

RECUPERACIÓN DE INFORMACIÓN EN INTERNET

Máster Universitario en Lengua y Usos Profesionales

Facultad de Filología, Universidad de A Coruña

Curso 2012/2013

TRABAJO PRÁCTICO

El siguiente trabajo práctico consta de tres ejercicios que se deben resolver utilizando las técnicas de recuperación de información que hemos visto a lo largo del curso.

El trabajo es individual. Debe elaborarse una breve memoria explicativa de los pasos seguidos para resolver cada ejercicio (aprox. 2 o 3 páginas para cada ejercicio). Posteriormente, debe enviarse dicha memoria a la dirección de correo electrónico del profesor Carlos Gómez Rodríguez (cgomezr@udc.es).¹ La fecha límite de entrega es el jueves 10 de enero de 2013, a las 23:59.

Los profesores de este curso revisarán los trabajos presentados, y comunicarán las calificaciones a los alumnos, reservándose el derecho de convocarlos previamente a una defensa personal, en caso de que algún aspecto del trabajo no esté expuesto de manera clara.

Por supuesto, desde hoy hasta la fecha de entrega, se puede consultar a los profesores del curso las dudas relacionadas con la realización del trabajo, o bien por correo electrónico, o bien en persona durante las horas de tutorías que tenemos reservadas para ello.

Los ejercicios son los siguientes:

1. Preparación de un viaje.

Queremos pasar una semana en Torremolinos, del 10 al 16 de diciembre de 2012². Queremos alojarnos en un camping de primera categoría o, en su defecto, en uno de segunda categoría. Queremos viajar en tren desde A Coruña hasta Málaga³ y, una vez en Málaga, queremos llegar hasta Torremolinos utilizando algún otro tipo de transporte público (autobús o tren de cercanías). Durante el tiempo que estemos alojados, para poder realizar alguna excursión, queremos alquilar un coche durante 2 o 3 días.

Averigüe si todo esto es posible. Indique aquellas búsquedas que han sido fructíferas y aquéllas que no. Haga mención explícita de aquellas páginas que permiten directamente efectuar reservas y de aquéllas que no. En caso de que necesite más información, y que la única manera de obtenerla sea ponerse en contacto con alguna persona de alguna de las entidades consultadas, puede hacerlo. Simplemente, hágalo constar y, por supuesto, no lo haga por teléfono, sino por correo electrónico (la idea es no gastar absolutamente ningún dinero en los preparativos) y procure no molestar en exceso, ni ser demasiado insistente (ya que al final se trata de un ejercicio, no de efectuar reservas reales).

¹El servidor web de la UDC puede rechazar correos provenientes de Hotmail. Se recomienda no utilizar estas direcciones para asegurar la recepción.

²Si estas fechas ya han pasado al hacer la práctica, se pueden cambiar por otra semana diferente, por ejemplo 11 a 17 de marzo de 2013. Se aconseja que la semana sea del año en curso, ya que de lo contrario puede haber dificultades para buscar billetes (es decir, algunas páginas web de trenes o autobuses pueden no mostrar billetes/horarios para 2013 hasta que empiece el año 2013).

³En el caso de que no haya trenes en el día de ida o vuelta, podemos ir el día siguiente o volver el anterior.

Si lo desea, puede cambiar el destino, las fechas, el tipo de alojamiento y el medio de transporte, por otros que le parezcan más atractivos o que le interesen más. Eso sí, intente que la *complejidad* del viaje y de las actividades que se van a realizar durante los días de estancia sea similar a la del supuesto que proponemos.

Al final, queremos también que nos transmita su opinión sobre si considera o no que Internet y las herramientas de recuperación de información son verdaderamente útiles para preparar este tipo de actividad, y queremos también que nos indique cuál es el tiempo medio que considera necesario para realizar toda la operación, desde que comienza las primeras búsquedas, hasta que confirma las reservas. Aunque puede ser muy variable, dependiendo de las fechas y del destino, este último dato le servirá para otra vez, para saber con cuánto tiempo de antelación se deben comenzar los preparativos.

2. Búsqueda de información personal.

En este ejercicio vamos a comprobar hasta qué punto las técnicas de recuperación de información en Internet se pueden emplear para encontrar información sobre personas individuales.

Para ello, tomaremos como referencia la siguiente lista de admitidos y excluidos:

http://www.iuii.ua.es/master_ti/Listadofinaladmitidos_junio.pdf

Se pueden encontrar numerosas listas similares en la Red, y hemos escogido ésta de forma totalmente aleatoria.

Escoja tres de las siete primeras personas de la lista, y averigüe toda la información personal sobre ellas que pueda. Por supuesto, debe limitarse a información que esté disponible públicamente en Internet (es decir, no es válido contactar con nadie a través de mail, msn, facebook o cualquier otro medio, ya que no se trata de molestar a nadie). Si sobre alguna persona en particular le resulta muy difícil encontrar información, puede cambiarla por otra dentro de las siete primeras.

Ejemplos de información personal que se puede intentar recopilar son: direcciones de correo electrónico, teléfonos, direcciones postales, aspecto físico/fotos de las personas, comunidades a las que pertenecen, trabajos que han llevado a cabo, asignaturas en las que se han matriculado, becas y otra información académica, participaciones en foros o blogs, gustos, o cualquier otra cosa que considere que sería relevante si quisiera elaborar un perfil de esa persona.

Para cada dato obtenido, incluya en la memoria de la práctica un enlace a la página web donde aparece, y explique la forma en que lo ha encontrado, incluyendo las consultas en buscadores en caso de utilizarlos.

Comente si buscar información sobre unas personas le ha resultado más difícil que sobre otras, y en tal caso cuáles cree que son los motivos de estas diferencias.

A continuación, puede experimentar buscando en Internet información personal sobre usted mismo/a; aunque no la incluya en la memoria.

Finalmente, queremos que nos transmita su opinión sobre la privacidad de los datos personales en Internet. ¿Cuáles son las ventajas y los inconvenientes de que esta información esté públicamente disponible en la Red? ¿Dónde cree que debería situarse el límite entre lo que se hace público y lo que no? ¿Piensa que en general deberíamos tomar precauciones para proteger nuestros datos personales en Internet, o simplemente aceptar que datos que solían considerarse privados ya no lo sean tanto?

3. Estudio comparativo del rendimiento de varios sistemas de recuperación de información.

Elija 3 buscadores. Elija una consulta relacionada con algún tema que conozca bien. Lance esa consulta en cada uno de los 3 buscadores. Fíjese en los 10 primeros resultados de cada buscador. Calcule la unión de esos documentos. Serán como mucho 30 documentos, aunque probablemente sean menos, ya que es de suponer que algún documento será recuperado por más de un buscador. Llamemos entonces D al número de documentos diferentes, siendo, como ya hemos dicho, $D \leq 30$. Imagine ahora que esos D documentos son toda la base de datos documental de la que disponemos. Acceda a cada uno de esos D documentos, y decida si es relevante o no para la consulta que ha hecho. En función de esas relevancias, calcule la precisión, el *recall* y la precisión no interpolada de cada buscador. Indíquenos qué buscadores ha utilizado, cuál ha sido la consulta, cómo son las tres listas sí/no de relevancias, cuál es el valor de cada una de las tres medidas de rendimiento, y qué conclusiones saca al respecto.

Repita de nuevo el mismo experimento, pero esta vez con una consulta relacionada también con algún tema que conozca, igual que antes, para que sea capaz de juzgar rápidamente la relevancia de cada documento recuperado, pero que sea de un tema completamente diferente. ¿Qué ha ocurrido esta vez? El buscador que más le satisface ahora, ¿es el mismo de antes o es otro? ¿Qué conclusiones saca al respecto?

Por supuesto, en cada uno de los dos experimentos, tomando inicialmente un único buscador, puede tantear con la consulta e intentar refinarla hasta que el buscador ofrezca resultados interesantes. Pero eso sí, una vez que encuentre la consulta que le gusta, láncela tal cual en cada uno de los tres buscadores, para que todos ellos compitan al mismo nivel y ninguno se vea perjudicado respecto a los demás.

Miguel A. Alonso Pardo (alonso@udc.es)
Carlos Gómez Rodríguez (cgomezr@udc.es)
Jesús Vilares Ferro (jvilar@udc.es)