

Uso de escenarios aplicados a la Ingeniería de los Requisitos para la creación de una hemeroteca digital

Perdrix, F.;
Piqué, E.; Cal, J.
Diari Segre S.L.U.
C/ Del Riu nº6
25007 Lleida
+(34) 973 24 80 00

{fperdrix,epique,jcal}@
diarisegre.com

Castells, P.;
Pulido, E.; Rico, M.
Universidad Autónoma de Madrid
Ctra. de Colmenar km 15
28049 Madrid
+(34) 91 348 22 84

{pablo.castells,estrella.pulido,
mariano.rico}@ii.uam.es

Granollers, T.; Lorés, J.;
González, M.; Badia, J.
GRIHO (UdL Lleida)
C/ Jaume II nº69
25001 Lleida
+(34) 973 70 27 20

{tonig,jesus}@griho.net

RESUMEN

La gestión de contenidos en una hemeroteca digital dentro de un grupo mediático es cada día más compleja y de mayor volumen. Las proporciones actuales de información que se deben gestionar se pueden comparar con el contenido de información disponible en la red. Uno de los principales problemas que entrañan las hemerotecas es la divergencia de criterios y modelos mentales entre los que catalogan la información y los que realizan búsquedas sobre esta inmensa fuente de información. Para mitigar estas divergencias se ha usado una metodología de desarrollo centrada en los usuarios. Este artículo se centra en observar el uso de la técnica de escenarios como parte de la fase de análisis de requisitos, demostrándose que es un método válido y que conjuntamente con el uso de ontologías de representación de la información permiten dar respuesta a los factores humanos que intervienen en el universo de aplicación.

Palabras clave

Análisis de requisitos, Escenarios, Diseño Centrado en el Usuario, Web semántica, Ontología, Hemerotecas Digitales.

1. INTRODUCCIÓN

Con la digitalización de contenidos y la introducción de las nuevas tecnologías de la información, las hemerotecas de los medios de comunicación se han convertido en valiosos repositorios información y noticias de calidad al alcance de un público cada vez más amplio [1].

La disponibilidad, eficiencia y facilidad de acceso de estos modernos archivos no tienen nada que ver con los arcaicos (aunque vigentes aún) sistemas de papel y microficha. Las hemerotecas digitales son una potente y versátil herramienta capaz de dispensar múltiples servicios a un público enormemente variado: un sistema de documentación para los periodistas, una fuente de referencia para investigadores en los más diversos

Se concede el permiso para la reproducción digital o impreso total o parcial de este trabajo sin contraprestación económica únicamente para la utilización personal o en clase. En ningún caso se podrán hacer o distribuir copias de para su explotación comercial. Todas las copias deben de llevar esta nota y la información completa de la primera página. Para cualquier otro uso, publicación, publicación en servidores, o listas de distribución de esta información necesitara de un permiso específico y/o el pago correspondiente.

Interacción 2004, 3-7 mayo, 2004, Lleida (España).

ámbitos (historia, cultura, política, estudiantes, administración de justicia, etc.), y un servicio de información a la carta sin barreras temporales para el lector de prensa diaria. Las empresas del sector mediático son conscientes hoy del valor añadido que esto supone, y de las consiguientes oportunidades de negocio. Un ejemplo de esta tendencia en la industria mediática ha sido la emergencia, en muy pocos años, de todo un nuevo mercado de servicios online para la redistribución, sindicación, agregación y consumo de noticias de archivo [2],[3].

La gestión de contenidos en una hemeroteca digital dentro de un grupo mediático es una actividad compleja que involucra a un grupo de personas de perfil muy diverso: redactores, reporteros, documentalistas, maquetadores, reporteros gráficos, técnicos, y eventualmente lectores. En torno a, y en interacción con, la hemeroteca, tienen lugar tareas de muy distinta naturaleza: búsqueda de información, descripción y catalogación de contenidos, elaboración de noticias, actualización, corrección, adición, eliminación y tránsito de contenidos, etc. Las hemerotecas en sí son cada día más complejas y de mayor volumen.

Para que un sistema de gestión de hemerotecas tenga éxito en un escenario complejo como éste, es esencial alcanzar, previamente al desarrollo del sistema, una comprensión profunda y sistemática de las tareas que se realizan en torno a la hemeroteca, las necesidades y objetivos de los usuarios, sus modelos mentales, y las prioridades de las organizaciones (proveedores, gestores y consumidores) en cuyo marco se desarrollan estas actividades. La complejidad del problema con todas sus variables requiere una labor de análisis seria, disciplinada y metodológica.

El proyecto Neptuno [4] aborda el desarrollo de una plataforma de gestión de archivos de material informativo desde la perspectiva de la web semántica, una nueva corriente de innovación en el ámbito de las tecnologías web [5]. Las proporciones actuales de información que se deben gestionar en una hemeroteca se pueden comparar con el contenido de información disponible en la red, y en este sentido, por la similitud del problema, se ha decidido usar la tecnología de la web semántica para la descripción y gestión del contenido de estas hemerotecas como objetivo del proyecto.

Las propuestas de la web semántica introducen técnicas, metodologías, modelos y planteamientos novedosos, tanto para los usuarios como para los desarrolladores, con un énfasis particular en la representación del conocimiento, y la presencia

explícita de la noción de semántica en todos los niveles de los sistemas: infraestructura y lenguajes, representación de contenidos, uso y explotación de la información. La novedad de esta forma de abordar el desarrollo de plataformas de gestión, las consecuencias en cuanto a la forma de entender la interacción con los usuarios, las nuevas complejidades que introduce, y la relativa escasez de experiencias previas, nos ha llevado a considerar la importancia de realizar una labor concienzuda de análisis y comprensión de requisitos previa al desarrollo propiamente dicho. Para ello se han empleado las técnicas y metodologías más avanzadas, basadas en escenarios y casos de uso, a cargo de los expertos en la materia del equipo del proyecto. La utilización de estas técnicas en combinación con los avances de la web semántica es una experiencia novedosa desde ambas perspectivas.

En este artículo se presentan los resultados relativos al uso de escenarios como método de adquisición de requisitos de los diversos agentes implicados en el sistema, siendo la información recogida la base para el desarrollo del primer prototipo de la nueva hemeroteca digital. La relevancia de los factores humanos describen un factor diferencial en este proyecto. Debido a las divergencias entre el modo de ver y de usar el sistema de varios perfiles de usuarios –como los documentalistas y los redactores–, que, por otro lado, deben interactuar conjuntamente sobre las mismas estructuras de información, hacen necesario articular un método que permita reflejar de forma exacta y clara los requisitos de cada grupo de usuarios y permita descubrir nuevos requisitos encaminados a desarrollar una estructura de información que permita esta flexibilidad deseada. En el proyecto Neptuno se han usado los escenarios para tal fin, de forma que con la necesaria evaluación por parte de los usuarios y siguiendo una metodología de desarrollo iterativo, se han logrado definir los requisitos de los diversos perfiles de usuario y encontrar un sistema de representación de la información, basado en una ontología, que responde a las expectativas y objetivos planteados.

A continuación se describe en primer lugar como se realiza la gestión documental en el Diari SEGRE, así como los problemas detectados en su uso diario por parte de los periodistas y los documentalistas. Después se pasa a detallar la metodología utilizada para desarrollar este proyecto y más concretamente qué son y como se deben usar los escenarios dentro del ciclo de desarrollo propuesto. También se hace especial hincapié en el uso de la web semántica para solucionar los problemas presentados y finalmente se muestra cómo se ha aplicado el uso de escenarios para adquirir los requisitos de los usuarios finales del sistema, así como las conclusiones a las que se ha llegado después de esta experiencia aplicada.

2. GESTIÓN DOCUMENTAL EN EL DIARI SEGRE

En la actualidad, la creación de noticias dentro de un grupo mediático como el Diari SEGRE se compone de muy diversas aportaciones de fuentes internas o externas a los como puede ser información proveniente de agencias de noticias, Internet, de los propios implicados, etc. Entre ellas cabe destacar el uso de las hemerotecas de noticias. En ellas se encuentra la información que permite matizar, poner en contexto, relacionar o enfatizar la nueva información que se debe redactar, además de permitir crear opinión o centrar a los redactores, que aunque sean neonatos o no

estén familiarizados en ciertos temas, deben desarrollar informativamente alguna noticia relacionada.

Es por ello que en la mayoría de los medios periodísticos actuales existen grandes repositorios de información basados en un sistema de bases de datos relacional y un cierto software que permite gestionar su contenido. Habitualmente los documentalistas clasifican diariamente los contenidos que aparecen en el diario. En el caso del Diari SEGRE, existe una base de datos con un cierto software de gestión y los documentalistas del periódico archivan las noticias aparecidas cada día. La gestión de la hemeroteca se realiza utilizando la herramienta MILLENIUM ARCANO de la empresa Protec [6], una de los más usadas y populares en el sector en España y Portugal. Los criterios que utilizan para esta clasificación son fundamentalmente dos: la sección a la que pertenece el contenido (Deportes, Sociedad,...) y el tema del que trata. Para esta última clasificación utilizan unos tesauros que han elaborado entre todos los documentalistas y que van actualizando conforme surgen nuevas necesidades.

Las noticias recogidas de la publicación diaria en medios de comunicación representa un enorme volumen de información (p.e. LexisNexis asegura majenar más de 3.300 millones de documentos [7]), muy vagamente organizada (p.e. comparada a una biblioteca de libros), que crece sin una estructura global a priori, a medida que nuevos contenidos se suman y evolucionan impredeciblemente. Este corpus, en constante crecimiento, de noticias archivadas resulta del trabajo coordinado, pero en buena medida autónomo, de un grupo de reporteros cuyo objetivo primario no es construir una hemeroteca, sino su consulta para poder proporcionar el mejor producto informativo posible para el consumo inmediato y que suelen estar asistidos por documentalistas que son los encargados de clasificar, indexar y anotar noticias cuando éstas pasan al archivo, según sus criterios y según las posibilidades que les permite la plataforma software que usen. El departamento de documentación y sus documentalistas usan para la clasificación de los contenidos un tesoro jerárquico donde se sucede una constante evolución de los conceptos que pretende aglutinar y que son ellos mismos quienes se encargan del mantenimiento de este índice.

Además una vez introducido el contenido en nuestro repositorio de información, se necesitan mecanismos potentes de búsqueda y exploración para que los consumidores puedan encontrar su camino en este espacio informativo. Los periodistas también realizan búsquedas cuando se auto asesoran sobre ciertos temas o sucesos, no siempre con la calma que se prevé necesaria para hacerlo, ni con la suficiente información de antemano como para cerciorarse de que buscan algo en concreto. La tecnología actual típicamente proporciona sistemas de búsqueda basada en palabras clave (a menudo por campos: titular, resumen, cuerpo, sección), funcionalidades de exploración dentro de un ejemplar de un periódico y, en ediciones web, navegación a través de hiperenlaces creados a mano entre noticias (p.e. enlaces a antecedentes de una noticia).

Uno de los problemas que se planteó cuando se analizó el trabajo diario de redactores y documentalistas es que el modo en que los redactores buscan la información en la base de datos difiere mucho de la forma en que los documentalistas archivan dicha información. Por este motivo se propuso a los documentalistas utilizar una clasificación basada en la jerarquía de temas de tres niveles [8] (con 17 categorías en el primer nivel) promovida como

estándar de facto por el consorcio internacional IPTC (International Press Telecommunications Council) que agrupa las principales agencias de noticias, editores y distribuidores de periódicos. Estos estuvieron de acuerdo en que el uso de una nueva jerarquía consensuada y especializada en el sector podría ser una solución al problema. Es por todo ello, que la diferencia de los modelos mentales (puesta en evidencia en el transcurso de este proyecto) entre los/las documentalistas y los/las periodistas hacen necesario articular un sistema más flexible de categorización y de búsqueda de contenidos, que es lo que se pretende lograr en el proyecto Neptuno, ayudándonos de las posibilidades que nos ofrece la web semántica.

Para lograr este objetivo es preciso conocer de antemano cuáles son estos modelos mentales de los dos grupos en discordia, o sea, conocer cómo piensa un/a redactor/a cuando desea localizar una noticia y cómo piensa un/a documentalista ante el reto de clasificar una información. Para ello se propone, como se detalla a continuación, aplicar el modelo de proceso MPIu+a y en concreto el uso de escenarios dentro de la fase de análisis de requisitos.

3. METODOLOGÍA DE DESARROLLO UTILIZADA

El contexto de desarrollo de este proyecto se encuentra en la disciplina de la Interacción Persona-Ordenador (IPO) donde los aspectos de la usabilidad y de la accesibilidad del sistema a implementar tienen una destacada consideración. En este sentido se ha tomado como referencia metodológica el modelo de diseño de sistemas interactivos centrado en los usuarios conocido como Modelo de Proceso de la Ingeniería de la Usabilidad y de la Accesibilidad, MPIu+a, [9], [10], el cual puede verse reflejado en la figura 1.

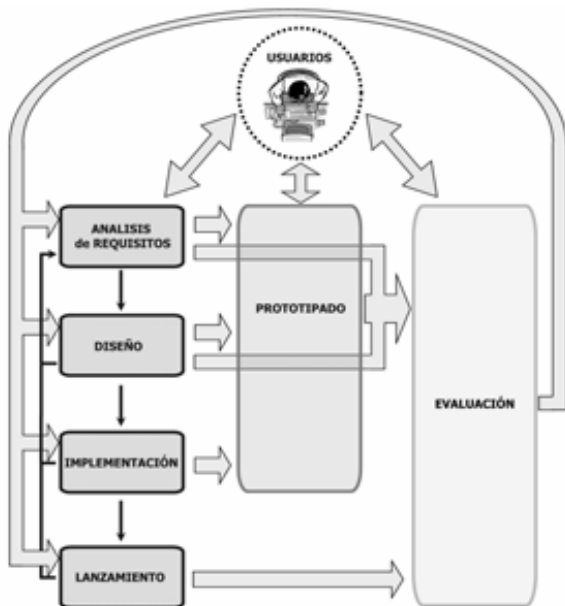


Figura1. Modelo de proceso MPIu+a.

Este modelo de proceso que se fundamenta, como hemos mencionado, en la disciplina de la IPO y también en la Ingeniería del Software, proporciona la manera de proceder organizadamente para conseguir la usabilidad y la accesibilidad en el diseño de

interfaces de usuario durante el desarrollo de un sistema interactivo, tratándose de una materia que es, como resultado de ser una consecuencia de la IPO, multidisciplinar que tiene sus raíces en otras disciplinas básicas (psicología cognitiva, psicología experimental, etnografía, diseño gráfico, etc.).

Las principales características que definen el MPIu+a son:

- es tecnológicamente independiente (se adecua a cualquier cambio tanto tecnológico como de paradigma),
- es aplicable a todo tipo de proyectos, independientemente de su clase y envergadura,
- se adapta a los diferentes modelos mentales de los integrantes de los equipos multidisciplinares,
- es muy simple,
- sigue los principios del Diseño Centrado en el Usuario,
- fomenta el desarrollo de sistemas evolutivo: iterativo e incremental,
- integra la metodología y los formalismos necesarios de la Ingeniería del Software con la de la Usabilidad,
- integra a su vez la Accesibilidad como componente fundamental de todo el proceso, y
- es consistente con los estándares de calidad relacionados.

El modelo de proceso está organizado en una serie de fases que repetidamente se van realizando, siendo el prototipado y la evaluación dos fases claves que se aplican desde el mismo instante que comienza cada nuevo desarrollo.

En relación al proyecto que nos incumbe, durante esta fase inicial, que mayoritariamente se ha centrado en la definición de los requisitos tanto funcionales como de usabilidad del sistema, se ha utilizado la técnica de los escenarios [11] como herramienta que ha permitido conocer la manera de trabajar de los usuarios del sistema y ser capaces así de incrementar el conocimiento acerca del sistema para poder desarrollar un sistema que sea realmente útil para las personas que lo utilizarán.

Utilizar los escenarios ha permitido pasar de la especificación de procesos y funcionalidades en lenguaje natural en forma de narración textual a especificaciones técnicas de flujos de información descritos con la formalidad de la notación UML.

4. LA WEB SEMÁNTICA

La web semántica es un área pujante en la confluencia de la Inteligencia Artificial y las tecnologías web, que propone la introducción de descripciones semánticas explícitas de los contenidos en la web para facilitar su localización, compartición e integración [5]. Se trata de describir los recursos con representaciones procesables no sólo por personas, sino por máquinas que puedan asistir, representar, o reemplazar a las personas en tareas rutinarias o inabarcables para un humano.

La web semántica rescata la noción de ontología del campo de la Inteligencia Artificial como vehículo para alcanzar un entendimiento entre las partes: usuarios, desarrolladores y programas de muy diverso perfil, que participan, comparten recursos, e interactúan en la web. Una ontología es una jerarquía de conceptos con atributos y relaciones, que proporciona un vocabulario consensuado para definir redes semánticas de unidades de información interrelacionadas [12].

Las tecnologías de la web semántica persiguen potenciar la facilidad de extensión de la web, aumentar la granularidad de la

web (permitiendo la integración de componentes a mayor escala), desarrollar una web que se preste más a la automatización de tareas, y que haga posible el funcionamiento de agentes autónomos más sofisticados. Alcanzar estas ambiciosas metas sin perder los principios que han hecho un éxito de la web, como son la distribución, descentralización, compatibilidad, y mínimos requisitos para los clientes, es un problema difícil que abre un campo de investigación interesante y atractivo por su impacto potencial en el desarrollo de la sociedad de la información.

Los inicios de la web semántica se están abordando como una transición suave [13], manteniendo la compatibilidad (al menos inicialmente) con la tecnología actual: navegadores, protocolos, servidores web y de aplicaciones, arquitecturas. En estos momentos el avance de la web semántica está muy necesitado del desarrollo de aplicaciones reales y experiencias piloto que consuman, motiven, y validen la tecnología desarrollada, la hagan madurar, generen masa crítica y pueblen la web con ontologías, para desencadenar la materialización de una web de nueva generación. El proyecto Neptuno representa una contribución en esta dirección.

4.1 Tecnologías de la web semántica en el proyecto Neptuno.

En Neptuno se adoptan las tecnologías de la web semántica como nexo de unión entre las dos vistas tan diferentes que interactúan en la hemeroteca digital. De este modo se pretende que el uso de una ontología de representación de los contenidos permita definir interrelaciones entre las noticias, que permitan anular las concepciones semánticas distintas que tienen los documentalistas y los redactores. Generalizando este concepto nos permitirá desarrollar un sistema que de satisfacción a cualquier usuario final que lo utilice y que permita un uso mucho más intuitivo y eficaz, así como unos métodos de búsqueda avanzada mejorados.

En concreto se ha construido una ontología para representar los archivos documentales del Diari SEGRE utilizando el editor de ontologías Protégé [14]. Algunos de los conceptos de esta ontología corresponden a las tablas de la base de datos existentes, como por ejemplo, Noticia, Foto, Gráfico y Página. Todos estos conceptos son subclases del concepto Contenido.

Además se han incluido tres formas de clasificar un contenido. En una de ellas se clasifican los contenidos de acuerdo al tema siguiendo la jerarquía de clasificación por temas del IPTC. Una clasificación alternativa puede realizarse de acuerdo al género que describe la naturaleza o característica de una noticia o fotografía y no su contenido específico. Por último, un contenido puede clasificarse de acuerdo a palabras clave que lo describen.

Estas tres clasificaciones sustituyen a las clasificaciones por sección y por categoría de tesauros que utilizaban hasta ahora los documentalistas. Se ha establecido una correspondencia entre estas categorías y las que aparecen en la clasificación por temas del IPTC. En algunos casos esta correspondencia no es biunívoca por lo que se permite que un contenido esté clasificado por más de un tema. Se han creado también otros conceptos que no se corresponden directamente con información almacenada en la base de datos actual y entre los que cabe destacar el de RelacionNoticia con el que se establecerían relaciones entre noticias como las de ampliación, antecedente, comentario, etc.

4.2 El factor humano en la web semántica.

En cierto modo la web semántica persigue reemplazar al usuario humano por programas en actividades rutinarias o inabarcables para una persona, o prescindir de la intervención humana en algunos aspectos del desarrollo de sistemas (p.e. semi-automatizando la integración de subsistemas). Esta substitución no significa sin embargo una reducción de la actividad del usuario, sino un cambio en sus funciones. Los cambios en el papel del usuario conllevan una transformación en el modo de entender la interacción del usuario con la web, que afecta tanto al desarrollador como al consumidor y los proveedores.

Las consecuencias de estos cambios para los usuarios, es decir el factor humano en la web semántica, ha sido señalado como una cuestión importante a tener en cuenta. La mayoría de los congresos y talleres internacionales relacionados con la web semántica reflejan esta preocupación, con tópicos directamente relacionados con la Interacción Persona-Ordenador en la propia llamada a la participación, pero hasta la fecha pocos trabajos abordan explícitamente estos aspectos en la práctica.

La toma en cuenta del usuario es relevante en cuestiones como las siguientes:

- ¿Cómo se ha de mostrar la web semántica al usuario?
- ¿Cómo se entiende y cómo se realiza la construcción de la web semántica?
- ¿Cómo representar la información y el conocimiento?
- ¿Cómo interactúan los usuarios entre sí? (Dentro y fuera del sistema.)

El estudio del factor humano ha sido un aspecto central en el desarrollo del proyecto Neptuno, se ha incluido en la propia metodología de desarrollo, aplicándolo de manera sistemática y formal. En este sentido el trabajo realizado implica la contribución de un experimento de aplicación de las metodologías de la usabilidad, llevadas al campo de la web semántica. Esta pauta metodológica se ha traducido en aspectos como el análisis de modelos mentales y su reflejo en la ontología, la comprensión de las necesidades del usuario, guiar el desarrollo de las funcionalidades basadas en semántica, comprender qué capacidades son realmente útiles, cómo se deben ofrecer al usuario, y cuáles se deben descartar.

5. ANÁLISIS DE REQUISITOS

Trabajar con ordenadores y sistemas interactivos en general es algo más que funcionalidades: reestructuran actividades humanas creando a la vez nuevas posibilidades y dificultades. Por otra parte en cada contexto en que el ser humano tiene experiencia y actúa proporciona unas restricciones para el desarrollo de sistemas de información

El trabajo presentado en este artículo está ubicado en la fase inicial del Modelo de Proceso, el análisis de los requisitos. Y cuando nos disponemos a analizar para diseñar software necesitamos una manera de ver como estos nuevos sistemas pueden transformar los contextos actuales de la actividad humana a la vez que pueden verse restringidos por ellos.

Una aproximación directa es imaginando y documentando las actividades típicas y significativas. Describiendo así, las actividades que se realizan durante el proceso de desarrollo, y los

escenarios son una forma apropiada de presentar estas descripciones.

Considerando que un *usuario* de un sistema interactivo es aquella persona que interactúa con el sistema, lo controla directamente y que utiliza los recursos (información, resultados, etc.) del mismo [15] y un *implicado* es cualquier persona el trabajo de la cual será alterado, aquel que proporciona u obtiene información de ello, o cuyo poder o influencia dentro de la organización incrementará o disminuirá [16], los escenarios se han desarrollado teniendo como actores tanto a usuarios representativos del sistema como de los principales implicados en el mismo y lograr así una hemeroteca digital útil y que responda a necesidades reales.

Observando sus reacciones, comentarios, aptitudes, etc., esto nos ha proporcionan uno de los principios vitales de la Ingeniería de la Usabilidad: testear pronto y a menudo [17] centrandó el proceso en el usuario.

En este sentido, previamente se han realizado entrevistas exhaustivas con todos los implicados captando sus conocimientos, estudios, aptitudes y actitudes, opiniones respecto al actual sistema de hemeroteca, la visión de su sistema ideal e incluso datos más personales que nos permitan definir la personalidad de los actores. De este estudio previo han surgido dos perfiles de usuario básicos y muy definidos: los redactores y los periodistas por un lado y los documentalistas por otro. Con estos perfiles y realizando grabaciones en video del trabajo diario de varios implicados se elaboran varios escenarios que tratan de revelar ciertas características controvertidas del actual sistema, así como nuevas funcionalidades que los usuarios nos han profetizado como ideales del sistema. Tras contrastar estos escenarios con los implicados de los diversos perfiles, se obtienen los primeros requisitos del nuevo sistema. Finalmente, y siguiendo el carácter iterativo del modelo MPIu+a usado, se ha desarrollado una segunda implementación de escenarios con la intención de captar los modelos mentales de los usuarios respecto a sus métodos de búsquedas de información, tratando en todo momento que se abstraieran del sistema actual. De esta segunda fase se han extraído valiosos requisitos funcionales y no funcionales del sistema que sin duda deben ser considerados en las fases posteriores de diseño e implementación.

5.1 Perfil de los Usuarios.

Una vez realizadas las entrevistas previas, se definieron los diversos perfiles de usuario necesarios como un primer paso dentro del modelo MPIu+a para poder luego utilizar estos perfiles como "input" en el resto de fases del modelo centrado en el usuario [18]. En el Diari SEGRE nos encontramos dos tipos de usuarios claramente diferenciados que se ven implicados en la gestión documental. Por una parte los documentalistas, que son personas de entre 30 y 40 años, licenciados en Historia Contemporánea mayoritariamente. Todos tienen más de 7 años de experiencia en su trabajo. En cuanto a su conocimiento de informática coinciden en tener un nivel medio bajo en el uso de ordenadores, conocen de 3 a 10 programas y no tienen nociones de programación. Las aplicaciones con las que trabajan diariamente son, entre otras, Millennium Editor (aplicación para visualizar el diario), Millennium Arcano (aplicación para documentar la información) y Microsoft Word. No les motiva descubrir nuevas aplicaciones ni aprender la plena funcionalidad de las que conocen, y consideran finalizado el aprendizaje de las mismas una vez el conocimiento adquirido les permite llevar a

cabo su trabajo. Su trabajo es muy especializado: una persona se encarga de documentar las fotografías, otra documenta las noticias y otra documenta las páginas (tal y como aparecen en el diario físico día a día). Alguna vez también llevan a cabo búsquedas de información que los redactores del diario les piden hacer al no encontrar ellos la información deseada. En su trabajo diario los documentalistas se preocupan realmente por buscar, para cada documento, un campo semántico en el que encajarlo, de forma que esta unión sea lo más lógica posible, pero coinciden en manifestar que el modo en que los redactores buscan la información en ARCANO poco tiene que ver con la forma en que ellos la documentan. Creen que este problema se debe a la poca comunicación entre documentalistas y redactores y en la forma en que los redactores buscan (según su semántica) los contenidos almacenados previamente según la semántica y el modo de operar de los documentalistas.

Por otra parte, los redactores y periodistas son los que realizan el trabajo de redacción de las noticias. Gran parte de ellos son licenciados en periodismo pero también nos encontramos licenciados en filología hispánica y redactores sin titulación universitaria. Su experiencia laboral oscila entre los dos años y más de 17. Normalmente están asignados a una de las diferentes secciones y no suele cambiar durante su vida laboral aunque si realizan trabajos de colaboración. En cuanto a los conocimientos informáticos son moderadamente bajos, tienen interés para conocer la funcionalidad de las aplicaciones que manejan en el trabajo así como por conocer nuevos programas. Diariamente suelen trabajar con Millennium Arcano y Microsoft Word, entre otros. El modelo mental del redactor al realizar las búsquedas es que, utilizando el campo de búsqueda "cualquier" (en lugar de los campos título, descripción, localización o fecha), el motor realiza un recorrido exhaustivo de toda la información almacenada en la BD. Por otro lado, les parece increíble cómo clasifican ciertas noticias los documentalistas, llegando incluso, a hacer imposible o improbable que las encuentren.

5.2 Uso de Escenarios.

En el proyecto Neptuno, se pretende que la representación del uso de un sistema o aplicación con un conjunto de escenarios de interacción, nos permita explicitar este uso y por tanto, que oriente en el análisis y posterior diseño del sistema.

Los primeros escenarios propuestos estaban encaminados a lograr reflejar los siguientes flujos de trabajo: a) flujo de información en el proceso de documentación de una noticia y su fotografía asociada, que fue publicada, plasmando el modelo semántico de los documentalistas, más que detallando el procedimiento completo que siguen con la aplicación para conseguir realizar su trabajo; b) reflejar la forma en que un documentalista realiza la búsqueda de una noticia y su fotografía asociada ya documentada y almacenada en la base de datos; c) reflejar como un fotógrafo busca una fotografía, que se publicó hace años en el diario, en demanda de un redactor que la necesita rápidamente; d) reflejar el flujo semántico en una búsqueda general realizada por un redactor de noticias publicadas en relación con un tema concreto y sin prisas; y e) reflejar como un documentalista busca una noticia almacenada en la base de datos y documentada por él mismo tiempo atrás.

A continuación se muestra un breve esquema general, a modo de ejemplo, de uno de los escenarios propuestos en esta primera fase:

Escenario: Documentación de una noticia

Objetivo: Reflejar el flujo de información en el proceso de documentación de una noticia, que fue publicada y plasmar el modelo semántico de los documentalistas.

Tareas:

- La documentalista de noticias visualiza en el sistema de documentación la noticia a documentar.
- La documentalista de noticias revisa el contenido de los campos del formulario que aparecen con información parametrizada en el sistema y rellenos por defecto, cambiando lo que no encuentra adecuado.
- La documentalista de noticias rellena el resto de los campos del formulario con la información de documentación (localizaciones, conceptos semánticos, etc.) que intuye más adecuada.
- La documentalista de noticias guarda los cambios y así la noticia se almacena en el sistema quedando asociada a la información adicional.

De estos escenarios, una vez evaluados con los implicados, usando los métodos de observación de campo, pensar en voz alta (*thinking aloud*) y varios *focus groups* (métodos introducidos por Nielsen, [19]), emergen las primeras conclusiones. La más destacada que se ha podido visualizar en este proceso, es el distanciamiento entre los modos de clasificar semánticamente la información. Difiere mucho la terminología utilizada por los redactores especializados en ciertas secciones y los conceptos o jerarquías en que se fundamenta el trabajo de los documentalistas. Para mitigar o acercar lo máximo posible estos dos modelos mentales, semánticos y lingüísticos, se propone el uso de la jerarquía IPTC.

Como hemos descrito anteriormente, se realizó una segunda fase de elaboración y evaluación de escenarios, centrandó ahora el objetivo en captar los procesos de raciocinio de los usuarios independientemente del sistema actual y intentando visualizar un posible puente entre el abismo que separa los dos modelos mentales de los dos perfiles estudiados basándonos en el uso común de la nueva jerarquía de clasificación. Así, los escenarios descritos en esta fase se fundamentan en proponer a varios documentalistas y a diversos redactores, la clasificación sobre la nueva jerarquía, del mismo conjunto predefinido de fotos y noticias de carácter claramente ambiguo. Del mismo modo se les proponen situaciones de búsqueda de información (tanto noticias como fotos) en la jerarquía, visualizando que procesos de búsqueda mental realizan, apoyándonos de nuevo en los métodos de pensar en voz alta y varios *focus groups*, prestando ahora especial interés en la primera idea intuitiva que se les revela como la más efectiva si el sistema pudiera recogerla y ejecutarla.

La conclusión general a partir de la observación de los escenarios aquí descritos es que el proceso de la documentación se lleva a cabo de forma minuciosa, utilizando todos los recursos semánticos que la aplicación utilizada ofrece. En cambio, no se observa una metodología determinada de documentación, es decir el documentalista aplica un criterio personal para decidir cual es el mundo semántico al que ha de pertenecer el documento. Como consecuencia de esto, nadie más que los documentalistas puede recuperar la información semánticamente y todos los otros usuarios recurren a la búsqueda masiva y al descarte o inspección

de grupos muy grandes de información. Por otro lado el uso de una nueva jerarquía, muy especializada en este entorno parece que responde a las expectativas de los dos perfiles, aumentando en gran medida las coincidencias, usando esta jerarquía como referencia, en la clasificación y búsqueda de la información. Otro requisito observado de los escenarios es la necesidad de que el sistema de respuesta rápida y concisa a las búsquedas de información siendo ágil y preciso, aunque también debe ser capaz de relacionar otras noticias asociadas con la principal de forma indirecta dando como resultado un conjunto mucho más amplio de información pero al mismo tiempo mucho más rico en matices, siendo difuso, aspecto muy valorado en caso de disponer de tiempo para elaborar una nueva noticia.

6. CONCLUSIONES

Como resultado de la realización del proyecto Neptuno se ha desarrollado una plataforma cuyas principales componentes son:

- Una ontología para noticias de archivo, basada en la experiencia y práctica de los periodistas y documentalistas, e integrando estándares dominantes del consorcio IPTC.
- Una base de conocimiento donde el material de archivo se describe utilizando la ontología. Un módulo de mapeo de base de datos a la ontología integra automáticamente el material antiguo de la hemeroteca en la base de conocimiento.
- Un módulo de búsqueda semántica que permite expresar necesidades informativas en términos de la ontología, y proporcionar respuestas más precisas o más amplias.
- Un módulo de visualización y navegación para a) mostrar elementos individuales del archivo, partes o combinaciones de elementos, y grupos de elementos, y b) proporciona funcionalidades de navegación semántica basadas en enlaces inferidos automáticamente entre materiales.
- Un módulo de personalización para mejorar la búsqueda y adaptar la navegación a cada usuario.

Cabe destacar que el uso del modelo de proceso MPIu+a centrado en el usuario se muestra efectivo, eficaz y eficiente en el desarrollo de sistemas interactivos como este. Su carácter iterativo y el hecho de estar centrado en el usuario permite involucrarlo en todo el proceso de desarrollo, incluso sin que él sea consciente de ello, logrando crear un producto software que colme sus expectativas.

En este sentido, el uso de escenarios se ha mostrado como un método válido para extraer requisitos del sistema de forma estructurada y subyacente a la actividad de los implicados. Para que los escenarios sean realmente útiles es muy importante observar primero el trabajo de los usuarios, comprender los mecanismos de interacción que se producen y qué los motivan. En este caso se ha procedido a grabar en video y analizar, en directo primero y luego en diferido, los usuarios in-situ, de modo que se han descrito escenarios totalmente focalizados a los puntos clave de dónde se debían extraer los requisitos. Llegados a este punto, se puede concluir que la evaluación de estos escenarios por parte de los implicados es muy sencilla y intuitiva, suscitando sus reacciones reales ante el caso descrito y vislumbrando así sus necesidades tanto funcionales como no funcionales.

Finalmente, en este proyecto se demuestra la utilidad y la viabilidad del uso de la denominada web semántica en entornos

como el aquí mostrado, usando las estructuras ontológicas de la información como nexo de unión y como arquitectura que soporta nuevos sistemas más ricos y avanzados de búsqueda y consolidación de información.

7. AGRADECIMIENTOS

El trabajo presentado en este artículo está financiado por el Ministerio de Ciencia y Tecnología (programa PROFIT), proyecto nº FIT-150500-2003-511.

8. REFERENCIAS

- [1] Gil, Q. *Diseñando el periodista digital*. Revista electrónica Sala de Prensa, nº 13, noviembre 1999. Disponible en <http://www.saladeprensa.org>.
- [2] <http://www.mynewsonline.com/>
- [3] <http://www.diarisegre.com/>
- [4] <http://pegasus.ii.uam.es/~neptuno/>
- [5] T. Berners-Lee, J. Hendler, O Lassila. The Semantic Web. *Scientific American*, May 2001.
- [6] <http://www.protec.es/>
- [7] <http://www.lexisnexis.com/>
- [8] <http://www.iptc.org>
- [9] Granollers, T. *User Centred Design Process Model. Integration of Usability Engineering and Software Engineering*. INTERACT 2003 (Doctoral Consortium), Zurich (Switzerland), September 2003.
- [10] Granollers, T.; Lorés, J. Perdrix F. *Usability Engineering Process Model. Integration with Software Engineering*. Long paper. HCI International 2003. Crete (Greece), June 2003.
- [11] Carroll, J. *Making use: Scenario-based design of human-computer interactions*. MIT Press., Cambridge, MA (2000)
- [12] Gruber, T. R. *A Translation Approach to Portable Ontology Specifications*. *Knowledge Acquisition*, 5(2), pp. 199-220, 1993.
- [13] Haustein, S.; Pleumann, J. *Is Participation in the Semantic Web too Difficult?*. International Semantic Web Conference (ISWC'2002). Sardinia, Italy, 2002.
- [14] <http://protege.stanford.edu/>
- [15] Sharp H.; Finkelstein A.; Galal G. *Stakeholder Identification in the Requirements Engineering Process*. (December 1999).
- [16] Dix A., Finlay, J., Abowd G., Beale R. *Human Computer Interaction*. Prentice Hall, Englewood Cliffs, NJ (1st edition, 1993)
- [17] Redish J., Bias R.G., Bailey R., Molich R., Dumas J., Spool J.M., *Usability in Practice: Formative Usability Evaluations – Evolution and revolution, Proceedings of CHI2002* (Minneapolis, Minnesota, April 20 - 25, 2002)
- [18] Mayhew, D. J. *The Usability Engineering Lyfecycle. A practitioner's handbook for user interface design*. Morgan Kaufmann Publishers (1999)
- [19] Nielsen J. *Usability Engineering*. Academic Press, Pág. 195-198, 1993