

Trabajo número 2
Interfaces adaptativas para aplicaciones web:
*Acquisition, updating and management of user
models*
Interacción Persona-Ordenador

Alejandro Bellogín Kouki
alejandro.bellogin@uam.es

3 de diciembre de 2007

1. Introducción

Actualmente la cantidad de información accesible a través de distintos medios (sobre todo, internet) por los usuarios es inmensa, y aumenta a diario. Esto provoca que cuando un usuario tiene una necesidad de información recurra a buscadores, la mayoría de los cuales no sabe nada del usuario salvo las palabras introducidas en la consulta.

Como una consecuencia directa de esto el usuario debe gastar más tiempo en encontrar lo que estaba buscando, siendo posible llegar hasta el límite de abandonar después de varios intentos infructuosos. Es por ello que en la Interacción Persona Ordenador se investigan distintas técnicas de evitar esto último, entre ellas encontramos las Interfaces adaptativas y los Sistemas de recomendación, áreas que necesitan, en una gran mayoría de las veces, un modelo de usuario para poder adaptar la interfaz al mismo, generar explicaciones y ser más flexibles.

Nuestra propuesta consiste en guardar información del usuario (en un perfil), donde él mismo puede configurar sus preferencias, complementado con técnicas implícitas para recuperar información sobre los gustos del usuario y acompañado por una ontología que se utilizará para encontrar documentos relacionados semánticamente con otros que haya visto el usuario previamente; en concreto, algunas técnicas utilizadas son: mantener un contexto en la sesión, analizar los clicks y tiempos que consume el usuario en cada parte de la aplicación, y la posibilidad de que el usuario diferencie entre los items devueltos (favorito, subrayado, votación) todo ello consiguiendo resultados en la propia sesión.

De esta manera, conseguimos motivar la creación de perfiles de usuario con la mínima participación del mismo (aunque permitiendo modificaciones por su parte), incrementando las probabilidades de éxito de que la información devuelta por la aplicación sea de interés por el usuario y en un tiempo de respuesta mínimo.

2. Trabajo relacionado

Los perfiles de usuario se empezaron a utilizar hace más de veinte años[13], en particular se emplearon para mejorar los resultados en recuperación de información para satisfacer diferentes carencias[18], como por ejemplo:

- Diferenciar consultas ambiguas (por ejemplo: *kiwi*).
- Devolver resultados más apropiados a las preferencias del usuario.
- Crear resúmenes personalizados[9].

En definitiva: adaptar la presentación de la información al usuario, tarea compartida con la rama de la Interacción Persona Ordenador y utilizado ampliamente en *e-learning*[3][16][1][2].

Para solucionar algunos de los problemas mencionados, una primera aproximación habitual es utilizar una capa semántica intermedia, por ejemplo la base de datos WORDNET[10]. Nosotros usaremos una ontología[11], lo cual nos permite ser tan concretos como queramos, dependiendo del dominio de la aplicación, y teniendo a nuestra disposición relaciones semánticas no limitadas a las de sinonimia y antonimia.

Al igual que [22] guardamos y analizamos las interacciones del usuario con la aplicación, pero a diferencia de éste, los efectos del análisis (semántico) tienen lugar en la misma sesión. Por otro lado, [9] también utiliza una ontología, pero trata al perfil de usuario como un vector de conceptos con un peso, considerando el historial de búsqueda como los elementos visitados y siendo el peso el tiempo gastado en ver cada elemento, posteriormente se utiliza la ontología para realizar clústers de objetos; nuestro perfil almacena información extra concerniente a la edad, lugar de nacimiento, trabajo, etc del usuario.

Por nuestra parte, la capa semántica nos permitirá obtener conceptos relacionados con los ya existentes en el perfil o con los que el usuario ya hubiera expresado su preferencia, es decir, nos permitirá ofrecer objetos nuevos, pero con una cierta seguridad de que serán del agrado del usuario, gracias a una definición precisa de la semántica involucrada; esto se consigue mediante la expansión de conceptos, que consiste en que dados unos conceptos y a partir de las relaciones existentes en la ontología, conseguimos conceptos nuevos que se utilizarán para afinar la personalización. Además, nos permitirá realizar inferencia y explicar al usuario por qué se le han recomendado determinados resultados o elementos.

Con respecto a la introducción del contexto del usuario para personalizar los resultados devueltos, [19] comparte con nosotros que el resultado de la personalización es inmediato, no obstante, existen diferencias evidentes, entre otras: sólo utiliza la última consulta y siempre empieza desde cero al no guardar un perfil de usuario. Además, nos basaremos en [15] y [21] donde se define el contexto tal y como lo usaremos nosotros, junto con la activación y expansión del mismo, es decir, el contexto semántico en tiempo de ejecución es la combinación acumulativa de los conceptos envueltos en las sucesivas consultas o peticiones del usuario, expandiéndose con conceptos relacionados (dadas estas relaciones por la ontología).

3. Propuesta

El núcleo de nuestra propuesta consiste en:

- Perfil de usuario, donde se guarda información del usuario concreto, implícita o explícitamente recogida.
- Ontología de dominio, ofrece los conceptos sobre los que se definirán las preferencias del usuario y que permitirá expandir a otros relacionados e inferir nuevos datos.

Estos dos elementos se pueden definir de manera general para cualquier ámbito, a pesar de ello, en aras de la simplicidad los aplicaremos a un ámbito concreto: el de noticias, lo que nos permitirá ser más precisos en cuanto a las herramientas o técnicas a utilizar para capturar la información del usuario. De este modo, el perfil de usuario almacenará información demográfica y laboral del usuario, junto con los elementos (noticias) que le hayan parecido interesantes y sus conceptos asociados. También contendrá los conceptos deducidos por el sistema gracias al contexto y a técnicas implícitas, lo cual nos permitirá explicar al usuario por qué se le han mostrado determinadas noticias.

Llegados a este punto hay que destacar un aspecto importante y del que también se hace eco la literatura[18]: la seguridad. Nosotros hemos optado por mantener los datos del usuario en el servidor (en cada sesión se mandan al cliente, pero de manera asíncrona se actualizan en el servidor) ya que esto permitirá al usuario conectarse desde distintas terminales, dándole mayor facilidad de uso y más flexibilidad, por otra parte, consideramos que proveer seguridad al servidor es una tarea más sencilla y razonable que permitir que sea el usuario el que proteja sus datos[14][20], además de que con un servidor dedicado se podrán calcular más rápidamente las recomendaciones y/o personalizaciones que en el ordenador del cliente.

Con respecto al modelo de usuario almacenado en el perfil, nos interesa guardar sus datos personales (demográfico, laboral, estudios) y sus preferencias, explícitamente definidas. No obstante, dado que el usuario puede producir inconsistencias con respecto a las votaciones, perder mucho tiempo[7], no ser coherente con respecto a algún formulario rellenado con anterioridad al compararlo con su comportamiento real[8] e incluso llegar a considerar algunas preguntas como privadas y rechazar su contestación[9], hemos combinado el método explícito con uno no invasivo (implícito). De esta manera, registramos y analizamos todas las acciones entre el usuario y la aplicación: en qué noticias hace click, qué cantidad de tiempo está leyendo cada noticia (o parte de ella), cuáles vota o marca como favoritas y qué conceptos de cada noticia le parecen más interesantes (subrayado). Con estos datos se va construyendo un contexto de usuario, que se verá influido por las preferencias almacenadas en el perfil y las creadas recientemente debido al análisis implícito de las acciones realizadas (por defecto tendrán más peso las preferencias más recientes, aunque se podrá modificar en la aplicación). Este contexto se puede utilizar en la misma sesión para personalizar las búsquedas o recomendar noticias nuevas.

No obstante, hay que tener en cuenta que los datos relativos al historial de clicks no pueden ser un juicio de relevancia absoluto[12], sino que dan una valoración relativa entre los documentos vistos por el usuario y los que podría llegar a haber visto (si nunca llegó a la página 4 de resultados, sobre los documentos

que ahí aparecen no podemos deducir nada). Además, el tiempo empleado en cada noticia hay que normalizarlo según el tamaño de la página[7].

Por último, como característica novedosa se permiten subrayar los conceptos (o frases) que más le han interesado al usuario e incluir el resultado obtenido como favorito con el objetivo de obtener del usuario más información pero sin exigir demasiada carga cognitiva.

Con todo lo anteriormente descrito, se llegan a las siguientes fórmulas:

$$\begin{aligned}
 S_P(D_i) &= T_P(D_i) \times \left(F_P(D_i) + R_P(D_i) + \alpha U_P(D_i) + \beta \sum_{c \in D_i} s_P(c) \right) \\
 F_P(D_i) &= \phi \text{ si el usuario ha marcado } D_i \text{ como favorito, } 0 \text{ en otro caso} \\
 R_P(D_i) &= r_P(D_i) \in [0, \psi] \text{ voto del usuario, } 0 \text{ si no ha votado} \\
 U_P(D_i) &= \sum_{c \in u(D_i)} s_P(c), \text{ siendo } u(D_i) \text{ los términos subrayados en } D_i \\
 T_P(D_i) &= \frac{\text{duracion}_P(D_i)/\text{tamaño}(D_i)}{\text{máx}_{D_i \in \text{docsVistos}_P} \{ \text{duracion}_P(D_i)/\text{tamaño}(D_i) \}}
 \end{aligned}$$

que hacen referencia a las puntuaciones (S) de un usuario o perfil determinado (P), donde $\phi > \psi > 0$, $\alpha, \beta \geq 0$ parámetros a determinar, D_i es un documento y $s_P(c)$ indica la preferencia del usuario al que se refiere el perfil P por el concepto c , para lo que se tienen en cuenta las preferencias del usuario, el contexto y las correspondientes expansiones mediante la ontología de dominio (depende del algoritmo empleado para personalizar/recomendar). Hay que tener en cuenta que esta fórmula se calcula para todos los documentos que se pueden mostrar en un determinado momento, no obstante, si el usuario ha votado o marcado como favorito un documento, bastaría con recalcular la puntuación para ese documento y volver a ordenar según los nuevos datos.

Por lo tanto, la máxima puntuación de un documento será la de aquel que haya sido marcado como favorito, con la mayor valoración al votarlo (ψ) y en el que el usuario tenga mucha preferencia tanto por los conceptos relacionados con el documento como por los conceptos subrayados por el propio usuario y haya seleccionado documentos parecidos recientemente (contexto); como además se tiene en cuenta el tiempo empleado en su visualización, cuanto mayor sea este, mayor será su puntuación.

Para encontrar los parámetros óptimos necesitaríamos realizar un estudio con usuarios reales, no obstante, nuestra intención es que el usuario pueda cambiarlos de una manera sencilla e intuitiva y quedarse con los valores que más le interesen, o, en última instancia, que dependa del tipo de aplicación.

4. Examen y evaluación

Para poder evaluar la validez de esta propuesta habría que complementarlo con distintos algoritmos de personalización y recomendación, ya que nos hemos centrado en el perfil de usuario y su generación, y comprobar si mejora el rendimiento de alguno de ellos o es más conveniente para un determinado (tipo de) algoritmo. Algunos ejemplos de algoritmos en los que se puede aplicar son los basados en contenido, basado en objetos (*items*), filtrado colaborativo, ...

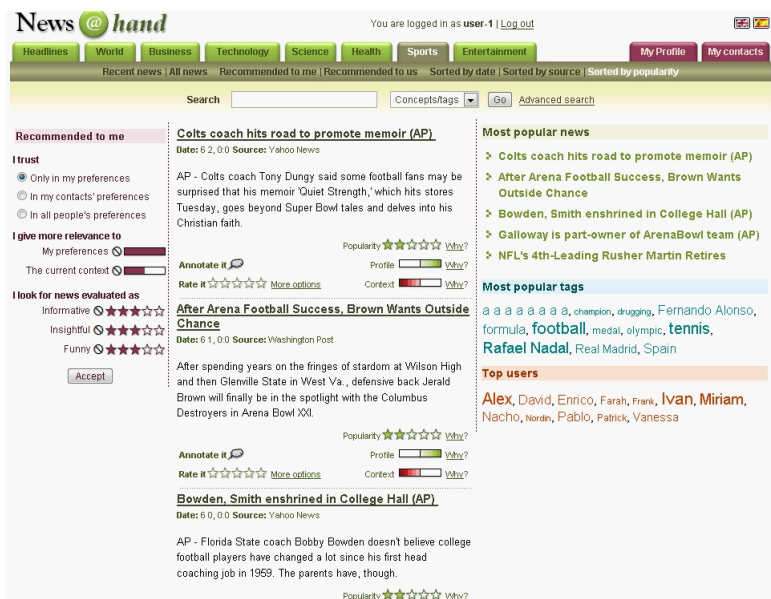


Figura 1: Captura de un sistema de recomendación que utiliza el modelo de usuario presentado

No obstante, la utilización de la historia pasada y el análisis de los clicks del usuario ha demostrado ser muy eficaz para mejorar los resultados en las búsquedas[17][12], por lo que esperamos que nuestra propuesta aplicada a sistemas concretos dé resultados positivos.

En particular, una parte de la propuesta presentada ha sido ya probada, integrada en un algoritmo multicapa de recomendación híbrido (utiliza filtrado colaborativo y basado en contenido) basado en ontologías[5]. En este momento, estamos trabajando en un sistema de recomendación de noticias que utiliza por completo esta propuesta, una captura de dicho sistema se puede ver en la figura 1.

Una vez esté el sistema terminado y en funcionamiento, para poder evaluarlo se necesitarán usuarios reales, comparando la satisfacción del mismo utilizando el sistema con y sin personalización, activando o no el contexto, etc. Para ello se les pueden asignar tareas concretas para medir la satisfacción a corto plazo o dar libertad al usuario y realizar encuestas al cabo de un tiempo para averiguar su rendimiento a largo plazo.

5. Conclusiones y trabajo futuro

Se ha presentado un modelo de usuario orientado a la personalización y recomendación de información, independiente del dominio gracias a la capa semántica intermedia utilizada (ontología) que permite dicha independencia.

En dicho modelo se tienen en cuenta las preferencias explícitas del usuario, pero aún más importantes son las implícitas (utilización de contexto, análisis de clicks, votación, favoritos), las cuales disminuyen la carga cognitiva del usuario y le permiten obtener resultados sin tener que haberse esforzado en expresarlas

de alguna determinada manera.

También se ha presentado un sistema sobre el que se probará dicho modelo de usuario junto con un algoritmo de recomendación, aunque distintos resultados previos nos permiten aventurar que se obtendrán resultados positivos.

Una mejora sobre la presente propuesta consistiría en distinguir situaciones de interés, dado que los usuarios pueden cambiar rápidamente de intereses dentro de un sesión, por lo que la primera tarea sería detectar un cambio de situación y dejar de recomendar noticias válidas para la situación anterior.

Con el objetivo de facilitar la tarea al usuario podrían recogerse preferencias por medio de tags[6], sustituyendo la ontología de dominio por una más ligera[4], reduciendo asimismo el esfuerzo de anotación de los documentos involucrados en el sistema.

Además, otro origen a partir del cual se podrían captar preferencias del usuario podrían ser los *bookmarks* o favoritos del explorador, de manera que los elementos (en nuestro caso, noticias) que contengan conceptos similares a los de los elementos incluidos en dichos favoritos serán propensos a recomendarse, incluyendo esos conceptos en el perfil.

Otra medida que podría incluirse a la hora de generar la puntuación de un documento sería dependiendo de en qué lugar aparece el concepto, es decir, que el sistema debería distinguir si aparece en el título, en el cuerpo, etc y ponderarlo en consecuencia.

Por último, el algoritmo de contextualización en que nos basamos es capaz de tener en cuenta el desinterés de un usuario por un elemento, por lo que sería útil poder utilizar esta característica, para esto el modelo debería tratar de manera distinta cuándo el usuario dice que un objeto no le interesa y almacenarlo de manera distinta que en el caso contrario (posiblemente bastaría admitiendo pesos negativos).

Referencias

- [1] Enrique Alfonseca, Rosa M. Carro, Estefanía Martín, Alvaro Ortigosa, and Pedro Paredes. The impact of learning styles on student grouping for collaborative learning: A case study. *User Modeling and User-Adapted Interaction. Special Issue: User Modelling to Support Groups, Communities and Collaboration*, 16(3-4):377–401, September 2006.
- [2] Terry Anderson and Denise Whitelock. The educational semantic web: Visioning and practicing the future of education. *Journal of Interactive Media In Education*, 1, May 2004.
- [3] Lora Aroyo, Ronald Denaux, Vania Dimitrova, and Michael Pye. Interactive ontology-based user knowledge acquisition: A case study. pages 560–574. 2006.
- [4] Scott Bateman, Chris Brooks, and Gord Mccalla. Collaborative tagging approaches for ontological metadata in adaptive e-learning systems, 2006.
- [5] Iván Cantador, Pablo Castells, and Alejandro Bellogín. Modelling ontology-based multilayered communities of interest for hybrid recommendations, June 2007.

- [6] Francesca Carmagnola, Federica Cena, Omar Cortassa, Cristina Gena, and Ilaria Torre. Towards a tag-based user model: How can user model benefit from tags? pages 445–449, 2007.
- [7] Philip K. Chan. A non-invasive learning approach to building web user profiles, 1999.
- [8] Kay Connelly, Ashraf Khalil, and Yong Liu. Do i do what i say?: Observed versus stated privacy preferences. pages 620–623. 2007.
- [9] H. Dai and B. Mobasher. Using ontologies to discover domain-level web usage profiles, 2002.
- [10] Marco Degemmis, Pasquale Lops, and Giovanni Semeraro. Learning semantic user profiles from text. In *ADMA*, pages 661–672, 2006.
- [11] Thomas R. Gruber. A translation approach to portable ontology specifications. *Knowl. Acquis.*, 5(2):199–220, 1993.
- [12] Thorsten Joachims. Optimizing search engines using clickthrough data, 2002.
- [13] Alfred Kobsa. Generic user modeling systems. *User Modeling and User-Adapted Interaction*, 11(1-2):49–63, 2001.
- [14] Rebecca T. Mercuri. Computer security: quality rather than quantity. *Commun. ACM*, 45(10):11–14, 2002.
- [15] Phivos Mylonas, David Vallet, Pablo Castells, Miriam Fernandez, and Yannis Avrithis. Personalized information retrieval based on context and ontological knowledge. 2008.
- [16] Pedro Paredes and Pilar Rodriguez. A mixed approach to modelling learning styles in adaptive educational hypermedia. *Advanced Technology for Learning*, 1(4):210–214, 2004.
- [17] Xuehua Shen, Bin Tan, and Chengxiang Zhai. Context-sensitive information retrieval using implicit feedback. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 43–50, New York, NY, USA, 2005. ACM Press.
- [18] Xuehua Shen, Bin Tan, and Chengxiang Zhai. Implicit user modeling for personalized search. In *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 824–831, New York, NY, USA, 2005. ACM Press.
- [19] Xuehua Shen, Bin Tan, and Chengxiang Zhai. Ucair: Capturing and exploiting context for personalized search, 2005.
- [20] Richard M. Stallman. *Free Software, Free Society: Selected Essays of Richard M. Stallman*. Gnu Press, 2002.

- [21] David Vallet, Pablo Castells, Miriam Fernandez, Phivos Mylonas, and Yannis Avrithis. Personalized content retrieval in context using ontological knowledge. *Circuits and Systems for Video Technology, IEEE Transactions on*, 17(3):336–346, March 2007.
- [22] Hui Zhang, Yu Song, and Han-Tao Song. Construction of ontology-based user model for web personalization. In *User Modeling*, pages 67–76, 2007.