necesidades de información que realmente no tienen. A esto se añade que las preguntas/situaciones se les han proporcionado redactadas, lo que puede condicionar bastante su forma de plantear las consultas, sobre todo en cuanto a la elección del vocabulario que usarán. Finalmente encontramos que otro sesgo estaría en el dominio del idioma en el que están las preguntas (inglés). Al no tratarse del idioma materno de los usuarios el vocabulario y las estructuras gramaticales no serán tan ricas como las que usuarios nativos con un conocimiento similar sobre tecnologías.

3.4. Formulación de las consultas a los servicios de preguntas y respuestas

De forma paralela a las consultas realizadas en los buscadores se han lanzado las mismas preguntas en lenguaje natural a 3 servicios de preguntas y respuestas: Yahoo! Answers, Live QnA y la sección «Answers» de LinkedIn. En este último, teniendo en cuenta que se trata de una comunidad profesional, hemos optado por no enviar algunas preguntas cuya respuesta puede obtenerse fácilmente por otros medios. El motivo de hacer esto ha sido molestar lo menos posible a las personas que amablemente y en detalle responden a los colegas. No hemos enviado las preguntas que esperan respuestas unirespuesta ni tampoco una de las que esperaban respuesta multirespuesta. Las preguntas enviadas han sido las número 5, 6, 8, 9, 10, 11 y 12.

Las preguntas en lenguaje natural también se introdujeron en los 3 servicios de respuestas y se guardaron estas respuestas. Hay que decir que no fueron muy numerosas, concretamente una media de 4 respuestas por cada pregunta en Yahoo! Answers, 2 en Live QnA y 5 en LinkedIn.

Para la fase de evaluación con el segundo grupo de usuarios elegimos 3 respuestas para cada pregunta. En el caso de que hubiera respuestas sin contenidos reales se eliminaron, y así se quedó algo más reducido. Nos referimos a respuestas como la que dio una persona en Yahoo! Answers, donde tan sólo decía «depende». En algunas ocasiones no se obtuvieron respuestas para algunas preguntas, o se obtuvieron menos de 3, por lo que el número de respuestas disponibles para evaluar en cada pre-

gunta no siempre pudo ser 3. En los casos en que había más de 3 respuestas válidas para cada pregunta en un mismo servicio hemos optado por escoger las que llegaron primero.

Para poder comparar las cifras del número de respuestas obtenidas para cada pregunta en cada servicio tenemos que considerar que en LinkedIn no se introdujeron las preguntas 1, 2, 3, 4 y 7. Por lo tanto, haremos los cálculos con la información que tenemos suponiendo que para estas preguntas que no hicimos en LinkedIn se hubiera obtenido el número promedio de respuestas. Sólo de esta forma podemos establecer comparativas entre los 3 servicios.

Tabla 1. Cantidad de respuestas por servicio de P/R.

	Y! Answers	Live QnA	LinkedIn
Nº de respuestas totales	45	29	60
Promedio de respuestas por pregunta	4	2	5
Nº mayor de respuestas obtenidas por pregunta	7	6	8
Nº menor de respuestas obtenidas por pregunta	2	0	2

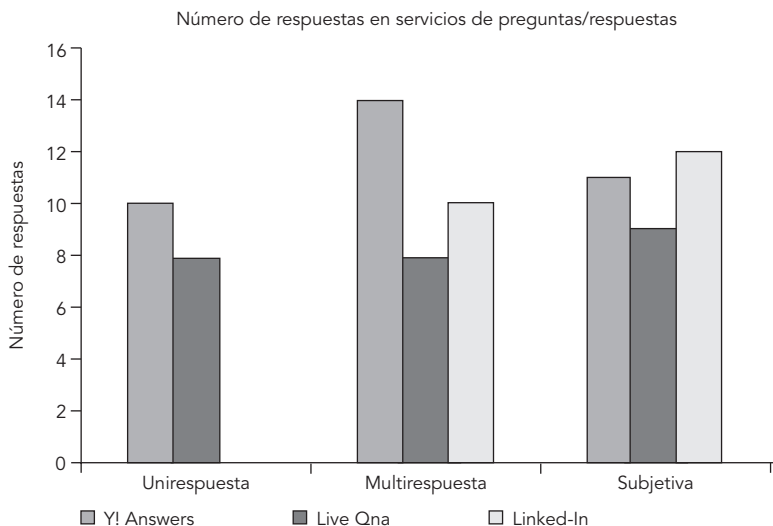
En la tabla puede verse cómo LinkedIn es el servicio que más respuestas generó, seguido de Yahoo! Answers. Live QnA es el que menos respuestas ha dado, con bastante diferencia con respecto a los otros dos.

Si observamos estos datos separando las respuestas por el tipo de preguntas y por los distintos servicios consultados, vemos que las preguntas que esperan respuestas de tipo unirespuesta tienen una cantidad ligeramente menor pero no muy significativa en relación con el número de los otros dos tipos de preguntas. En cambio sí se observa que Yahoo! Answers recibe considerablemente más respuestas que los otros servicios, en especial en el tipo multirespuesta, mientras que Live queda siempre en último lugar en los tres tipos de preguntas (Tabla 2, Figura 1).

Tabla 2. Número de respuestas obtenidas en los servicios de P/R

	Yahoo! Answers	Live Qna	Linked-In	Total	
Unirespuesta	10	8	(sin datos)	18	22%
Multirespuesta	14	8	10	32	39%
Subjetiva	11	9	12	32	39%
Total	35	25	22		
	42%	30,50%	27%		

Figura 1. Comparativa del número de respuestas obtenidas en los servicios de P/R



4. Evaluación de resultados y respuestas

4.1. Recogida de datos para la evaluación

Para la evaluación de los resultados y respuestas obtenidas de los 6 servicios hemos contado con un segundo grupo de usuarios del área de las Ciencias de la Documentación, esta vez han sido 8 usuarios, también con un conocimiento medio o avanzado de tecnologías web y de inglés. Hemos creado un sitio web vinculado a una base de datos al que los usuarios entraron con clave, de forma que cada uno podía ver sólo 3 consultas, una de cada tipo. El sistema se ha pensado de manera que los resultados y respuestas de cada pregunta (tanto URLs como respuestas humanas) seleccionados ha sido evaluada por 2 personas, lo que significa que las 8 personas han estado involucradas en la evaluación de resultados y respuestas de los 3 tipos de preguntas.

Para la evaluación se ha elaborado un formulario para cada pregunta en el que como máximo había 18 entradas: 9 URLs (3 de cada buscador, o menos en caso de que haya coincidencias) y 9 respuestas de servicios de P/R (a veces menos por no haber tenido al menos 3 respuestas para algunas preguntas).

Los usuarios no sabían la procedencia de las URLs ni de las respuestas. Cada uno de ellos ha determinado mediante un test si cada resultado o respuesta le sirve o no para su necesidad de información de forma excelente, suficiente, pobre o no le sirve (Figura 2).


Figura 2. Formulario de evaluación de páginas web obtenidas en los buscadores

Question 1 : Who was the Google creator and in what year created it?				
Web Pages				
<u>Web Page 1</u>	<input type="radio"/> Excellent	<input type="radio"/> Enough	<input type="radio"/> Poor	<input type="radio"/> Not Useful
<u>Web Page 2</u>	<input type="radio"/> Excellent	<input type="radio"/> Enough	<input type="radio"/> Poor	<input type="radio"/> Not Useful
<u>Web Page 3</u>	<input type="radio"/> Excellent	<input type="radio"/> Enough	<input type="radio"/> Poor	<input type="radio"/> Not Useful
<u>Web Page 4</u>	<input type="radio"/> Excellent	<input type="radio"/> Enough	<input type="radio"/> Poor	<input type="radio"/> Not Useful

Un problema que se nos planteó en la fase de evaluación era que los usuarios podían navegar en los resultados obtenidos de los buscadores y al evaluar no estarían considerando la página que les proporcionamos sino también otras a las que se llega desde ella. Para evitar esta posible navegación y con la intención de que todos se limitaran a las páginas dadas, optamos por convertir las páginas web objeto de evaluación a formato PDF. De esta manera conseguimos tener más controlado el experimento. Algo similar nos ocurría en algunas de las respuestas obtenidas en los servicios de preguntas y respuestas, donde los miembros de las comunidades a menudo proporcionaban URLs como respuesta, o bien acompañando a explicaciones propias. En estos casos hemos seguido la misma política: en el caso de haber URLs recomendadas, hemos convertido esas páginas a PDF.

En la figura 3 presentamos un ejemplo de la respuesta que da un miembro de una comunidad a la pregunta acerca de la creación de Google: *Who was the Google creator and in what year created it?*

Figura 3. Respuesta a una pregunta en la que la persona da una explicación y además recomienda visitar una página web

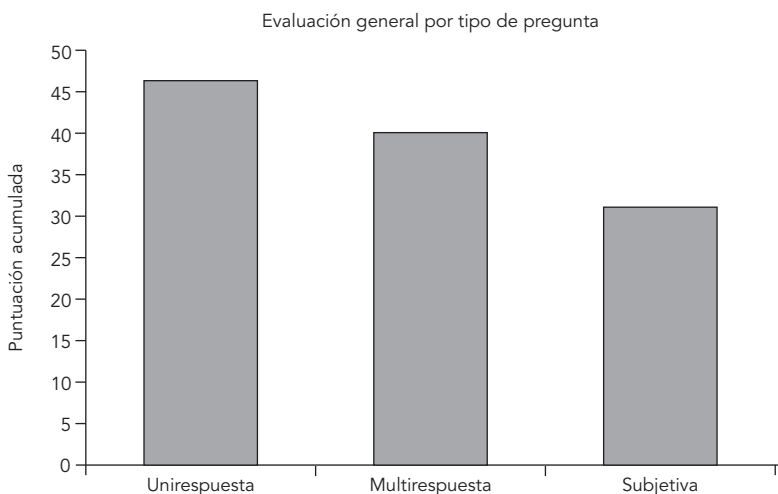
	EVALUATION TEST
<p>Actually, The founders of the Google internet search engine - Larry Page and Sergey Brin and it wasn't commercially launched until 1998.... from a friends garage. I remembered watching a show about it a couple years ago so I searched for something and found an ad on it. Here's a link that tells you a little more about them... http://news.bbc.co.uk/1/hi/business/3666241.stm</p>	

4.2. Resultados obtenidos

Con este experimento hemos obtenido una visión general de cómo los usuarios han evaluado las respuestas obtenidas, en promedio tenemos una calificación de 4 sobre 10, sorprendentemente baja. En cambio es más interesante cruzar información y descubrir diferencias según consideremos los tipos de preguntas planteadas y los tipos de servicios evaluados.

Curiosamente, las preguntas que esperaban una respuesta única posible han sido las mejor evaluadas, si bien la calificación no es muy alta, tan sólo un 5 sobre 10. Seguidas bastante de cerca, y prácticamente evaluadas con la misma puntuación, están las preguntas que esperan respuestas objetivas pero no necesariamente únicas (respuestas calificadas con un 3,7), y respuestas subjetivas (calificadas con un 3,5).

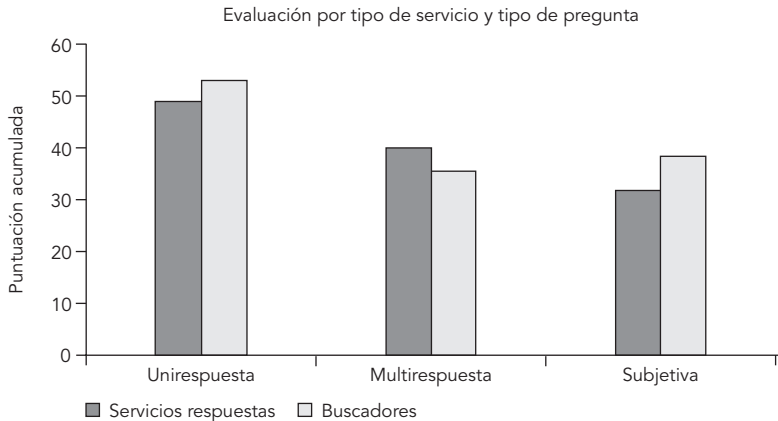
Figura 4. Evaluación realizada por los usuarios para los 3 tipos de preguntas



Tampoco encontramos una diferencia especialmente grande si consideramos los dos tipos de servicios por separado: los resultados obtenidos en los buscadores han recibido una puntuación media de 4,5, y las que proceden de servicios de respuestas un 3,9, apenas 3 décimas de diferencia sobre 10 puntos (Tablas 3 y 4).

En la figura 5 se observa cómo las respuestas dadas por los buscadores han sido algo mejor puntuadas en los casos de esperar respuestas únicas, mientras que en el caso de esperar varias posibles respuestas o respuestas subjetivas (que evidentemente también son a su vez múltiples) la diferencia apenas es significativa, 6 décimas.

Figura 5. Comparativa de la evaluación hecha por los usuarios para los distintos tipos de preguntas en buscadores y en servicios de P/R



Otros datos interesantes han surgido al comparar las calificaciones recibidas por los resultados obtenidos de los tres buscadores. Yahoo! queda en primer lugar, seguido muy de cerca por Google y a no mucha distancia de Live. En la tabla 3 pueden verse las puntuaciones sobre 10.

Llama la atención que las puntuaciones más altas en los tres casos se han obtenido en el tipo de preguntas unirespuesta (6,8 y 7,1 sobre 10), en cambio para los otros dos tipos de preguntas la puntuación en ningún caso llega a 5 puntos sobre 10, lo que indica que la satisfacción no ha sido alta con estos resultados en los buscadores.

Tabla 3. Evaluación de resultados dados por los buscadores

	Google	Yahoo!	Live	Promedio
Unirespuesta	6,8	7,1	4,0	6,0
Multirespuesta	3,6	4,4	2,8	3,6
Subjetiva	4,6	4,0	3,0	3,9
Promedio	5	5,2	3,3	4,5

Peores son los resultados de evaluación obtenidos por los servicios de preguntas y respuestas, que en ningún caso alcanzan el aprobado (tabla 4). En promedio, son las preguntas de tipo unirespuesta las que aparentemente salen favorecidas, si bien faltan datos de LinkedIn para poder afirmarlo con rotundidad. De todas formas, sí es cierto que las preguntas que esperan respuestas subjetivas han sido las peor valoradas. Si entramos en detalle en cada servicio, y siempre considerando la ausencia de datos en LinkedIn, vemos que Live QnA ha sido el servicio peor puntuado por los usuarios.

Tabla 4. Evaluación de respuestas dadas por los servicios de preguntas y respuestas

	Y! Answers	Live QnA	LinkedIn	Promedio
Unirespuesta	4,8	4,4	sin datos	4,6
Multirespuesta	3,4	4,9	4,0	4,1
Subjetiva	3,0	1,8	4,3	3,0
Promedio	3,7	3,7	4,2	3,9

Tabla 5. Puntuación obtenida por tipo de servicio

Tipo de servicio	Servicio	Puntuación sobre 10	Promedio
Buscadores	Google	5,0	4,5
	Yahoo!	5,2	
	Live	3,3	
Servicios P/R	Yahoo! Answers	3,7	3,9
	Live QnA	3,7	
	Linked-In	4,2	

5. Conclusiones

Aunque los servicios de respuestas suponen una nueva e interesante manera de ofrecer información a los usuarios a través de los resultados podemos ver que los servicios de búsqueda tradicionales siguen siendo una opción válida y acertada para resolver las diferentes necesidades de información de las personas.

Se puede establecer que los resultados para una consulta mejoran si la pregunta es específica, es decir, espera que la respuesta recibida sea objetiva; en cambio, cuando las preguntas esperan respuestas de tipo subjetivo, tanto los buscadores como los servicios de P/R obtienen resultados menos satisfactorios.

En la recogida de datos de los servicios de P/R, el servicio que más respuestas ha recibido, a pesar de no haber contemplado el grupo de preguntas unirespuesta, ha sido LinkedIn, con 60 respuestas, seguido de Yahoo! Answers con 45 y Live QnA con 29. Si revisamos estos datos considerando los tipos de preguntas, y de nuevo contando con que no tenemos datos de unirespuesta para LinkedIn, tenemos que el tipo con más respuestas ha sido el de multirespuesta con un 39% de participación.

A nivel individual los buscadores y servicios de respuestas tienen un comportamiento relativamente homogéneo, de hecho las diferencias entre los resultados ofrecidos por los diferentes tipos de servicios son muy cercanos y, según lo mencionado anteriormente, mejoran cuando las preguntas son de respuesta puntual.

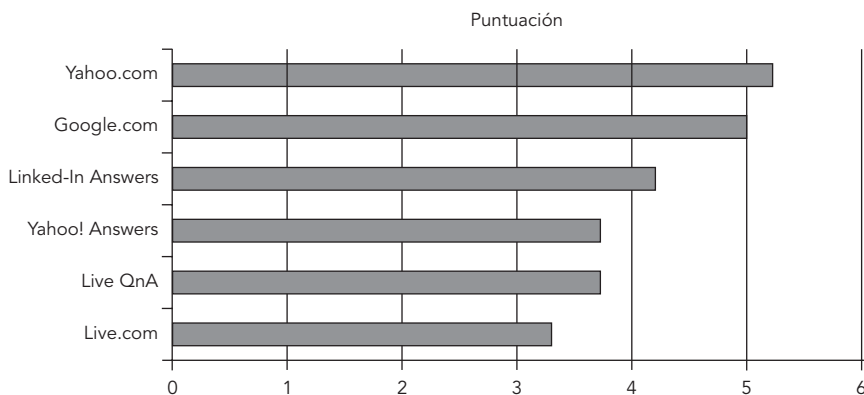
Si bien es cierto que los resultados globales han sido poco satisfactorios para los usuarios que evaluaron (4 puntos sobre 10), no se presenta una gran variación en la

evaluación de resultados y respuestas dados por buscadores (4,2 puntos) y servicios de P/R (3,9 puntos).

De la evaluación se desprende que los resultados obtenidos de Google y Yahoo! son mejores que los de Live (5; 5,2 y 3,3 respectivamente); y que en el caso de los servicios de P/R, las respuestas dadas en LinkedIn (sin las de tipo unirespuesta) han obtenido mayor puntuación (4,15) que las de Yahoo! Answers y Live QnA (3,7 respectivamente).

En conclusión, ninguno de los dos tipos de servicios, buscadores y P/R, han satisfecho en global las expectativas de los usuarios, es más, lo han hecho con un nivel de satisfacción igual. En general, podemos decir que las preguntas que esperan respuestas objetivas (en especial las unirespuestas) han quedado mejor contestadas que las que esperan respuestas subjetivas, y que los resultados de Google y Yahoo! han sido mejor valorados que los de Live, quedando éste por debajo de las puntuaciones recibidas por Yahoo! Answers y por Live QnA. El orden de mayor a menor en cuanto a puntuación recibida por estos 8 usuarios para esta batería de 12 preguntas es el que muestra la figura 6. Estos resultados son meramente orientativos, pues con otras preguntas y otros usuarios podrían producirse cambios en el ranking que ahora tenemos.

Figura 6. Puntuación obtenida para los 6 servicios estudiados



6. Trabajo futuro

A partir de los datos recogidos se va a continuar trabajando en varias líneas:

- Estudio acerca de la forma en que los usuarios transforman una necesidad presentada en forma de lenguaje natural en una consulta para buscadores, por ejemplo la frecuencia de uso de limitadores en las consultas, como las comillas, signos + y -, and/or/not, uso de paréntesis, etc.

- Estudio de la coincidencia que se da en las palabras usadas en las consultas realizadas por los usuarios.
- Estudio con usuarios de otros perfiles interesantes para la batería de preguntas con la que estamos trabajando. Estos perfiles son los profesionales de áreas Sociales (periodismo, diseño de interacción, sociología, etc.) y de áreas informáticas.
- Estudio con otras baterías de preguntas de otra temática.
- Estudio con otros buscadores y servicios de P/R

7. Bibliografía

- BAEZA-YATES, Ricardo; GONZÁLEZ-CARO, Cristina; CALDERÓN-BENAVIDES, Liliana. The intention behind web queries. *SPIRE 2006, 11-13 October, Glasgow, Scotland*, LNCS 4209, p. 98-109.
- BAEZA-YATES, Ricardo; MARCOS, Mari-Carmen (2007). Las distintas caras de la Web. En: Schultz, Margarita (ed.). *El factor humano en la cibercultura*. Buenos Aires: Alfagrama, 2007, p. 65-114.
- COX, Anthony; FISHER, Maryanne (2004). Expectation as a Mediator of User Satisfaction. *WWW 2004 Conference workshop (New York, May 18, 2004)*, http://users.cs.dal.ca/~watters/www2004WorkShop/coxfisher_paper.pdf
- JANSEN, Bernard J.; SPINK, Amanda; SARACEVIC, Tefko (2000). Real life, real users, and real needs: a study and analysis of user queries on the web, *Information Processing and Management: an International Journal*, v. 36, núm. 2, p. 207-227.
- LEE, Uichin; LIU, Zhenyu; CHO, Junghoo (2005). Automatic identification of user goals in Web search. *Proceedings of the 14th international conference on World Wide Web, May 10-14, 2005, Chiba, Japan*.
- MARCOS, Mari-Carmen (2008). La cara social de la Web: la Web 2.0. En: Codina, Lluís; Rovira, Cristòfol; Marcos, Mari-Carmen (Eds.). *Web semántica*. Gijón: Trea, 2008. En prensa.
- ROSE, Daniel E.; LEVINSON, Danny (2004). Understanding user goals in web search. *Proceedings of the 13th international conference on World Wide Web, May 17-20, 2004, New York, NY, USA*.