

Segmentación de consultas de usuario en un dispositivo móvil: Un enfoque basado en frecuencias

Segmentation of user queries on a mobile device: A frequency-based approach

María Valencia, MSc., Santiago Eibe, PhD. y Ernestina Menasalvas, PhD.

Facultad de Informática, Universidad Politécnica de Madrid, España

mvalencia@alumnos.fi.upm.es, {seibe,emenasalvas}@fi.upm.es

Recibido para revisión 03 de Octubre de 2008, aceptado 25 de Agosto de 2009, versión final 10 de Septiembre de 2009

Resumen— Identificar la intención de los usuarios es una actividad de creciente interés para las organizaciones debido a que les permite proporcionar a sus usuarios servicios de valor añadido en respuesta a las acciones del usuario. Este problema ha sido abordado en diferentes enfoques, sin embargo aun existen algunos problemas abiertos. En concreto, planteamos abordar el problema de identificar la intención del usuario que realiza búsquedas en un sitio Web desde su dispositivo móvil. Esto representa un reto en dispositivos móviles debido al impacto de la ubicuidad, en particular, el contexto cambiante del usuario influye en las necesidades del usuario. En este artículo presentamos un modelo que permite analizar las consultas emitidas por un usuario en base a su relación con consultas emitidas previamente y con contenido del sitio Web. Presentamos el proceso y un modelo de metadatos que permite la sistematización del proceso. Este modelo de metadatos constituye la base para la automatización del proceso.

Palabras Clave— Intención del Usuario, Motores de Búsqueda, Minería de Datos Ubicua

Abstract— Identify the user's intention is an activity of increasing interest due to the fact that it makes possible for organizations to provide users with value-added services in response to their actions. This problem has been addressed from different approaches, however, there are still some open problems. Specifically, we address the problem of identifying the intention of users that searches in a Website from their mobile device. This is challenging in mobile devices due to the impact of ubiquity, in particular the users' changing context that influences their information needs. In this paper, we present a model to analyze the queries issued by a user based on their relationship with previously issued queries and the proper Website content. We present the process and a metadata model that allows its systematization, actually forming the basis for its automatization.

Keywords— User Intention, Search Engines, Ubiquitous Data Mining

I. INTRODUCCIÓN

Los avances tecnológicos acontecidos en los últimos años, y en concreto la computación ubicua, permiten que usuarios móviles con pequeños y potentes dispositivos utilicen aplicaciones en casi cualquier lugar. En concreto, en estos dispositivos es posible realizar procesos de análisis que explotan información relativa al contexto del usuario: localización, tiempo, características del entorno del usuario y, en general, variables del estado del usuario. Como resultado surge la oportunidad de ofrecer servicios personalizados para cada usuario basados en este análisis.

Identificar la intención del usuario en un sitio web es un problema abierto que ha sido abordado en diversos trabajos. Los enfoques planteados en [14, 9, 2, 10, 7] se caracterizan por: i) calcular el modelo utilizando información global de los usuarios, y ii) ejecutar el proceso de manera centralizada. Sin embargo, en entornos móviles el análisis local de la información de cada usuario resulta de interés puesto que puede ayudar a entender la necesidad del usuario en su entorno. En entornos móviles surgen dos posibles modelos: i) enviar la información del usuario y calcular modelos de comportamiento de manera centralizada, ii) computar modelos de comportamiento local del usuario. Este segundo modelo que puede resultar interesante e incluso ser la única opción en algunos casos en los que la comunicación con un servidor central no sea posible, plantea la necesidad de analizar los retos del proceso de descubrimiento en dispositivos posiblemente con limitaciones de recursos y en los que se requiere autonomía.

Por otra parte, limitar el análisis a la información de cada usuario en lugar de emprender un análisis global de la información de todos los usuarios plantea los siguientes retos:

- Determinar la información de la que se dispone: su disponibilidad y sus requisitos.
- Requerimientos derivados del escenario ubicuo como la posible heterogeneidad de los datos, la privacidad de la información o la fiabilidad y la escalabilidad de los sistemas

Son numerosas las aplicaciones y dominios que se benefician del análisis de datos locales de usuario: la tecnología asistencial, la Web y domótica se encuentran entre ellos. En el caso particular de la Web, algunos enfoques integran diferentes tipos de información disponibles en el dispositivo móvil del usuario. En [15], los autores presentan un sistema de búsqueda móvil que dividen en dos partes lógicas: el motor de búsqueda local y un meta rastreador. En concreto, el motor de búsqueda Web local proporciona al usuario la capacidad de adaptar los resultados de la búsqueda a sus propias necesidades, mediante la actualización del índice de búsqueda a partir del análisis conjunto de la información generada localmente por cada usuario y el cambio de contenido. En la misma línea, en [12] se presenta un enfoque de búsqueda personalizada. Esta personalización se propone del lado del cliente y se basa en el mapeo de los intereses a una taxonomía y en la categorización de los resultados, lo que permite desplegar los resultados de acuerdo a los intereses del usuario. Estos intereses son obtenidos a partir de información local que describa los intereses profesionales y habilidades del usuario.

Por otra parte, los autores describen en [8] un enfoque para predecir el contexto del usuario aplicado a entornos asistidos, para lo cual analizan la información que describe la ejecución de un conjunto de tareas por parte de cada usuario. Por otro lado, en [3] se describe el framework SPE (Secure Persona Exchange), que proporciona servicios personalizados a los usuarios en entornos de computación ubicua en base a las preferencias de cada usuario. La información analizada en este enfoque corresponde a las preferencias del usuario que son almacenadas en su dispositivo móvil, y son proporcionadas de manera explícita por el usuario. Sin embargo, este enfoque no tiene en cuenta el contexto dinámico del usuario como localización o la actividad actual y la personalización está restringida a los datos indicados por el usuario.

En este artículo se presenta un enfoque que permite analizar la intención de búsqueda del usuario y para ello se propone analizar la relación de la consulta que emite con consultas pasadas y la información que visualiza en el sitio Web. En concreto, nos centramos en el análisis de la frecuencia de aparición de los términos que aparecen en las consultas de los usuarios y en el análisis de la relación de las consultas con los contenidos visualizados en el site. El artículo presenta las fases y la metodología de análisis de este proceso. Para la sistematización del proceso proponemos un modelo de metadatos que permita en un futuro la automatización de parte del proceso.

El artículo se ha organizado de la siguiente manera. En la sección 2 se define el proceso de análisis de consultas y su relación con el contenido del site. En la sección 3 se describen los retos que se presentan como consecuencia de la ejecución local del proceso de minería en base a información local. El modelo de metadatos presentado en la sección 4 permite la sistematización del proceso. Para concluir la sección 5 presenta la discusión y las posibles líneas futuras.

II. ANÁLISIS DE CONSULTAS BASADO EN FRECUENCIAS

Proponemos identificar la intención de un usuario en un dispositivo móvil mediante el análisis de las consultas emitidas por el usuario al buscador de un sitio Web desde su dispositivo (ordenador portátil, PDA, u otro similar). En concreto, analizamos los términos contenidos en la consulta con el fin de identificar: 1) la relación de la consulta con consultas anteriores y 2) la relación de la consulta con el contenido Web.

- Relación de la consulta con consultas anteriores: En este enfoque consideramos el hecho de que una consulta emitida por un usuario refleja su intención de búsqueda, y por tanto la relación existente entre ésta y otras consultas anteriores determina la tendencia local de búsqueda del usuario en el sitio Web. Con el fin de identificar esta relación proponemos analizar la frecuencia de aparición de los términos de una consulta en consultas emitidas previamente por el mismo usuario.

- Relación de la consulta con el contenido Web: El contenido publicado en el sitio Web puede ser muy cambiante dependiendo de la naturaleza del portal y refleja la información acerca de la dinámica de los usuarios y sus intereses. Por ello, proponemos analizar la relación entre las consultas emitidas por un usuario y el contenido Web desplegado en un momento determinado en un sitio Web a fin de identificar la influencia de los contenidos sobre la tendencia de búsqueda del usuario y viceversa. Para identificar esta relación, de forma análoga a la anterior, analizamos la frecuencia de aparición de los términos de una consulta en el contenido publicado en la página principal del sitio Web.

Dado que la frecuencia es una función que depende del tiempo, definimos períodos para los cuales se calcula la frecuencia de aparición de un término. En base a estos coeficientes de frecuencia obtenemos conjuntos de términos de acuerdo a los siguientes aspectos: Frecuencia, Estabilidad, Novedad, Evolución. Con respecto a estos conceptos se definen los siguientes conjuntos de términos:

- Frecuencia

Término frecuente - $ft(p)$: un término pertenece a este conjunto si en el periodo p el número de veces que aparece supera el radio medio de aparición.

- Estabilidad

Término estable - $st(p)$: los términos estables en el periodo π son aquellos que aparecen de forma uniforme a lo largo del periodo comprendido entre $\pi-n$ y $\pi+m$

- Novedad

Término nuevo - $nt(p)$: un término pertenece a este grupo si aparece por primera vez en el periodo

- Evolución

Término que reaparece - $rt(p)$: los términos que reaparecen en el periodo π son aquellos que aparecen en consultas que fueron emitidas durante el periodo $\pi-n$ o antes y vuelven a ser utilizadas en consultas emitidas en el periodo π

Término de frecuencia creciente - $ift(p)$: un término pertenece a este grupo si su frecuencia de aparición en el periodo π se

incrementa respecto al periodo inmediatamente anterior

Término de frecuencia decreciente - $dft(p)$: un término pertenece a este grupo si su frecuencia de aparición en el periodo π se incrementa en el periodo inmediatamente posterior

2.1. Relación de una consulta de usuario con consultas pasadas y con el contenido visualizado

El proceso de análisis de las consultas de usuario en base a la frecuencia de aparición de los términos que contiene, consiste en obtener conjuntos de términos en base a la frecuencia de aparición de los mismos en consultas pasadas y en el contenido Web, posteriormente etiquetar las consultas en base a esta información y finalmente segmentar las consultas. Este proceso es representado en el DFD presentado en la figura 1.

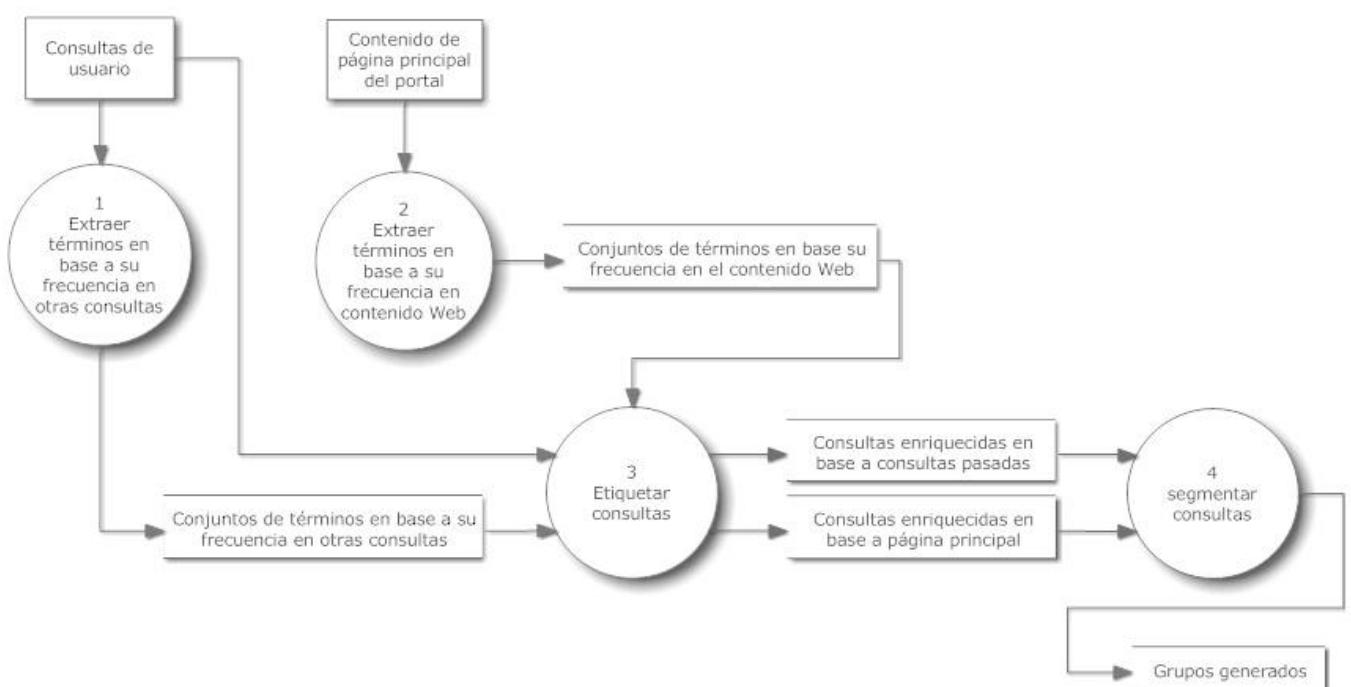


Figura 1. Proceso

Como se muestra en la figura 1, en el proceso se utilizan como entrada tres fuentes externas de datos:

- Consultas de usuario: Consultas emitidas por el usuario que son almacenadas en el dispositivo desde el cual el usuario realiza el proceso de búsqueda.

- Contenido de la página principal del portal: Porciones de contenido capturados de la página principal del sitio Web, para cada uno de los cuales se almacena la fecha y hora en que se capturó, la posición dentro de la página y el texto asociado.

A continuación describimos las fases del proceso:

1. Extraer términos en base a su frecuencia en otras consultas.

El objetivo de este proceso consiste en calcular conjuntos de términos en base a su frecuencia de aparición en consultas pasadas en diferentes períodos. Para ello examinamos las consultas emitidas por el usuario en diferentes períodos y calculamos para cada uno su frecuencia de aparición. En base a este valor evaluamos para cada par término-período su frecuencia, estabilidad, novedad y evolución y generamos los conjuntos correspondientes para cada período. La tabla 1 presenta el pseudocódigo del proceso.

Este proceso requiere que las consultas de usuario hayan sido previamente preprocesadas. En la tabla 1 incluimos las tareas de preprocesamiento de las consultas, las cuales pueden ser ejecutadas en el momento en que calculamos los conjuntos de términos o en una fase previa.

2. Extraer términos en base a su frecuencia en el contenido Web

En esta fase pretendemos calcular conjuntos de términos en base al contenido Web en diferentes períodos: En forma análoga al cálculo anterior, evaluamos cada uno de las piezas de contenido capturados de la página principal del sitio Web y

calculamos la frecuencia de aparición de cada término durante un periodo determinado, y posteriormente calculamos los conjuntos de términos basados en su frecuencia, estabilidad, novedad y evolución. La tabla 2 muestra el pseudocódigo del proceso.

De forma similar al proceso anterior, es necesario capturar y preprocesar el contenido Web. En la tabla 2 especificamos esta tarea aún cuando se puede llevar a cabo bien sea en el dispositivo previamente a calcular los conjuntos de términos, o en un servidor y asumir que un servicio Web proporciona esta información.

Entradas:	Consultas de usuario registradas en el dispositivo
Salidas:	Conjuntos de términos en base a su frecuencia en otras consultas
<hr/>	
1. Preprocesar consultas	
<ul style="list-style-type: none"> - Para cada búsqueda registrada en el dispositivo <ul style="list-style-type: none"> - Identificar los parámetros de búsqueda - Obtener la consulta q - Limpiar la consulta y obtener los terminos de la consulta y <ul style="list-style-type: none"> - Eliminar diferencias entre mayúsculas y minúsculas - Reemplazar signos de puntuación con espacios en blanco - Comprimir espacios en blanco en solo uno - Descartar stopwords 	
2. Calcular conjuntos de términos basados en otras consultas	
<ul style="list-style-type: none"> - Determinar los períodos para los cuales se calcula la frecuencia de los términos - Para cada periodo p <ul style="list-style-type: none"> - Para cada término t <ul style="list-style-type: none"> - Calcular $f_{consultas}(t,p)$ - Generar los “conjuntos de términos” para el periodo: $ft_{consultas}(p)$, $st_{consultas}(p)$, $nt_{consultas}(p)$, $rt_{consultas}(p)$, $it_{consultas}(p)$ y $df_{consultas}(p)$ 	
<hr/>	

Tabla 1. Cálculo de conjuntos de términos basado en consultas pasadas

Entradas:	- Contenido de página web (página principal del site)
Salidas:	- Conjuntos de términos en base a su frecuencia en el contenido Web
<hr/>	
1. Preprocesar Contenido	
<ul style="list-style-type: none"> - Definir la frecuencia de captura del contenido p - Capturar contenido de página web $wp_content$ <ul style="list-style-type: none"> - Para cada $wp_content$ <ul style="list-style-type: none"> - Eliminar tags de HTML - Eliminar imágenes - Obtener porciones de contenido C - Para cada C <ul style="list-style-type: none"> - Eliminar case differences - Reemplazar signos de puntuación por espacios en blanco - Comprimir espacios en blanco en uno solo - Descartar stop words - Identificar el momento en que es capturado $m_content$ - Determinar la localización física de C en la página web $\rightarrow page_section$ - Obtener los términos incluidos en la porción de contenido $T(C) = \{t_1, t_2, \dots, t_n\}$ 	
2. Calcular conjuntos de términos basados en el contenido de página principal	
<ul style="list-style-type: none"> - Determinar los períodos para los cuales se calcula la frecuencia de los términos - Para cada periodo p <ul style="list-style-type: none"> - Para cada término t <ul style="list-style-type: none"> - Calcular $f_{contenido}(t,p)$ - Generar los “conjuntos de términos basados en página principal” para el periodo: $ft_{contenido}(p)$, $st_{contenido}(p)$, $nt_{contenido}(p)$, $rt_{contenido}(p)$, $it_{contenido}(p)$ y $df_{contenido}(p)$ 	
<hr/>	

Tabla 2. Cálculo de conjuntos de términos basado en el contenido Web

3. Etiquetar consultas

El propósito de esta fase consiste en etiquetar las consultas emitidas por el usuario de acuerdo a la frecuencia de aparición de los términos que contiene. Para ello, una vez obtenidos los conjuntos de términos basados en la frecuencia de aparición

de términos en otras consultas y en el contenido Web, añadimos información a cada consulta que indica para cada conjunto de términos generado, si la consulta contiene o no términos pertenecientes a dicho conjunto. La tabla 2 muestra el pseudocódigo del proceso.

Entradas:	<ul style="list-style-type: none"> - Conjuntos de términos en base a su frecuencia en otras consultas - Conjuntos de términos en base a su frecuencia en el contenido Web
Salidas:	<ul style="list-style-type: none"> - Consultas enriquecidas en base a consultas pasadas - Consultas enriquecidas en base a página principal

1. Etiquetar consultas en base a frecuencia de aparición de los términos en otras consultas
 - Para cada consulta q
 - Para cada periodo p
 - Para cada conjunto de visibilidad $cv_{consultas}(p)$ obtenido en el proceso 1: $fl_{consultas}(p)$, $st_{consultas}(p)$, $nt_{consultas}(p)$, $rt_{consultas}(p)$, $ifl_{consultas}(p)$ y $dfl_{consultas}(p)$
 - Indicar si la consulta contiene o no un término incluido en $cv_{consultas}(p)$
2. Etiquetar consultas en base a frecuencia de aparición de los términos en página principal
 - Para cada consulta q
 - Para cada periodo p
 - Para cada conjunto de visibilidad $cv_{página-principal}(p)$ obtenido en el proceso 2: $fl_{contenido}(p)$, $st_{contenido}(p)$, $nt_{contenido}(p)$, $rt_{contenido}(p)$, $ifl_{contenido}(p)$ y $dfl_{contenido}(p)$
 - Indicar si la consulta contiene o no un término incluido en $cv_{página-principal}(p)$

Tabla 2. Proceso de enriquecimiento de consultas

4. Etiquetar consultas

Tomando como datos de entrada el conjunto enriquecido de consultas, ejecutamos un proceso de segmentación de consultas que permite identificar los tipos de consultas realizadas por el usuario en base a su relación con otras consultas y con el contenido de página principal.

III. CONSECUENCIAS DE LA EJECUCIÓN LOCAL DEL MODELO EN UN DISPOSITIVO MÓVIL

En minería de datos encontramos diversos trabajos que abordan el problema de identificar la intención de búsqueda de los usuarios [10, 7, 11, 1]. Los enfoques descritos en estos trabajos se caracterizan por ejecutar el proceso de manera centralizada utilizando para ello información de las búsquedas realizadas por todos los usuarios.

Sin embargo, en entornos móviles se presenta el caso en el que es necesario identificar la intención del usuario en base a la información de cada usuario.

Desde este punto de vista, pasamos de entornos en los que la computación de los modelos se realizaba analizando toda la información disponible acerca de los usuarios y su comportamiento, a un modelo en el que sólo es necesario analizar el comportamiento local de cada usuario a fin de identificar sus intereses.

En este entorno, surgen dos posibles escenarios para el cálculo de los modelos. El primero consiste en enviar la información local del usuario a un servidor centralizado y

ejecutar el proceso de manera centralizada. El segundo consiste en ejecutar el proceso localmente en el dispositivo móvil.

No obstante, ejecutar el proceso de identificar la intención del usuario en el dispositivo móvil constituye un reto debido a que se enfrenta a las dimensiones a lo largo de las cuales se presenta la ubicuidad: dispositivos, datos, proceso, usuarios y contexto. En estos entornos requerimos de componentes de minería que tengan en cuenta la limitación de recursos y que sean adaptables. Pero ello no es suficiente sin incorporar autonomía, puesto que el reto de la ubicuidad plantea además un entorno en el que no podemos asumir que los ingenieros de minería de datos estén disponibles durante el proceso de descubrimiento de conocimiento. Estos aspectos se describen brevemente en las secciones siguientes.

a. Adecuación a los recursos

Cuando la ejecución del modelo se realiza en un dispositivo móvil, la limitación de recursos plantea una problemática que hace necesario que nuestro modelo sea consciente de los recursos disponibles. En particular, el modelo debe considerar las restricciones que determinan los recursos de memoria, almacenamiento, conectividad y suministro de energía para el correcto cálculo del modelo.

b. Adaptabilidad

Adaptabilidad es definido como la cualidad de ser capaz de acomodarse o ajustarse a nuevas condiciones o usos [13]. En este sentido, las condiciones de los entornos móviles hacen que sea necesario que el modelo sea adaptable, por un lado, a los intereses del usuario y, por otro, a las características del entorno.

Adaptabilidad con respecto a los intereses del usuario

Los intereses y el contexto del usuario no son factores estáticos y pueden variar en el tiempo. Por ello, se requiere que el modelo que describe el comportamiento del usuario sea adaptable a estos cambios. Consecuentemente, el modelo debe ser actualizado y validado constantemente, lo cual permite garantizar su validez.

Adaptabilidad con respecto a las características del entorno:

Como se vio en la sección anterior, se requiere que el sistema sea consciente de los recursos que están disponibles. Sin embargo, es necesario además que el sistema emplee una estrategia que le permita adaptarse en las situaciones en que varíe la disponibilidad de los recursos.

Por otro lado, debemos considerar el coste computacional que implica la actualización de los modelos en cuanto a almacenamiento de datos, memoria y tiempo de respuesta. Debido a que las restricciones de almacenamiento no permiten en muchas ocasiones almacenar toda la información histórica del usuario, se deben establecer mecanismos que permitan desechar información a medida que se recogen nuevos datos del usuario. Así mismo, también es requerido hacer una actualización incremental de los modelos teniendo en cuenta las acciones recientes del usuario.

c. Autonomía

La ubicuidad plantea la necesidad de contar con componentes autónomos capaces de procesar la información recibida y extraer conocimiento sin la intervención del minero. Esto implica que el componente de minería, o un controlador que lo gestione, sea capaz de conocer en cada momento de la ejecución el estado de la misma, las precondiciones de cada proceso, y las restricciones asociadas a los recursos, el modelo y los datos de entrada. En base a esta información, el componente debe ser capaz de decidir si puede continuar la ejecución o no, así como decidir el paso siguiente. Esta situación hace que se minimice, y en el caso ideal que no sea necesaria, la presencia humana que supervise el proceso de minería de datos.

Pero ello requiere sistematizar y automatizar el proceso. En base a esto proponemos un modelo de metadatos que presentamos en la siguiente sección, como soporte al proceso presentado en este artículo.

IV. UNA SOLUCIÓN BASADA EN METADATOS

Los retos descritos en la sección anterior plantean la necesidad de disponer de información detallada que describa los datos y el proceso. Ello motiva el desarrollo de una base de datos independiente que dé soporte a la automatización y que almacene esta información que, en adelante, llamamos metadatos. En este modelo hacemos énfasis en los retos asociados a la autonomía y la limitación de recursos.

Aunque este diseño se ha realizado para un proceso en particular, las entidades y los conceptos modelados son

suficientemente generales como para ser aplicados en cualquier fase del proceso o dominio.

Un componente autónomo se caracteriza por la capacidad de auto-gestionarse, lo que supone controlar el flujo del proceso y las restricciones asociadas a éste. Para ello es necesario, por un lado, contar con información que permita decidir en cada momento si es posible o no lanzar la ejecución de una tarea y, por otro lado, registrar información a partir de la cual el sistema aprenda y se adapte a nuevas condiciones del entorno.

El modelo de metadatos permite automatizar el proceso en la medida en que proporciona información al sistema que guía y controla la ejecución de las tareas que constituyen el proceso. De esta manera, el sistema consulta los metadatos para determinar si puede o no ejecutar una tarea dada.

En este sentido, identificamos dos planos de información que deben ser integrados en el modelo:

- Plano de descripción:

En este plano se define la información que da soporte a la automatización y que describe en detalle cada tarea. El soporte a la automatización está dado por la especificación de parámetros y restricciones que permiten la ejecución de la tarea sin supervisión humana. Para definir los elementos que representamos en este plano nos basamos en los estándares existentes, en concreto, en PMML que define el diccionario de datos [16]. También definimos las entradas y salidas asociadas a cada tarea, así como las restricciones sobre cada una de ellas.

Por otra parte, dado que requerimos la ejecución de las tareas en un entorno adecuado a los recursos, incluimos en el modelo información que describa los recursos y las restricciones sobre éstos que permitan garantizar dicha ejecución.

La figura 2 muestra la parte del modelo de metadatos propuesto correspondiente al plano de descripción, el cual define la información mínima que debe contener. Este modelo permite representar la división de un proceso en tareas, de las cuales se especifican las entradas, salidas, recursos y parámetros asociados.

- Plano de monitorización:

Este plano corresponde al registro de la ejecución de las diferentes tareas. La dinamicidad de esta información permite contar con información actualizada acerca del comportamiento de cada tarea y es utilizada en el proceso de mejora. Un ejemplo de ello es presentado en la figura 3.

El modelo de metadatos correspondiente a este plano permite describir cada ejecución de una tarea. Esta descripción se presenta mediante la especificación de los parámetros con que se realiza la ejecución y el estado de finalización de la misma. Por otro lado, incluimos en el modelo la marca de ejecución, que permite identificar las ejecuciones que corresponden a un mismo ciclo de ejecución (ejecución de todas las tareas).

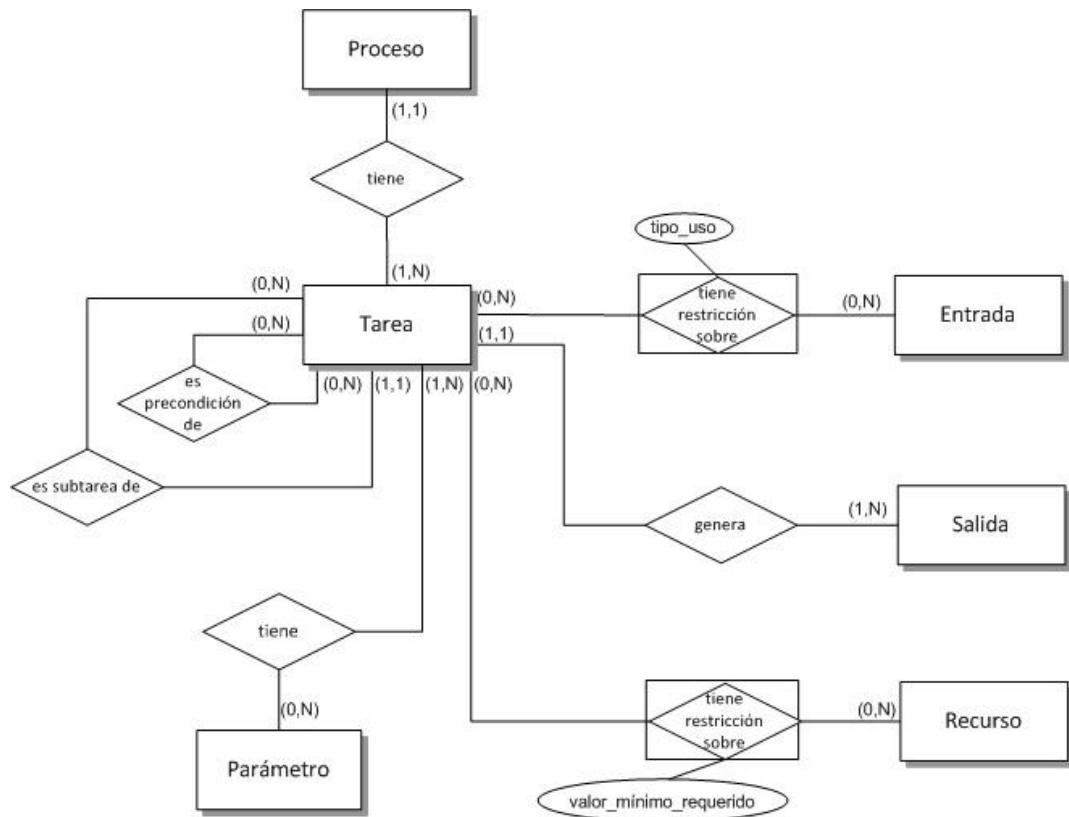


Figura 2. Definición del plano de descripción

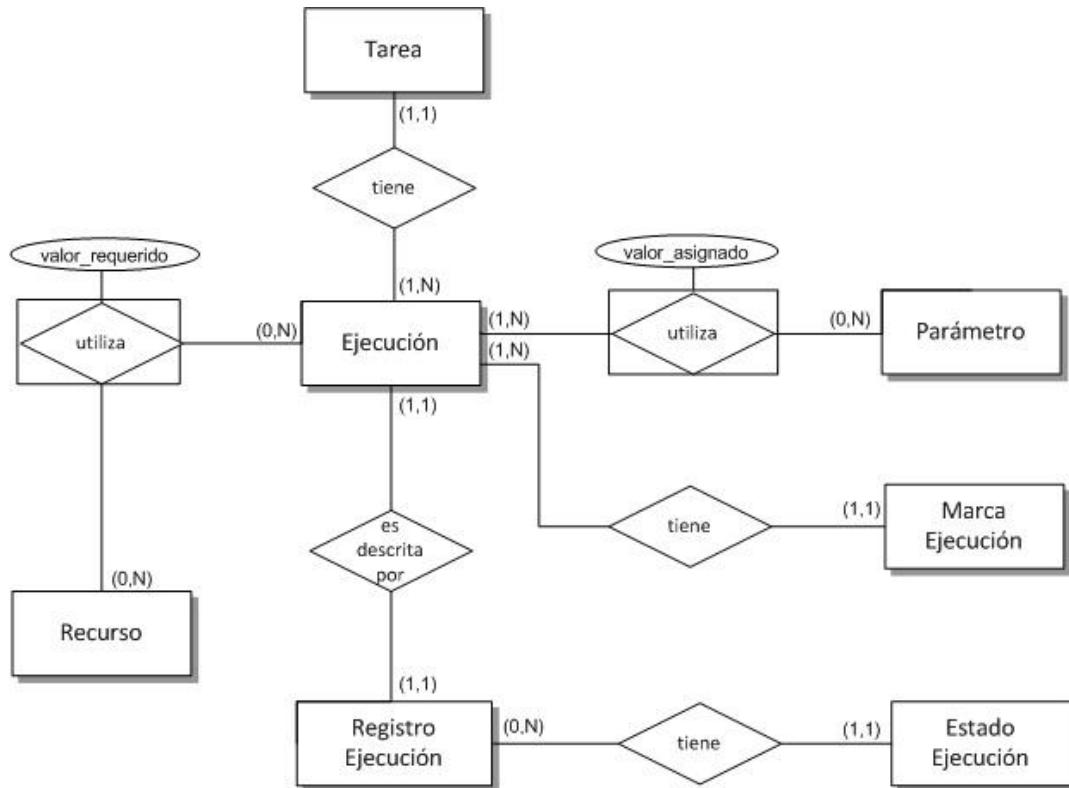


Figura 3. Definición del plano de monitorización

Finalmente y de manera complementaria a la especificación de recursos que incluimos en el plano de descripción, se establece una relación entre cada ejecución y los recursos, lo cual permite registrar la cantidad real de recursos utilizados y, de esta manera, monitorizar las condiciones del dispositivo en las que se ejecutan las tareas.

V. CONCLUSIONES

En este artículo abordamos el problema de identificar la intención del usuario mediante el análisis local de información local en dispositivos móviles. En primer lugar presentamos un modelo de análisis de las consultas del usuario basado en la frecuencia de aparición de los términos que contiene y la influencia que los contenidos del sitio Web pueden tener en ellas. En segundo lugar analizamos los retos asociados a la ejecución local del modelo. Estos retos están asociados con la urgente necesidad de desarrollar componentes que reduzcan la intervención manual del experto y que sean capaces de autogestionarse. Se quiere ir entonces hacia la definición de componentes autónomos en dispositivos móviles, lo cual requiere la automatización previa del proceso. Para ello, proponemos un modelo basado en metadatos, que permita al componente conocer en cada momento si puede o no puede ejecutar el proceso. Este trabajo constituye un primer paso hacia la automatización del proceso en dispositivos móviles. En este momento, estamos trabajando en la implementación del modelo de análisis de consultas en un dispositivo móvil, usando para ello los metadatos propuestos. Esta implementación nos ha permitido valorar la generalidad del modelo propuesto. Una de las razones es que facilita la especificación de las tareas que componen el proceso de categorización independientemente de la fase del proyecto a la que corresponda. Por otra parte, las primeras pruebas del sistema conducen a que los tiempos de ejecución se reducen, debido a que la ejecución de las tareas se realiza de manera secuencial en los momentos en que los recursos están disponibles. Todo ello, en contraposición al esquema tradicional, en el que se requería que un usuario verificara la correcta finalización de una tarea y el estado del dispositivo para lanzar la ejecución de la siguiente tarea.

REFERENCIAS

- [1] Ashkan A.; Clarke C.; Agichtein E. y Guo Q., 2008. Characterizing query intent from sponsored search clickthrough data. En: Proceedings of the SIGIR Workshop on Information Retrieval for Advertising, pp. 15-22
- [2] Baeza-Yates R.; Calderón-Benavides L. y González-Caro C., 2006. The intention behind Web queries. En: Proceedings of String Processing and Information Retrieval (SPIRE 2006). Glasgow, Scotland, 98-109.
- [3] Brar A. y Kay J., 2005. Privacy and security in ubiquitous personalized applications. En: User Modelling Workshop on Privacy-Enhanced Personalization, pp. 47-54, Edinburgh, UK, 25 July, 2005
- [4] DMG. PMML Version 3.2, <http://www.dmg.org/pmml-v3-2.html>
- [5] Jansen B. J.; Booth, D. L. y Spink, A., 2007. Determining the user intent of web search engine queries. En: Proceedings of the 16th international Conference on World Wide Web (Banff, Alberta, Canada, May 08 - 12, 2007). WWW '07. ACM, New York, NY, 1149-1150.
- [6] Lee S. W.; Bien Z.; Mokhtari M. y Feki M. A., 2007. Context aware life pattern prediction using fuzzystate q-learning. En: Okadome T., Yamazaki T., Makhtari, M. (eds.), ICOST, volume 4541 of Lecture Notes in Computer Science, pp 188-195. Springer.
- [7] Lee U.; Liu Z. y Cho J. 2005. Automatic identification of user goals in Web search. In Proceedings of the 14th international Conference on World Wide Web (Chiba, Japan, May 10 - 14, 2005). WWW '05. ACM, New York, NY, 391-400.
- [8] Li X.; Wang Y. y Aceri A., 2008. Learning query intent from regularized click graphs. En: Proceedings of the 31st Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Singapore, Singapore, July 20 - 24, 2008). SIGIR '08. ACM, New York, NY, 339-346.
- [9] Liu N.; Yan J.; Fan W.; Yang Q. y Chen Z., 2009. Identifying vertical search intention of query through social tagging propagation. En: Proceedings of the 18th international Conference on World Wide Web (Madrid, Spain, April 20 - 24, 2009). WWW '09. ACM, New York, NY, 1209-1210.
- [10] Ma Z.; Pant G. y Sheng, O. R., 2007. Interest-based personalized search. En: ACM Trans. Inf. Syst., 25(1):5.
- [11] Real Academia Española. Diccionario de la Lengua Española. Editorial Espasa
- [12] Rose D. E. y Levinson, D., 2004. Understanding user goals in web search. En: Proceedings of the 13th international Conference on World Wide Web (New York, NY, USA, May 17 - 20, 2004). WWW '04. ACM, New York, NY, 13-19.
- [13] Tiago P.; Kotilainen N.; Kokkinen H.; Vapa M. y Nurminen J. K., 2008. Mobile search - social network search using mobile devices. En: Consumer Communications and Networking Conference, 2008. CCNC 2008. 5th IEEE, pages 12011205, 10-12 Jan. 2008