

2009-30

Anexo C. Trabajos relacionados  
con el perfil de usuario, perfil de  
contexto, reescritura de  
consultas y enriquecimiento de  
consultas

Fernando Antonio Aragón

Maria Claudia Higuera

Versión 1.0

## Tabla de contenido

TRABAJOS RELACIONADOS .....	3
PERFIL CONTEXTUAL .....	3
PERFIL DE USUARIO.....	7
REESCRITURA DE CONSULTAS.....	9
ENRIQUECIMIENTO DE CONSULTAS .....	12
Referencias .....	18

## Lista de Tablas

Tabla 1 Comparación de artículos de contexto.....	6
Tabla 2 Comparación trabajos perfil de usuario .....	9
Tabla 3Tabla de correspondencia, adaptado de [2].....	10
Tabla 4 Reescritura a partir de la Tabla 5 de correspondencia, adaptado de [7] .....	10
Tabla 5 Comparación de trabajos de reescritura de consultas .....	12
Tabla 6 Comparación trabajos relacionados enriquecimiento. ....	17

## Lista de figuras

Figura 1. Definición general de reglas, tomado de [7] .....	11
Figura 2. Enriquecimiento de consultas por medio de perfil de usuario, tomado de [30] .....	14
Figura 3. Sistema de filtrado, adaptado de [26] .....	16

## TRABAJOS RELACIONADOS

A continuación se presenta un resumen de los trabajos relacionados, que se investigaron para el desarrollo del sistema de enriquecimiento de consultas.

### PERFIL CONTEXTUAL

El contexto de uso hace referencia a las características de la interacción del usuario con el sistema. Dey *et al.* [13] define el contexto como “*todos los elementos de información que pueden ser utilizados para caracterizar la situación de una entidad, siendo una entidad toda persona, lugar u objetos considerada pertinente para la interacción entre el usuario y la aplicación*”.

A su vez, Kirsch-Pinheiro *et al.*[17] sugieren que el contexto está compuesto de “información física y conceptual que proviene de recursos que van más allá de lo que se ve a simple vista, como recursos digitales, conceptos, estados mentales, tareas, relaciones sociales y cultura local; así mismo, si se cuenta con trabajos cooperativos es necesario tener en cuenta la comunidad”. Este perfil contextual, es propuesto para aplicaciones en las cuales se realiza trabajo cooperativo, es decir, que cada usuario cuenta con unas tareas asignadas dependiendo de su rol en la comunidad y a su vez necesita que algunas personas le ayuden con su trabajo (relaciones sociales). Sin embargo, este modelo no considera las restricciones, qué por ejemplo, podrían determinar información que no debería ser mostrada a algunos usuarios. Así mismo, las actividades no cuentan con periodos de vigencia ni periodicidad. Si estos datos se incluyeran se podría determinar cuándo es el plazo máximo en que se debe iniciar y terminar una tarea asignada, así como cada cuánto se repite. Por ejemplo, se puede determinar que todos los lunes se debe trabajar sobre la documentación de un código, que tiene un periodo de vigencia que finaliza el 2 de junio, y que para ese entonces debe estar finalizada dicha tarea.

Ad-me (acrónimo de Advertising for the mobile e-commerce user) [15] es un sistema que brinda al usuario información turística relacionada con su contexto y su perfil de usuario. El contexto tenido en cuenta en Ad-me busca contestar unas preguntas: ¿Dónde estoy?, ¿Dónde está el objeto deseado? y ¿Cómo puedo alcanzar el objeto deseado? A partir de dichas preguntas Ad-me considera un contexto que se divide en: *i)* Contexto del usuario que contiene identificación, localización y relaciones con el perfil de usuario, *ii)* contexto computacional relacionado con conectividad a red, ancho de banda y características del dispositivo, *iii)* contexto físico que corresponde a características del entorno y los objetos

cercanos al usuario y, *iv*) contexto histórico que se compone de datos consignados en un histórico, por ejemplo, las fechas en que iniciaron y terminaron las estaciones en un año específico. Sin embargo, Ad-me no considera características como la privacidad, es decir, a qué sitios el usuario prohíbe o restringe el uso de su información, ni restricciones. Por ejemplo, si el usuario desea obtener información sobre algún lugar turístico, pero no desea que se conozca su ubicación exacta, se debería dar la posibilidad de mantener oculta dicha información. Además, no se contempla la definición de actividades, lo que ayudaría a hacer un estimado de qué objetos busca alcanzar el usuario en determinados lapsos del día con respecto al quehacer durante los mismos.

Los sistemas basados en la localización como el propuesto por Burak *et al.* [5] y el sistema propuesto por Chen [7] centran sus estudios en la localización del usuario, y en algunos casos se tienen en cuenta características de privacidad de datos y formas de adaptación al despliegue de la información [5] pero a su vez descuidan otros aspectos que podrían ser de gran importancia en estos sistemas, como las restricciones legales dependiendo del lugar donde se encuentra el usuario, sus gustos y las actividades que se están realizando.

AmbieAgents [18] muestra una guía turística que cuenta con la personalización como un valor agregado. Esta personalización se realiza dependiendo de las características del usuario y del contexto de uso. En este caso se consideran los siguientes aspectos como perfil contextual: *i*) Contexto social, *ii*) Tareas, *iii*) Ambiente, *iv*) Características personales, *v*) Espacio temporal y *vi*) Privacidad. No obstante, no se tienen en cuenta restricciones legales ni características estructurales del lugar en el que se encuentra el usuario, lo que traería como beneficio mejores mecanismos de búsqueda de información turística (para los casos en los cuales la parte estructural pueda limitar la conectividad), y, por ejemplo, la exclusión de los resultados de los destinos turísticos a los cuales el usuario no tiene acceso (*e.g.* si el usuario no cuenta con la visa para Estados Unidos, no podría visitar este país).

El trabajo de Neff *et al.* [23], describe un sistema que presenta a los usuarios ayuda sobre las aplicaciones de forma auditiva, y no como texto. Este sistema considera un contexto conformado por las características de las aplicaciones que están siendo usadas en un momento específico, es decir, las ventanas activas y el sistema operativo en uso. También se consideran las partes del cuerpo que el usuario utiliza para interactuar adecuadamente con la aplicación (ojos, oídos, manos). Se considera que al presentar la ayuda en forma de texto (normalmente en una ventana nueva), se modifica el contexto puesto que la ventana del programa en el que está trabajando el usuario es cambiada por la ventana de ayuda, lo que hace que el usuario se desubique y pierda tiempo en volver a sus labores, disminuyendo así

su eficiencia de trabajo. Para evitar este cambio de contexto se propone un sistema que presenta ayudas en forma auditiva, para que el usuario mantenga el contexto de trabajo y pueda recibir la información de la ayuda por otro canal que no se esté usando con la aplicación (el auditivo en este caso), pudiendo seguir trabajando con los ojos y con las manos. No obstante, este trabajo no contempla otras características que podrían influir en el despliegue de información como son las características espacio temporales, las actividades y las restricciones, que podrían establecer pautas sobre si se puede mostrar o no la ayuda de forma auditiva, o se opta por la presentación de la ayuda en forma textual (*e.g.* en una reunión, el usuario no tendrá audífonos y no querrá que toda la reunión escuche lo que consulta en la ayuda, o en otro caso, la actividad que realiza puede dar tiempo a leer la ayuda con detenimiento para comprenderla mejor).

Gandon *et al.*[11] proponen una arquitectura para la web semántica, llamada *e-wallet*. Esta arquitectura busca ser utilizada en aplicaciones sensibles al contexto, dado que soporta el almacenamiento (por medio de perfiles) y uso de información contextual de los usuarios. El contexto incluido en *e-wallet* está compuesto por: *i*) un componente estático, que contiene la información básica de los usuarios (*e.g.* nombre, identificación, correo electrónico); *ii*) un componente dinámico, el cual contiene las preferencias que varían de acuerdo a la actividad que se esté realizando; y *iii*) las preferencias de privacidad, que describen qué información puede ser usada en diferentes condiciones. *E-wallet* tiene como prioridad el componente de preferencias de privacidad. Este componente es el encargado de realizar la comunicación con otras aplicaciones o sistemas de información; para enviar mensajes, primero se verifica si se cumple con las reglas de privacidad establecidas; de ser así se envía el mensaje, de lo contrario se negará el envío del mismo. Esta arquitectura fue probada en una aplicación que usa conexión inalámbrica. Sin embargo, este trabajo no incluye características estructurales, que podrían ser importantes a la hora de contar con una conexión inalámbrica, puesto que dependiendo de la cantidad de paredes u obstáculos, la señal podría disminuir y por esto, la conectividad se vería limitada. Por otra parte, tampoco se cuenta con características climáticas, que podrían influir tanto en las actividades a realizar como en las preferencias de las mismas; por ejemplo si está lloviendo, es probable que un usuario prefiera comer en la casa que buscar un restaurante.

Por otra parte, Bai *et al* [3] proponen un modelo, en el cual se busca expandir la consulta inicial de un usuario; para dicha expansión, se realiza una asociación de términos. La propuesta dada por Bai *et al.* consiste en realizar dicha asociación a partir de un conjunto de términos. Por ejemplo, si una consulta contiene el conjunto de términos “Java, **Programación**”, y estos dos términos están relacionados con el tema “Computación”,

entonces al realizar la búsqueda de “Java, **Programación**”, la consulta se enriquece agregando el término “Computación”. Pero si se tiene el conjunto de términos “Televisión, **Programación**”, vale destacar que dicho conjunto NO está relacionado con el tema “Computación”. Por tal motivo, las relaciones entre términos evitan una posible ambigüedad en la utilización de ciertos términos (para este ejemplo Programación) utilizados en varios conjuntos establecidos. De esta manera se evitaría realizar el mismo enriquecimiento para dos consultas semánticamente diferentes. A esta asociación se le conoce como el *contexto dentro de la consulta*. Por su parte, este modelo no tiene en cuenta características como la actividad que se está llevando a cabo al momento de realizar la consulta, con lo cual, los términos relacionados podrían variar dependiendo de las preferencias de los usuarios (*e.g.* un programador puede estar interesado en el término Java como lenguaje de programación, cuando se encuentra realizando actividades laborales; mientras que ese mismo usuario en actividades de ocio puede consultar la palabra Java con el fin de obtener resultados acordes a la variedad de café); tampoco se tienen en cuenta restricciones, que podrían influir en algunos términos que no se puedan agrupar, por ejemplo si se restringe la palabra violencia, este término no se debería agregar a ninguna consulta.

Después de expuestos los trabajos relacionados, se presenta una tabla de comparación entre los datos tenidos en cuenta en cada uno de los perfiles contextuales para cada trabajo. Los trabajos relacionados se encuentran enunciados en las filas; en las columnas se muestran los diferentes aspectos a comparar. A su vez se cuenta con las siguientes conversiones: *i)* el carácter “+” indica que el aspecto nombrado en la fila, se encuentra contemplado en el trabajo de la columna, *ii)* el carácter “-” simboliza que el aspecto no está considerado en el trabajo y por último *iii)* el carácter “?” alude a que el componente es nombrado, pero no se especifica cómo es tratado.

	[17]	[15]	[5]	[7]	[11]	[4]
<b>Espacio temporal</b>	+	+	+	+	?	?
<b>Actividad</b>	?	-	+	-	+	?
<b>Cultural</b>	+	-	-	-	?	-
<b>Restricciones</b>	-	+	-	-	?	-
<b>Privacidad</b>	-	+	+	-	+	-
<b>Estructural</b>	-	+	-	-	-	-
<b>Términos relacionados</b>	-	-	-	-	-	+

Tabla 1 Comparación de artículos de contexto

De la Tabla 1 se puede concluir que los trabajos relacionados contemplan en su gran mayoría las características espacio temporales; a su vez se tienen en cuenta datos sobre la privacidad de la información a utilizar. Sin embargo, ninguno de los trabajos cumple con varios aspectos importantes a la hora de utilizar el contexto, como por ejemplo, las restricciones. Tal componente sólo se encuentra en Ad-me [15], al igual que las características estructurales del lugar donde se encuentra el usuario. Este último es de gran importancia para las aplicaciones que utilicen conexiones inalámbricas, puesto que éstas se podrían ver afectadas por la parte estructural.

## PERFIL DE USUARIO

A continuación se presentarán algunos trabajos que proponen o usan un perfil de usuario como modelo de adaptación al contenido.

El trabajo de Berkovsky [4], propone un mecanismo para unir perfiles de usuario generados a partir de modelos distintos. Estos pueden haber sido generados en servidores, a partir de los datos históricos de la interacción del usuario con el sistema, o aquellos que se han almacenado en los dispositivos de acceso de los usuarios. Berkovsky afirma que cada uno de estos perfiles es parcial, y por eso proponen un sistema que se encarga de unificarlos, se busca que el perfil resultante de dicha unión considere más características que cada uno de los perfiles parciales con el fin de tener una información más completa del usuario. Para esto, propone un mediador entre los dos sistemas, que conozca ambos modelos y pueda resolver ambigüedades mediante interpretaciones de lo que significa cada ítem en cada modelo. Para ejemplificar se presenta un perfil procedente de una tienda de libros y otros de una tienda de *DVD*; a partir de la unión de estos perfiles, se podrían inferir las películas que un usuario puede estar interesado en comprar basándose en los libros leídos, bien sea porque a partir de dichos libros se han creado películas o porque el género de los libros es el mismo de las películas que ha comprado o alquilado. Sin embargo, y aunque el resultado de relacionar los perfiles entre sí da como resultado un perfil que considera varias características (las contempladas en cada perfil), este trabajo contempla solamente formas de establecer relaciones entre fuentes distintas de información y sigue siendo un problema las posibles ambigüedades que se podrían presentar (*e.g.* considerar el mismo componente con nombres distintos “ubicación” y “localización” o que dos componente iguales, estén definidos con diferentes tipos de datos o diferente con valores) debido a que los perfiles de usuario son utilizados en aplicaciones distintas. Tampoco se considera la existencia de perfiles diferentes al de usuario, como el perfil de contexto, o de despliegue, con los cuales se podría contar con un modelo de adaptación más completo y genérico.

Otro perfil de usuario es el propuesto por Wang *et al.* [28], quienes definen un sistema que busca entregar a los visitantes de un museo rutas personalizadas, basándose en el conocimiento que tienen sobre las obras de arte y sus preferencias. Para realizar estas rutas, se le presentan al usuario obras de arte con características que deberá seleccionar como de su preferencia o no y una encuesta de conocimiento artístico, luego se realiza un cuestionario de conocimiento sobre obras de arte. A partir de esta información los usuarios se categorizan en: *i) novatos ó ii) expertos*. Luego se les muestra la ruta sugerida, realizada con base en las preferencias del usuario y la categoría asociada. Después de que el usuario ha realizado la ruta se vuelve a aplicar la encuesta y se observa un aumento de las respuestas acertadas. Los modelos se definen utilizando ontologías representadas en *OWL*. No obstante, Wang *et al* [28] modelan a los usuarios con respecto a sus preferencias artísticas, y no tienen en cuenta otras características del usuario como su profesión, sus pasatiempos o incluso datos personales que permitirían refinar sus preferencias artísticas y hacer más amplio el rango de aplicaciones para las cuales se puede usar dicho perfil .

Michalski *et al.* [22] proponen la creación de un perfil de usuario a partir de las interacciones que éste tenga con el sistema, las aplicaciones que utiliza, el tiempo durante el cual las usa y los cambios que tengan las ventanas de la aplicación (por ejemplo, título, tamaño, *etc.*). Con el fin de proveer seguridad en el acceso a equipos de cómputo, el modelo identifica usuarios particulares, lo que sirve para detectar intrusiones o impostores. Para esto, se realiza una comparación entre lo que los usuarios normalmente realizan en el sistema, y lo que están haciendo durante la sesión actual (definida como el tiempo que transcurre entre el ingreso de un usuario o *login*, hasta la desconexión o *logout*). Sin embargo, el modelo propuesto no contempla la inclusión de datos básicos de los usuarios que podrían ser validados a través de preguntas para identificar a impostores. Por ejemplo se podría preguntar: ¿Cuál es el número de su identificación? o ¿Cuál es el nombre de su mascota? Tampoco se incluyen los roles que tienen los usuarios dentro del sistema, lo que podría mejorar la seguridad, puesto que se podrían poner restricciones y/o asignar permisos dependiendo del tipo de usuario. Además, para este trabajo se podría incluir un modelo de adaptación de las actividades que el usuario realiza. Teniendo en cuenta las actividades del usuario se podría restringir el acceso a recursos o a información solamente cuando esté realizando una determinada actividad (*e.g.*, una actividad que se defina como administración del servidor).

Después de expuestos los diferentes perfiles de usuario, sus componentes y algunos trabajos relacionados, donde se utilizan dichos perfiles, se presenta una tabla de comparación entre los datos tenidos en cuenta en cada uno de los perfiles de usuario. Los trabajos relacionados se encuentran enunciados en las filas; en las columnas se muestran los diferentes aspectos a

comparar. A su vez se cuenta con las mismas conversiones utilizadas para la tabla de perfil contextual.

	[6]	[24]	[14]	[8]	[28]	[20]	[4]	[22]
<b>Localización</b>	+	?	?	-	-	-	?	-
<b>Preferencia actividad</b>	+	+	-	-	-	-	+	?
<b>Dispositivo de acceso</b>	+	?	-	-	?	-	?	?
<b>Preferencias</b>	?	+	-	-	+	-	+	-
<b>Productos</b>	-	+	-	+	-	-	?	-
<b>Preferencia de Resultados</b>	-	+	-	-	-	-	?	-
<b>Información relacionadas</b>	-	+	-	+	-	-	+	-
<b>Gustos</b>	?	+	-	-	+	-	+	?
<b>Información básica</b>	?	+	?	+	?	?	+	+
<b>Estado</b>	-	-	+	-	-	-	?	-
<b>Aplicación</b>	-	-	-	-	-	+	?	+
<b>Tiempo</b>	-	?	+	-	-	+	?	+
<b>Seguridad</b>	-	-	-	-	-	+	?	+

Tabla 2 Comparación trabajos perfil de usuario

Como se puede ver en la anterior tabla, los perfiles de usuario propuestos en los diversos artículos, consideran diferentes aspectos que incluyen desde la información básica del usuario (nombre, identificación, teléfono, entre otros), gustos, sus preferencias de actividad hasta preferencias de resultados (indican cómo desea el usuario recibir la información). Sin embargo la mayoría de los trabajos no incluyen características de seguridad, que podría ser definidas como preferencias de quiénes pueden acceder a cierta información o realizar ciertas funciones o solicitar servicios dentro del sistema.

## REESCRITURA DE CONSULTAS

A partir de los conceptos presentados anteriormente se procede a presentar trabajos relacionados con la reescritura de consultas.

Yang *et al.*[29], proponen un modelo de reescritura de consultas, basado en un conjunto de consultas base<sup>1</sup> y reglas lógicas. Por medio de éstas se busca inferir información y realizar

---

<sup>1</sup> Definidas como un conjunto de consultas realizadas anteriormente, y las respuestas que éstas tuvieron. Con respecto a esto se realiza la reescritura de las consultas.

asociaciones de dicha información, empleando la minería de datos, para luego añadirlas a la consulta ingresada por el usuario ( $Q$ ) y generar una consulta enriquecida ( $Q'$ ).

Así mismo, Aslam *et al.* [2] presentan un sistema de reescritura semántica, en el cual se traducen las consultas, para que sea equivalente en diferentes esquemas<sup>2</sup>. Para esto utiliza una tabla de correspondencias con un identificador y la descripción del mismo (ver Tabla 3), a cada atributo se le asigna un identificador de correspondencia, con el que se determina si es necesario realizar modificaciones a un consulta, bien sea concatenando o reemplazando los atributos (equivalente al *SELECT* en una consulta *SQL*) o modificar la entidad (equivalente al *WHERE* en una consulta *SQL*).

<b>Id</b>	<b>Entidad esquema global</b>	<b>Atributo</b>	<b>C1 Entidad esquema local</b>	<b>S1 Atributo</b>	<b>ID correspondencia</b>
01	Estudiante	Nombre	Estudiante	Nombre, Apellido	1

<b>ID Correspondencia</b>	<b>Descripción</b>
0	Iguales
1	Concatenar S1
2	Reemplazar S1
3	Modificar C1
4	No valido

Tabla 3 Tabla de correspondencia, adaptado de [2]

A partir de la tabla de correspondencias, mostrada anteriormente se procede a realizar la reescritura, tal como se muestra en Tabla 4.

<b>Consulta <math>Q</math></b>	<b>Consulta reescrita <math>Q'</math></b>
<i>SELECT</i> Nombre <i>FROM</i> Estudiante	<i>SELECT</i> Nombre, Apellido <i>FROM</i> S1.Estudiante

Tabla 4 Reescritura a partir de la Tabla 5 de correspondencia, adaptado de [7]

<sup>2</sup> Representaciones gráficas o simbólicas de alto nivel.

Por su parte, Finance *et al.* [10] formulan un lenguaje de reglas, con el fin de realizar la reescritura de consultas, tanto semántica como sintácticamente. Ellos definen, de forma general las reglas, de tal manera que si se cumplen ciertas condiciones indicadas en la parte izquierda, se realiza la reescritura de la forma indicada en la parte derecha de la misma. Se presenta un lenguaje de definición de reglas (ver Figura 1), el cual puede ser instanciado para especificar las reglas de reescritura, de diferentes tipos (*e.g.* semánticas, sintácticas). El trabajo fue utilizado para reescribir consultas en *ESQL* (extensión de *SQL*, para bases de datos relacionales, orientadas a objetos y deductivas).

```

<Rule> ::= <Term> / [ <Constraint> ] ->
           <Term> / [ <Method List> ]
<Term> ::= <Functional Expression>
           | <Complex Term>
<Functional Expression> ::=
  <Function Variable> ( <Term> [ { , <Term> ... } ] )
  | <Function Name> ( <Term> [ { , <Term> ... } ] )
<Constraint> ::= <Formula>
<Method List> ::= <Functional Expression>
                 [ { , <Functional Expression> ... } ]
<Complex Term> ::=
  <Constant>
  | <Formula>
  | <Term Variable>
  | <Collection Variable>
  | <Collection> ( <Complex Term List> )
  | TUPLE ( <Named Complex Term List> )
<Constant> ::= <Real> | <Integer> | <String>
<Formula> ::= NOT ( <Formula> )
            | ( <Boolean Expression> AND <Formula> )
            | ( <Boolean Expression> OR <Formula> )
            | <Boolean Expression>
<Boolean Expression> ::= <Functional Expression>
                       | <Predicate> | true | false
<Predicate> ::= <Num_exp> <Comparator> <Num_exp>
<Comparator> ::= = | < | ≤ | ≥
<Num_exp> ::= <Constant> | <Functional Expression>
<Collection> ::= COLLECTION | SET | BAG
               | LIST | ARRAY
<Term Variable> ::= u | v | w | x | y | z
<Function Variable> ::= F | G | H | I | J | K
<Collection Variable> ::= u* | v* | w* | x* | y* | z*
<Complex Term List> ::= <Complex Term>
                       [ { , <Complex Term> ... } ]
<Named Complex Term List> ::= <Name> : <Complex Term>
                             [ { , <Name> : <Complex Term> ... } ]

```

Figura 1. Definición general de reglas, tomado de [7]

Por otra parte, Kim *et al.* [16] presentan un sistema de enriquecimiento basado en consultas históricas y la frecuencia con que éstas son hechas. A partir de esto y teniendo en cuenta el patrón que debería tener una pregunta (Adverbios relativos<sup>3</sup>, sujeto, verbo), se aplica el teorema de Bayes para clasificar los términos de las consultas y definir una lista de conceptos; a partir de dicha lista se determinan las palabras que se añaden o cambian en la consulta original.

<sup>3</sup> Palabras como: Dónde, Cómo, Qué, Por qué

Liu *et al.* [21] proponen un sistema de reescritura semántica, con el fin de reparar información que presente conflictos, con las condiciones previamente establecidas en el modelo entidad relación. Para esto, se crea una tabla de índices, en la cual se especifica el valor 0 para los atributos que presentan conflictos y 1 para los que no lo presentan. A partir de esta tabla se transforma la consulta inicial, verificando en la cláusula *WHERE* que el valor del atributo a buscar no se encuentre en 0, es decir que no presente conflictos y pueda entregar al usuario información consistente.

A partir de los trabajos relacionados con la reescritura de consultas, se procede a presentar una tabla comparativa de los mismos.

	[29]	[2]	[16]	[21]
<b>Reglas lógicas</b>	+	+	?	+
<b>Consultas históricas</b>	+	-	+	-
<b>Reglas de equivalencia</b>	?	+	?	-
<b>Semántica</b>	+	+	+	+
<b>Sintaxis</b>	+	+	+	-
<b>Minería de datos</b>	+	-	?	-

Tabla 5 Comparación de trabajos de reescritura de consultas

Como se observa en la Tabla 7, la mayoría de los trabajos utilizan reglas lógicas para realizar la reescritura de consultas, y muchas veces se tienen en cuenta similitudes semánticas y/o sintácticas para modificar los términos que poseen las consultas. En cambio, se puede observar que casi ninguno de los anteriores trabajos considera algoritmos de minería de datos para realizar tal reescritura.

## ENRIQUECIMIENTO DE CONSULTAS

Al realizar una consulta, generalmente los resultados retornados no son los que el usuario espera, esto se debe a que el usuario no utiliza las palabras adecuadas para expresar lo que realmente necesita, bien sea porque no conoce mucho del tema o por el uso de términos ambiguos; además, las preferencias de usuario y el contexto en el que se desenvuelve no son tenidas en cuenta. Para esto se propone el uso de ontologías que ayudan al usuario a conocer más sobre un tema [27]. Por ejemplo, si un usuario es pedagogo y se encuentra buscando trabajo en Bogotá, podría realizar una búsqueda que sea “trabajos en Bogotá”; y debido a su profesión se podría sugerir trabajos de enseñanza.

Storey *et al.* [27] proponen un algoritmo para el uso del perfil de usuario con el fin de enriquecer consultas. Este algoritmo comienza por una clasificación, en la cual se identifica si el usuario tiene dominio sobre el tema que consulta; también, si quien realiza la consulta la hace por el mismo o si alguien le delegó realizarla y si las palabras usadas son ambiguas. Si la consulta es delegada, se utiliza el perfil de usuario; luego se escogen los datos que serán añadidos a la consulta inicial. Si la consulta enriquecida no trae resultados, se eliminan los datos que tienen menor prioridad<sup>4</sup> para el usuario y se vuelve a ejecutar la consulta.

Por otra parte, Li *et al.* [19] describen que los trabajos de enriquecimiento de consultas, en su mayoría se basan en diccionarios de sinónimos y palabras relacionadas (*Thesauri*) o en la retroalimentación de los resultados, en donde se genera una especie de *TOP* de los resultados más relevantes en una consulta. Por medio del diccionario se agregan palabras relacionadas a las incluidas en la consulta inicial, y luego se realiza una retroalimentación con el fin de buscar las respuestas sean las que en ocasiones anteriores fueron las más útiles para los usuarios que realizaron una búsqueda. Li *et al.* proponen un algoritmo donde se unen tanto el diccionario de sinónimos y palabras relacionadas, como la retroalimentación de resultado. Este algoritmo se describe a gran escala como: *i*) un usuario realiza una consulta ( $Q_0$ ), *ii*) mediante el diccionario se presentan una lista de términos relacionados, de los cuales el usuario escoge cuáles desea agregar a  $Q_0$ ; a partir de esto se crea la consulta enriquecida ( $Q_1$ ), *iii*)  $Q_1$  es ejecutada, *iv*) los resultados son ordenados por relevancia<sup>5</sup> *v*) y por último se realiza la retroalimentación de la consulta.

Así mismo, Perez *et al.* [25] y Gelbukh *et al.* [12] presentan trabajos que centran su investigación en aumentar las consultas por medio de términos relacionados y sinónimos de las palabras a buscar. Por ejemplo, al tener en una consulta la palabra “sismo”, ésta se puede relacionar con palabras como “servicios sismológicos”, “desastres naturales” y tendría como sinónimo “terremoto”. Así mismo, se buscan palabras relacionadas con las que el sistema ha generado para la lista. Por ejemplo, el sistema buscaría palabras relacionadas con “desastres naturales” como “inundaciones”, “avalanchas”, entre otros. A partir de esos datos se enriquece la consulta, agregando los términos relacionados a la consulta original, la cual es enviada a los motores de búsqueda. Sin embargo, estos dos trabajos se centran únicamente en el enriquecimiento a partir de datos lingüísticos, descuidando así las preferencias del usuario y su contexto. Además, genera un mayor volumen de resultados que pueden no ser importantes para

---

<sup>4</sup> Esta prioridad es el peso que el usuario asigna a los datos consignados en el perfil de usuario.

<sup>5</sup> Mediante el Top establecido en las retroalimentaciones.

los usuarios (*e.g.*, la búsqueda del terremoto puede terminar en resultados de otro tipo de desastres naturales).

Zayanil *et al.* [30] proponen un algoritmo para realizar el enriquecimiento, en consultas de documentos *XML*. Éste se basa en una comparación semántica entre la consulta ingresada y los datos consignados en un perfil de usuario. El perfil se crea a partir de las consultas que el usuario ha realizado previamente, y como se observa en la Figura 2, se representa por medio de ontologías semánticas; es decir, la palabra carro estaría definida con un nodo padre que sea vehículo y unos nodos hijos que podrían ser jeep, deportivo, entre otros. El algoritmo propuesto para el enriquecimiento se divide en dos partes básicas: la primera es la actualización del perfil de usuario, en la cual se verifica si la consulta contiene palabras que se encuentran dentro del perfil; de ser así, se aumenta la frecuencia<sup>6</sup>, en caso contrario se agregan al perfil. La otra parte que constituye, es el enriquecimiento de la consulta en sí; para esto, si las palabras de la consulta son similares semánticamente, se añade a la consulta sus elementos y propiedades (nodo padre y nodos hijos).

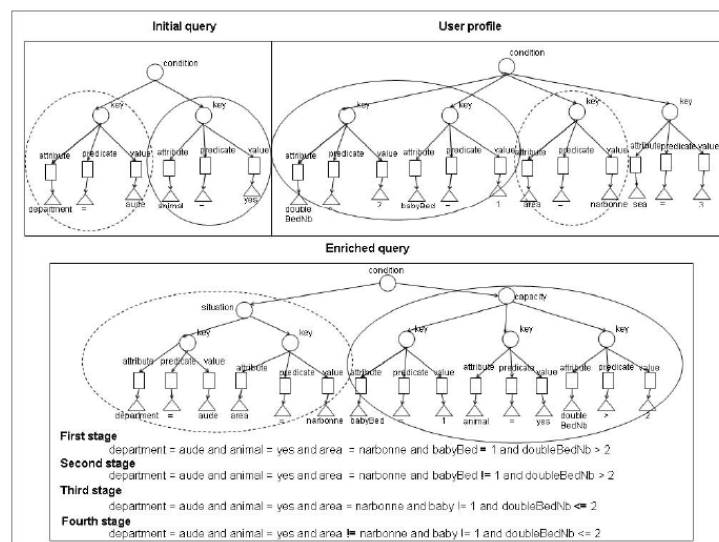


Figura 2. Enriquecimiento de consultas por medio de perfil de usuario, tomado de [30]

Por su parte Zhengyu *et al.* [31] proponen un sistema de enriquecimiento de consultas basado en los datos consignados en el perfil del usuario. Este enriquecimiento se hace agregando palabras a la cadena de caracteres de la consulta inicial. Los datos que se toman del perfil de usuario, son determinados mediante datos estadísticos sobre la correlación de temas. Este trabajo tiene como punto negativo el no considerar el perfil contextual, dado que las necesidades del usuario

<sup>6</sup> La frecuencia hace referencia a la cantidad de veces que se consulta algo consignado en el perfil. Al realizar una consulta sobre un tema que se encuentra en el perfil, éste aumentará su frecuencia.

pueden cambiar dependiendo del contexto que lo rodee. Por ejemplo, si un usuario se encuentra trabajando es probable que sus consultas estén encaminadas a temas laborales (*e.g.* se busca Java como lenguaje de programación), mientras que si el usuario se encuentra descansando, sus necesidades podrían variar, realizando consultas que no sean laborales (Por ejemplo, la consulta de Java podría relacionarse con la isla de Indonesia).

Cui *et al.* [9] proponen un sistema que se encarga de enriquecer las consultas de los usuarios a partir de datos consignados en un registro. Este registro contiene la información de las consultas realizadas por el usuario y los documentos vistos en dichas consultas. Esto con el fin de buscar las relaciones existentes, entre los datos del documento consultado y la consulta en sí. Cuando el usuario busca un término que se encuentre en el registro, se aumentará la consulta inicial, y se incluirán los temas relacionados que se adquieren de los documentos visitados previamente.

Como se ha presentado anteriormente, varios autores como Shapira *et al.*, Abbas *et al.*, Li *et al.*, entre otros, consideran el mecanismo de enriquecimiento de consultas como base para entregar a los usuarios información adaptada. Así mismo, también se cuenta con propuestas de filtrado de información, para así poder realizar una adaptación al contenido que le será entregado al usuario. Por ejemplo, Shapira *et al.* [26] describen que un sistema de filtrado básico (ver Figura 3), está compuesto por *i)* un analizador de información, el cual es el encargado de adquirir información de los proveedores; esta información es analizada y enviada a un *ii)* componente que se encarga del proceso de filtrado. A su vez, se cuenta con *iii)* un perfil de usuario, que contiene la información que permite identificar las necesidades del usuario; esta información es enviada al proceso de filtrado. Luego, en el proceso de filtrado se compara la información del perfil de usuario y del analizador de información, con el fin de identificar los resultados que podrían ser relevantes para los usuarios. Para finalizar, el usuario realiza una evaluación sobre los resultados retornados, que será enviada *iv)* al proceso de aprendizaje, el cual es el encargado de actualizar el perfil de usuario.

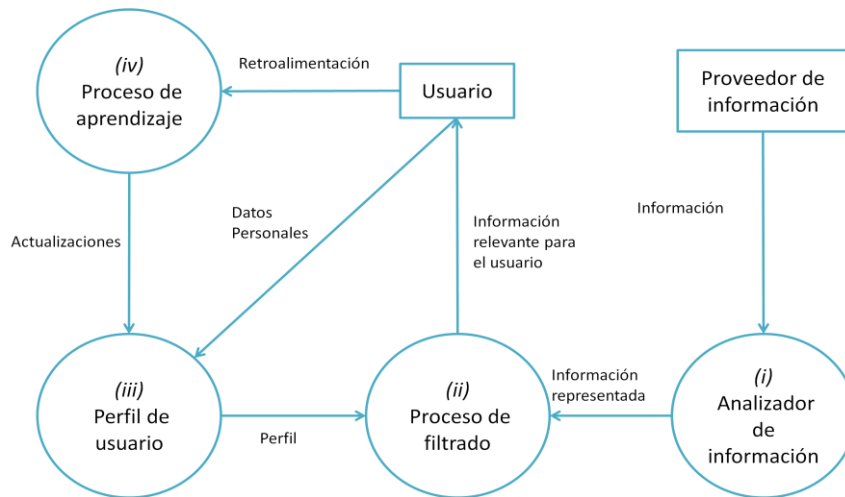


Figura 3. Sistema de filtrado, adaptado de [26]

Abbas *et al.* [1] proponen un sistema de filtrado que utiliza un perfil de usuario y un perfil contextual, para adaptar los resultados que son entregados al usuario. Éste trabaja sobre documentos *XML*. Para el proceso de filtrado, se procede en primera instancia a realizar la consulta original del usuario, buscando los documentos *XML* que la respondan. Con base en dichos documentos se realiza el filtrado de resultados, comparando los nodos del documento, con los datos consignados en los perfiles, y así obtener los resultados que sean acordes a las necesidades de un usuario. A pesar de la utilización de los perfiles para la adaptación, el proceso de filtrado exige la recepción de todos los posibles resultados que contesten la consulta inicial del usuario, por lo cual sería más costosa la recepción y comparación de cada resultado con los perfiles, que enriquecer la consulta y así obtener los resultados adaptados sin la necesidad de comparar los datos.

A continuación se presenta una tabla comparativa de los trabajos de enriquecimiento de consultas.

	[3]	[27]	[19]	[25]	[12]	[30]	[31]	[9]	[26]	[1]
<b>Perfil Contextual</b>	+	?	-	-	-	-	-	-	?	+
<b>Perfil de usuario</b>	+	+	-	-	-	+	+	+	+	+
<b>Registro de cambios</b>	?	-	-	-	-	-	?	-	+	?
<b>Semántica</b>		+	+	+	+	+	?	+	?	?
<b>Términos relacionados</b>	+	+	+	+	+	+	?	+	-	-
<b>Consultas históricas</b>	-	-	+	?	?	+	-	+	-	-
<b>Filtros</b>	-	-	-	-	-	-	-	-	+	+
<b>Uso de ontologías</b>	?	-	?	-	+	+	?	-	?	+

Tabla 6 Comparación trabajos relacionados enriquecimiento.

Como se puede ver en la Tabla 6, la mayoría de los trabajos basan la adaptación de las consultas en un enriquecimiento semántico, en su gran mayoría basados en diccionarios de sinónimos y términos relacionados; otros pocos trabajos utilizan un perfil de usuario para poder entregar al usuario resultados acordes con sus gustos. Sin embargo, los trabajos en los que se considera el uso de un perfil contextual son pocos, y como se había mencionado anteriormente, el uso de un perfil de usuario y un perfil contextual son aspectos fundamentales para la adaptación del contenido. Por otra parte, se cuenta con el uso de consultas históricas lo que se ve en la mayoría de los trabajos; estas son utilizadas en su gran mayoría para relacionar términos que podrían ser incluidos en el momento de enriquecer una consulta.

## Referencias

- [1] Abbas, K., Verdier, C., Flori. A.: Exploiting Profile Modeling for Web-Based Information System, WISE 2007 Workshops, pp. 313--324. Springer-Verlag Berlin Heidelberg (2007)
- [2] Aslam, A., Khan, S., Latif, K.: Semantic Based Query Rewriting in Heterogeneous Source. In: 4<sup>th</sup> International Conference on Emerging Technologies, pp. 292—297, Islamabad (2008)
- [3] Bai, J. , Nie, J., Bouchard, H., Cao, G.: Using Query Contexts in Information Retrieval. In: The 30th annual international ACM SIGIR conference on Research and development in information retrieval, pp.12--22. Amsterdam, The Netherlands (2007)
- [4] Berkovsky, S.: Ubiquitous User Modeling in Recommender Systems. In: User Modeling 2005, LNCS, vol 3538, pp. 496-498. Springer Verlag, Berlin Heidelberg (2005)
- [5] Burak, A., Sharon, T.: Usage Patterns of friendzone - Mobile Location-Based Community Services. In: 3rd International Conference on Mobile and Ubiquitous Multimedia, vol. 83, pp. 93–100. New York, USA (2004)
- [6] Carmichael, J., Kay, J., Kummerfeld, B.: Consistent Modelling of Users, Devices and Sensors in a Ubiquitous Computing Environment. Vol 15, pp. 197-234. Springer Netherlands (2005)
- [7] Chen, X., Chen, Y., Rao,F.: An Efficient Spatial Publish/Subscribe System for Intelligent Location-Based Services. In: 2nd International Workshop on Distributed Event-Based Systems (DEBS '03); Disponible en: [http://www.eecg.toronto.edu/debs03/papers/chen\\_etal\\_debs03.pdf](http://www.eecg.toronto.edu/debs03/papers/chen_etal_debs03.pdf).
- [8] Cristancho, J., Romero, A., MAICO: Modelo de adaptación de la información en comunidades, Proyecto especial Pontificia Universidad Javeriana (2008)
- [9] Cui, H., Wen, J., Nie, J., Ma, W.: Query Expansion by Mining User LOGS. In: IEEE Transactions on Knowledge and Data Engineering, vol 15, pp. 829 -- 839. (2003)
- [10] Finance, B., Gardarin G.: A Rule-Based Query Rewriter in an Extensible DMBS. In: Seventh International conference on Data Engineering, pp. 248—256. Kobe, Japon (1991)
- [11] Gandon, F., Sadeh, N.: Semantic Web Technologies to Reconcile Privacy and Context Awareness. In: Journal of Web Semantics. Vol. 1, no. 3. (2004).
- [12] Gelbukh, A.: Lazy Query Enrichment: A Method for Indexing Large Specialized Document Bases with Morphology and Concept Hierarchy. In: Database

- and experts System Application, LNCS, vol 1873, pp. 526--535Springer.Verlang, Berlin Heidelberg (2000)
- [13] Gregory, D., Abowd, D., Dey, A., Orr, R., Brotherton, J.: Context-Awareness in Wearable and Ubiquitous Computing. In: 1<sup>st</sup> IEEE International Symposium on Wearable Computers (ISWC), pp. 179 --180. IEEE Computer society, Washington (1997)
- [14] Heckmann, D., Krueger, Antonio.: A User Modeling Markup Language (userml) for Ubiquitous Computing. In: User modeling 2003, LNCS, Vol 2702, pp. 148--152. Springer Berlin Heidelberg (2003)
- [15] Hristova, N., P. O'Hare, G.M.: Ad-me: Intelligent Context-Sensitive Advertising within a Mobile Tourist Guide. In: 12th Irish AI and Cognitive Science Conference. Ireland (2001)
- [16] Kim, H., Bu, K., Kim, J., Lee, S.: Exploiting Question Concepts for Query Expansion. Computational Linguistics and Intelligent Text Processing, LNCS, vol 3106, pp. 624--627. Springer-Verlag (2005)
- [17] Kirsch-Pinheiro, M., Gensel, J., & Martin, H. Representing Context for an Adaptive Awareness Mechanism. Proceedings of the 10th International Workshop on Groupware (CRIWG 2004), Lecture Notes In Computer Science, Springer-Verlag, Vol. 3198, (2004), pp. 339-348.
- [18] Lech, T., Wienhofen, L.: Ambieagents: A Scalable Infrastructure for Mobile and Context-Aware Information Services. In: Aarts, H., Westra, J. (eds.): Proceedings of the 4th International Conference on Autonomous Agent and Multi-Agent Systems (AAMAS 2005) (Utrecht, Netherlands, July 25-29, 2005), pp. 625--631. ACM Press. New York, NY (2005)
- [19] Li, J., Guo, M., Tian, S.: A New Approach to Query Expansion. In: IV International Conference on Machine Learning and Cybernetics, IEEE, vol 4, pp. 2302, 2306. (2005)
- [20] Li, Y., Yao, Y. Y.: User Profile Model: A View from Artificial Intelligence. In: Rough Sets and Current Trends in Computing, LNCS, vol 2475, pp. 83 – 87. Springer Berlin / Heidelberg (2002)
- [21] Liu, J., Huang, F., Ye, D., Huang, T.: Efficient Consistent Query Answering Based on Attribute Deletions. In: International Symposium on Computer Science and its Applications, 2008, pp. 222-227. IEEE (2008)
- [22] Michalski, R., Kaufman, K., Pietrzykowski, J., 'Snie'zy'nski, B., Wojtusiak, J.: Learning Symbolic User Models for Intrusion Detection: A Method and Initial Results. Advances in Soft Computing. Vol 35, pp. 273–285. Springer Berlin (2006)

- [23] Neff, F., Kehoe, A., Pitt, I.: User Modeling to Support the Development of an Auditory Help System. In: Text, Speech and Dialogue, LNCS, vol 4629, pp.390--397, Springer Berlin / Heidelberg (2007)
- [24] Orozco, A., Cárdenas, J., Flórez, L., Carrillo, A.: PAMPU: Modelo de Preferencia de Actividades para la Definición de un Perfil. In: IV Congreso Colombiano de Computación, Bucaramanga, Colombia (2009)
- [25] Pérez, J., Araujo L.: Query Expansion with an Automatically Generated Thesaurus. In: Intelligent Data Engineering and Automated Learning – IDEAL 2006, LNCS, vol 4224, pp. 771--778, Springer Berlin (2006)
- [26] Shapira, B., Hanani, U., Raveh A., Shoval, P.: Information Filtering: A New Two-Phase Model Using Stereotypic User Profiling. In: Journal of Intelligent Information Systems, vol 8, pp. 155--165. Springer Netherlands (1997)
- [27] Storey, V., Sugumaran, V., Burton, A.: The Role of User Profiles in Context-Aware Query Processing for the Semantic Web. In: Natural Language Processing and Information System, LNCS, vol 3136, pp.45--65. Springer Berlin (2004)
- [28] Wang, Y., Aroyo, L. M., Stash, N., Rutledge, L.: Interactive User Modeling for Personalized Access to Museum Collections: The Rijksmuseum Case Study. In: Semantic Web and Peer to Peer, LNCS, Vol 4511, pp. 385--389. Springer Berlin (2007)
- [29] Yang, X., Wang, B., Wang, G., Yu, G.: A Query Rewriting System for Enhancing the Queriability of Form-based Interface. In: Digital Libraries: International collaboration and Cross-Fertilization, LNCS, vol 3334, pp. 462--472. Springer Berlin (2004)
- [30] Zayanil, C., Péninou, A., Canut, M., Sedes, F.: An adaptation approach: query enrichment by user profile. In: Advanced Internet Based system and application, vol 4879, pp. 351 -- 361. Springer Berlin (2009)
- [31] Zhu, Z., Xu, J., Ren, X., Tian, Y.: Query Expansion Based on a Personalized Web Search, In: Third International Conference on Semantics, Knowledge and Grid, pp. 128-133 (2007)