

Universidad de la República – Facultad de Ingeniería

Instituto de Computación

Año 2004

Informe de Proyecto de Grado:

Integrador de Contenidos

***Estudiantes:
Joaquín Arambarri
Pablo Berger***

***Tutor:
Juan José Prada***

***Empresa responsable:
El País Digital – El País S.A.***

***Usuario responsable:
Guillermo Pérez Rossel.***

RESUMEN

Este proyecto fue presentado por la empresa El País S.A. con la finalidad de afrontar la dificultad que presentan los lectores para acceder a noticias de su interés debido a la gran cantidad de información existente en el mundo de hoy.

Este problema empeora con las facilidades que proporciona Internet para la publicación de contenidos. Los periódicos digitales por si solos no aportan una solución para esta problemática, por lo que se plantea el desarrollo de un sistema que permita realizar una integración de contenidos provenientes de múltiples periódicos. Adicionalmente se desarrolla un sistema capaz de ofrecer a los usuarios de El País un resumen personalizado de noticias de acuerdo al interés de los mismos.

La implementación se lleva a cabo en la plataforma de desarrollo .NET. Esto permite integrar un ambiente web con uno orientado a servicios como lo son los XML web services. Una plataforma con estas características brinda la posibilidad de integrar estos dos ambientes de forma que cooperen entre sí para desarrollar un sistema genérico, escalable y extensible sin perder robustez ni confiabilidad.

Al momento de integrar contenido de varios periódicos es natural encontrar diferentes categorizaciones de noticias por lo que se propone la creación de una categorización propia que incluya a las anteriores. Cada periódico que se registra en el sistema debe definir una correspondencia entre las categorías que utiliza para agrupar sus contenidos y esta categorización

Debido a que se dispone solamente de un avance y link de cada noticia, para la realización de búsquedas sobre contenidos integrados se crea un sistema basado en "criterios de búsqueda" que se forman a partir de preferencias de consumidores de los contenidos integrados.

La implementación resultante satisface los requerimientos propuestos y de acuerdo a las pruebas realizadas se observan buenos resultados, con lo que se puede concluir que se logró el objetivo de integración planteado.

Palabras claves:

Integrador de Contenidos, noticias, integración, periódicos, categorización, búsquedas, perfiles de usuario.

CONTENIDO

Capítulo 1 - INTRODUCCIÓN.....	6
1.1. Motivación.....	6
1.2. Objetivos.....	6
1.3. Cronograma.....	7
1.4. Organización del documento.....	8
Capítulo 2 - MARCO TEORICO.....	9
2.1. Modelo de datos.....	9
2.2. Metadatos.....	10
2.3. Semántica.....	10
2.4. Sobrecarga Semántica.....	11
2.5. Potencia Semántica.....	11
2.6. Sistemas de múltiples bases de datos.....	12
2.7. Autonomía.....	12
2.8. Interoperabilidad.....	13
2.9. Arquitectura de Bases de Datos Heterogéneas Distribuidas.....	14
2.9.1. Arquitectura de "wrappers y mediadores".....	14
2.9.2. Integración en un Esquema Global.....	15
2.9.3. Sistemas de Bases de Datos Federados.....	16
Capítulo 3 – DESARROLLOS SIMILARES Y HERRAMIENTAS UTILIZADAS	17
3.1. Desarrollos similares.....	17
3.2. Contraste del IC frente a estos desarrollos.....	19
3.3. Herramientas utilizadas – Tecnología .NET.....	20
Capítulo 4 - DEFINICIONES Y ANÁLISIS DEL PROYECTO.....	23
4.1. Definición de alcance.....	23
4.2. Análisis de requerimientos	23
4.2.1. Requerimientos funcionales.....	23
4.2.2. Requerimientos no funcionales.....	25
4.3. Casos de Uso.....	25
4.3.1. Diagrama de Casos de Uso IC	25
4.3.2. Detalle de Casos de Uso.....	27
4.3.3. Diagrama de Casos de Uso Consumidor IC para ELPD.....	35
4.3.4. Detalle de Casos de Uso.....	36
4.3.5. Flujo de información.....	37
4.3.6. Diagramas de Secuencia.....	38
Capítulo 5 – DISEÑO Y CONSTRUCCIÓN.....	42
5.1. Decisiones tecnológicas	42
5.2. Decisiones de Diseño.....	43
5.3. Decisiones de implementación.....	48
5.4. Subsistemas del IC.....	49
5.5 Ambiente Web.....	51
5.5.1 Front-End.....	51
5.5.2 Back-End.....	52
5.6 Ambiente de Servicios.....	53
5.6.1 Fuentes - IC.....	53
5.6.2 Consumidores - IC.....	55
Capítulo 6 – PRUEBAS Y RESULTADOS OBTENIDOS.....	58
6.1. Pruebas IC.....	58

6.2. Pruebas de trasferencias e integración.....	59
7.1. Conclusiones.....	63
7.2. Pasos a seguir.....	64
8.1. Glosario.....	66
8.2. Referencias.....	67

Capítulo 1 - INTRODUCCIÓN

1.1. Motivación

El mundo actual se nos presenta como un océano de información en el cual nos encontramos navegando y recibimos continuo bombardeo de noticias. Desde la televisión, diarios, revistas y hasta los medios digitales tan de moda últimamente como son el correo electrónico, grupos de noticias, sitios Web, etc. se presentan como origen de un flujo continuo al cual nos tenemos que enfrentar tratando de concentrar nuestra limitada capacidad de atención en aquello que presenta algún atractivo o utilidad ya sea en forma personal o profesional.

No solo es complejo tratar de discernir que es lo que nos interesa sino que también se vuelve difícil la búsqueda específica. Para esto tenemos que leer y decidir que información descartar hasta llegar al punto deseado (si es posible ya que también podemos cometer errores al descartar algo que podría haber sido de utilidad).

Parece necesario contar con una herramienta que conozca el interés de los usuarios para poder poner a disposición de los mismos un resumen de noticias que sea accesible en el momento deseado y sin tener que descartar o leer información que no es de su interés y así evitar pérdidas de tiempo y desgaste al realizar dicha tarea.

Esta herramienta sería de utilidad a nivel nacional pero mucho mayor impacto y utilidad es la que tendría si se aplica internacionalmente, manejando información de múltiples fuentes y países. El valor de la misma se puede apreciar desde muchos puntos de vista como ser económico, legal, político, informativo, etc. Un ejemplo de ello podría ser en el ámbito agropecuario, donde una persona vinculada a la ganadería podría disponer de un panorama claro y actualizado a nivel regional.

1.2. Objetivos

Como objetivo principal se plantea desarrollar una herramienta que permita realizar la integración de contenidos provenientes de periódicos. Para esto se considera el desarrollo de un sistema que integre los contenidos proporcionados por los periódicos participantes.

Adicionalmente se espera que el sistema sea capaz de ofrecer a los usuarios de El País Digital (ELPD: sección de El País S.A. que presenta el proyecto) un resumen personalizado de noticias de acuerdo al interés de los mismos, exhibiéndolas en formato web según las especificaciones de la empresa.

1.3. Cronograma

Se plantea como cronograma inicial el desarrollo de las siguientes tareas:

- Interiorización con el problema a resolver.
- Definición de los requerimientos.
- Relevamiento de esquemas y categorización según las fuentes.
- Diseño del sistema.
- Validar con usuarios de ELPD que los requerimientos establecidos para el sistema estén de acuerdo con las necesidades especificadas.
- Implementar prototipo.
- Evaluación del prototipo.
- Implementación final

De acuerdo al cronograma planteado, se detalla en la Figura 1 mediante un gráfico de Gantt la precedencia y la duración de las tareas a medida que se desarrolló le proyecto.

Con esto se representa que el desarrollo del proyecto se realizó en forma normal con pequeños ajustes sobre lo planteado como cronograma inicial.

Dentro de la etapa de implementación se realizaron prototipados y varias validaciones con personal de diseño gráfico y usuarios de ELPD de forma de ir implementando en forma gradual sin tener que realizar reestructuras o rediseños grandes que podrían causar demoras en el cronograma.

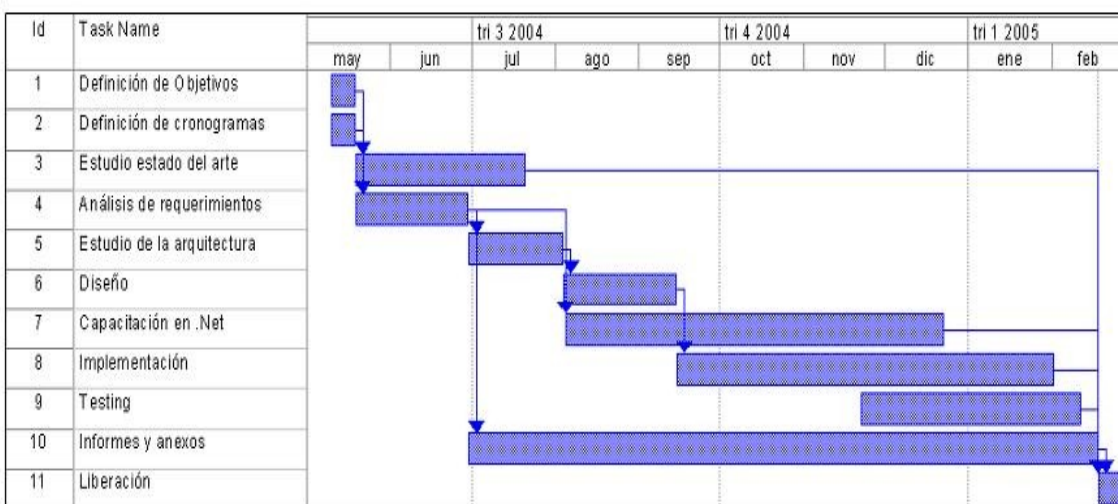


Figura 1: Gráfico Gantt para cronograma de proyecto.

Tarea	Descripción	Duración	Comienzo	Fin	Previa
1	Definición de Objetivos	7d	lun 10/05/04	mar 18/05/04	
2	Definición de cronogramas	7d	lun 10/05/04	mar 18/05/04	
3	Estudio estado del arte	45d	mié 19/05/04	mar 20/07/04	1,2
4	Análisis de requerimientos	30d	mié 19/05/04	mar 29/06/04	1,2
5	Estudio de la arquitectura	25d	mié 30/06/04	mar 03/08/04	4
6	Diseño	30d	mié 04/08/04	mar 14/09/04	5
7	Capacitación en .Net	100d	jue 05/08/04	mié 22/12/04	4
8	Implementación	100d	Mié 15/09/04	mar 01/02/05	6
9	Testing	60d	lun 22/11/04	vie 11/02/05	
10	Informes y anexos	167d	mié 30/06/04	jue 17/02/05	4
11	Liberación	7d	vie 18/02/05	lun 28/02/05	3,7,8,10,9

1.4. Organización del documento

En el Capítulo 2 del documento se presentan fundamentos teóricos sobre temas relacionados al manejo de datos e información vinculada a la integración de contenidos.

En el Capítulo 3 se hace un análisis de sistemas relacionados a la integración de contenidos. También se analizan las semejanzas y diferencias de estos con el sistema a desarrollar y se presenta la tecnología .NET como entorno de desarrollo a ser utilizado.

En el Capítulo 4 se define el alcance para el sistema a desarrollar y se detallan los requerimientos identificados en base a los objetivos propuestos.

En el Capítulo 5 reúne las decisiones de diseño más significativas y se detallan algunos aspectos específicos del sistema. Se presenta el mismo como la composición de subsistemas y para cada uno de ellos se brinda una breve descripción.

En el Capítulo 6 se documentan las pruebas realizadas y resultados de las mismas.

En el Capítulo 7 se realiza una evaluación de los resultados obtenidos teniendo en cuenta los objetivos planteados y se proponen pasos a seguir.

En el Capítulo 8 se encuentra el glosario de términos y la bibliografía utilizada.

Capítulo 2 - MARCO TEORICO

2.1. Modelo de datos

Conocimiento es el entendimiento de la naturaleza, cualidades y relaciones de las cosas.

La constante adquisición de conocimiento constituye una de las características esenciales de nuestra naturaleza humana.

Permanentemente buscamos aumentar nuestro conocimiento acerca de nosotros mismos y del ambiente que nos rodea.

Cuando percibimos u observamos un fenómeno, obtenemos una porción incremental de conocimiento. A estos incrementos de conocimiento denominamos "información".

La información valiosa tiene que ser registrada y también comunicada a otras personas en una forma simbólica que pueda ser entendida. Para estos fines resulta de importancia fundamental contar con una adecuada representación simbólica de la información.

La pieza más elemental de información es el "dato" (por ejemplo una cifra de sueldo). Un dato elemental no es de utilidad por sí mismo, es útil sólo cuando está vinculado con otros datos (por ejemplo la identificación de un empleado), por lo que el almacenamiento y la comunicación de información requieren la construcción de "estructuras de datos".

La estructura y el contexto le dan "significado" a los datos posibilitando su entendimiento. El lenguaje natural es nuestra forma primaria de representación y comunicación de información. Si bien es el medio más generalizado para representar y comunicar información, no siempre constituye el mejor medio debido a la ambigüedad que este presenta.

En muchas situaciones resulta más útil establecer formas especializadas de representación de la información. Las fórmulas matemáticas, los mapas carreteros y las partituras musicales, constituyen ejemplos de "modelos" especializados de representación simbólica de la información.

El almacenamiento, procesamiento y distribución de información mediante computadores y redes de comunicación ha creado la necesidad de encontrar modelos de representación adecuados a estos fines. Se han explorado y desarrollado diferentes "modelos de datos" (Relacionales, Jerárquicos, de Red, Semánticos, etc.) para la representación de la información en los sistemas computarizados [1].

El objetivo de la modelización de datos es el conocimiento de los datos que se van a manejar y su agrupación en entidades. El modelo de datos debe ser una fiel representación del sistema de información objeto de estudio.

El contenido del modelo de datos debe representar el estado final al que quiere llegar el sistema. Cualquier cambio en el sistema de información se debe reflejar en el modelo y viceversa.

En el modelo de datos debe aparecer representada toda parte lógica de la información necesaria. Por ello, es independiente del entorno físico y debe proporcionar a los usuarios toda la información que necesitan.

Por tanto, podríamos decir que el objetivo fundamental del modelo de datos es la obtención de estructuras no redundantes, sin inconsistencias, seguras e íntegras.

Los modelos de datos poseen diferentes niveles de abstracción a la hora de representar datos. Los de mayor nivel de abstracción y potencia semántica se orientan al diseño de las estructuras de datos que compondrán las bases de datos. Este tipo de modelo de datos recibe la denominación de modelo conceptual o semántico.

2.2. Metadatos

Los metadatos, literalmente “datos sobre datos”, consisten en información que describe el contenido, calidad, condiciones y otras características de datos. Permiten y ayudan al entendimiento total de información almacenada. Como ejemplo de metadatos se podrían nombrar: el catálogo de una biblioteca, la lista de canciones de un CD de música impreso en su cara posterior, etc.

La utilización de metadatos es muy amplia, se puede ver su uso frecuentemente en sistemas de gestión y administración de recursos de información, para describir o identificar recursos de información, para describir funcionamiento de sistemas, para describir niveles y clasificar tipos de uso de recursos, etc.

2.3. Semántica

Relativo al significado de las palabras [3].

Cuando se intenta almacenar información que representa algún aspecto de la realidad, es necesaria la utilización de modelos de datos orientados a la descripción de la realidad que esta representada en el repositorio de datos. Estos modelos se denominan modelos “conceptuales” o “semánticos” y se orientan a la descripción de la realidad captando en mayor medida el significado de los datos y sus vinculaciones con el mundo real.

2.4. Sobrecarga Semántica

En la actualidad, puede que los sistemas de almacenamiento de datos no respondan a modelos de elevado nivel de abstracción como los que se utilizan durante el diseño, por lo que es necesario traducir el modelo conceptual al modelo de almacenamiento que se utilizará (por ejemplo el modelo relacional).

Esta traducción se hace con pérdida de información semántica. Se dice que el modelo de datos de menor nivel de abstracción resulta "sobrecargado semánticamente" ya que al contar con una menor variedad de elementos de modelización cada elemento debe representar varios elementos del modelo de mayor nivel de abstracción y potencia semántica [1]. Los modelos de datos en general especifican:

- La estructura de los datos.
- Restricciones sobre la estructura.
- Operaciones sobre la estructura.

Un modelo de datos establece reglas para la estructuración de los datos (mediante tablas, árboles, redes, etc.), define las operaciones permitidas sobre las estructuras (en general para extraer nuevas estructuras de las existentes), y permite establecer restricciones sobre las estructuras para que los datos mantengan su coherencia informativa o "integridad semántica". Estas restricciones pueden abarcar una gran cantidad de aspectos. Por ejemplo, pueden especificar restricciones provenientes del mundo real tales como:

- Que los empleados no puede ganar más que sus jefes.
- Que para inscribirse en un curso, un estudiante debe satisfacer ciertos prerrequisitos.
- Que la cantidad de pacientes internados en un hospital no debe ser mayor que la cantidad total de camas existente.

Además, de restricciones explícitas, como las anteriores, un modelo de datos introduce restricciones implícitas como consecuencia de sus propias características y limitaciones de modelización.

2.5. Potencia Semántica

En la sección anterior se introdujo, sin definirlo, el concepto de "potencia semántica". La potencia semántica de un modelo está determinada por su capacidad de representar no sólo la estructura de los datos sino también "la interpretación o significado" de los mismos y de sus interrelaciones.

2.6. Sistemas de múltiples bases de datos

Organizaciones distribuidas por todo el mundo cuentan con una amplia variedad de sistemas de bases de datos para manejar información vinculada a sus actividades diarias. Se utilizan variadas plataformas y modelos de datos.

Con el transcurso de los años y avances de tecnologías y herramientas para almacenar y manejar datos también crecen las dimensiones de la información que se maneja. A medida que las organizaciones se sofistican, surge la necesidad de compartir información. Al mismo tiempo, avances en sistemas computacionales interconectados a través de redes, combinado con razonables niveles de conectividad, ayudan a una realidad donde se comparte información a través de bases de datos heterogéneas. Cuando nos referimos a heterogeneidad hablamos de que la información se puede almacenar en diversidad de formas. A modo de ejemplificar en forma simple, un sistema podría almacenar información en archivos de texto plano sin ningún tipo de estilo mientras que otro podría utilizar un sistema de páginas web con toda la dinámica que esto implica.

El desafío subyacente a todo esto es brindar a los usuarios que utilizan estos sistemas la sensación de que acceden a una base de datos simple que contiene prácticamente todo lo que necesitan mientras que se preserva la integridad de los sistemas preexistentes. En estos sistemas, es necesario, no solo proveer acceso uniforme a toda la información y recursos sino que también hay que permitir que las bases de datos cooperen a través de intercambio de datos y sincronizando ciertos procesos de ejecución.

Esta necesidad creciente de proveer a organizaciones recursos de datos y software en forma amplia crea la demanda de interconectar sistemas de software que anteriormente se encontraban aislados. Un usuario en un ambiente de computación heterogéneo debería estar habilitado no solo para invocar múltiples sistemas de software y dispositivos de hardware existentes sino para coordinar su interacción. Estos sistemas pueden correr en forma autónoma en diferentes computadoras, pueden soportar diferentes sistemas operativos, pueden haber sido designados para diferentes propósitos y pueden utilizar formatos de datos diferentes.

2.7. Autonomía

Las entidades organizacionales que gerencian diferentes sistemas de bases de datos son a menudo autónomas. En otras palabras, las bases de datos están a menudo controladas en forma separada e independiente. Quienes controlan una base de datos pueden aceptar compartir datos pero con la condición de continuar manteniendo el control de sus bases de datos.

A continuación se lista un resumen de diferentes aspectos que implica la autonomía [2]:

- *Autonomía de diseño:* Las bases de datos locales eligen el modelo de datos con el que van a trabajar, como también el lenguaje, la interpretación

semántica de los datos, restricciones, funciones/operaciones soportadas, etc.

- *Autonomía de comunicación:* Las bases de datos locales deciden cuando y como responder a las solicitudes desde otras bases de datos.
- *Autonomía de ejecución:* El orden de ejecución de transacciones u operaciones locales/externas es controlada por el propio sistema de base de datos. Las bases de datos locales no necesitan informar a otros sistemas de bases de datos (DBMSs) del orden de ejecución local o de operaciones externas.
- *Autonomía de asociación:* Las bases de datos locales pueden decidir cuanto de sus funciones/operaciones y datos compartir con cierta clase de usuarios. La liberación de información estadística, como ser costos, eficiencia, y velocidad de ejecución para procesar información, es también determinado para cada base de datos en forma individual; lo que hace más difícil poder realizar optimizaciones y consultas a nivel global entre múltiples sistemas autónomos. Cada base de datos local tiene la habilidad de asociarse o desasociarse de la red de bases de datos.

2.8. Interoperabilidad

Interoperación implica la habilidad de solicitar y recibir servicios y funcionalidades entre los sistemas que interoperan. Una forma limitada de interoperación es aquella donde el intercambio de datos entre sistemas debe realizarse en forma periódica enviando datos desde un sistema a otro. Interdependencia implica que datos y funciones en diferentes sistemas están relacionados o dependen unos de otros, aunque el usuario o aplicación final puede no estar enterado de este relacionamiento de dependencia. De este modo, el gerenciamiento de la interdependencia de datos implica reforzar la consistencia de restricciones de múltiples bases de datos. En forma genérica, se considera que sistemas de información son interoperables si se cumplen las siguientes condiciones:

- Pueden intercambiar mensajes y solicitudes.
- Pueden recibir servicios y operar como unidades en la resolución de problemas comunes.

Estas condiciones sugieren que para que sistemas de información sean interoperables, deben tener las siguientes características:

- Utilizar funcionalidades entre ellos.
- Habilidades de cliente-servidor.
- Comunicación a pesar de la incompatibilidad de detalles internos de los componentes.
- Distribución.
- Extensibilidad y fácil evolución.

2.9. Arquitectura de Bases de Datos Heterogéneas Distribuidas

2.9.1. Arquitectura de “wrappers y mediadores”

Un problema común que deben enfrentar las organizaciones en el día de hoy es la diversidad de fuentes de información y repositorios distribuidos con las que deben tratar. Alguno de estos ejemplos son: bases de datos, sistemas de archivos planos, documentos estructurados, etc.

El proceso de toma de decisiones necesita información de fuentes de datos múltiples, heterogéneas y distribuidas, por lo que la integración de fuentes de datos heterogéneos tiene una importancia considerable. La integración de múltiples fuentes requiere crear una vista integrada que permita realizar consultas en forma distribuida.

Las características para una integración de múltiples fuentes de datos pueden resumirse como:

- El número de fuentes de datos puede ser muy alto y hacer de la integración en una vista y la resolución de conflictos un gran desafío.
- Agregar y quitar fuentes de datos del sistema debería realizarse con mínimos impactos en la vista integrada.
- Las fuentes de datos varían en capacidades de computación. Por ej; archivos de texto plano, hojas de cálculo, sistemas de bases de datos.
- La naturaleza de las fuentes de datos estructuradas, semiestructuradas y nada estructuradas puede no aportar información para la vista integrada.

Como forma de atacar estas características se plantean varias arquitecturas, dentro de las cuales encontramos la arquitectura de wrappers y mediadores.

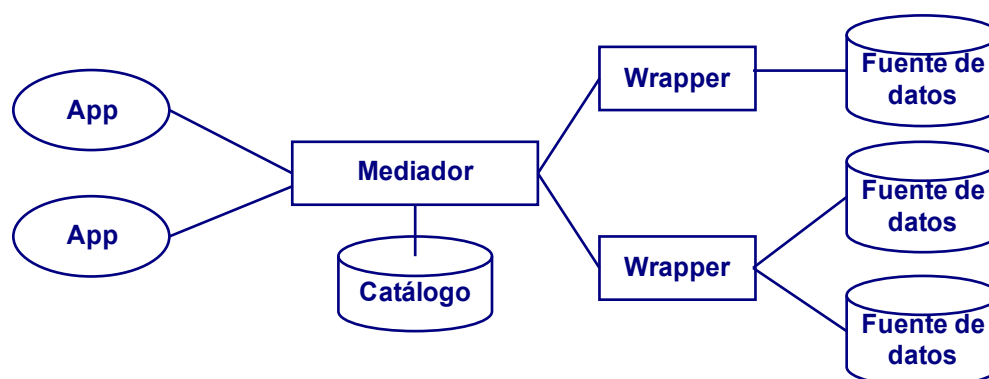


Figura 2: *Arquitectura de Wrappers y Mediadores.*

En la Figura 2 se muestra la arquitectura de wrappers y mediadores, donde los wrappers se utilizan para proveer acceso a las fuentes de datos heterogéneas. Para cada fuente de datos, un wrapper exporta información a cerca de su esquema fuente, sus datos y las capacidades de consultas. Un mediador almacena la

información aportada por los wrappers en una vista unificada de todos los datos disponibles en un diccionario o catálogo de datos central. Cuando una aplicación realiza una consulta, el mediador descompone la misma en consultas más pequeñas que se envían a los wrappers que corresponda para que estos las ejecuten en sus propias fuentes de datos, luego se obtienen los resultados de dichas consultas desde los wrappers implicados y se unen estos resultados de forma de satisfacer la consulta original de la aplicación.

2.9.2. Integración en un Esquema Global

La integración en un esquema global fue uno de los primeros intentos de compartir datos a través de arquitecturas de bases de datos heterogéneas distribuidas (HDDBSs). Está basado en la integración completa de múltiples bases de datos con la idea de proveer una vista simple (esquema global) de las fuentes integradas.

La ventaja de este enfoque es que los usuarios tienen una vista y acceso a los datos en forma uniforme y consistente. Los usuarios ignoran que las bases de datos que están utilizando están distribuidas y son heterogéneas. Múltiples bases de datos figuran en forma lógica como una base de datos simple. Sin embargo, la integración en un esquema global tiene ciertas desventajas [2]:

- Es difícil de automatizar por la complejidad en identificar vínculos y relacionamiento entre diferentes esquemas a integrar. Es necesaria intervención humana para resolver diferentes temas semánticos, estructurales o conflictos.
- Es común tener que sacrificar la autonomía de las fuentes para la resolución de conflictos semánticos. Todas las bases de datos involucradas necesitan revelar información a cerca de sus esquemas conceptuales o diccionarios de datos ya que el proceso de integración global requiere un conocimiento semántico completo. Incluso ocurre que una base de datos local necesita alterar su esquema para facilitar la integración.
- Si existen más de dos bases de datos, existen varios métodos de integración: algunos consideran todos los esquemas de una vez y otros consideran de a dos esquemas por vez. Dependiendo en el orden en que los esquemas son integrados, solo conocimiento semántico incompleto es utilizado en cada paso. Como resultado, puede perderse algún conocimiento semántico a menos que la integración se realice en un único paso y considere la exportación de todos los esquemas en forma simultánea.

Es claro que la integración en un esquema global consume mucho tiempo y es propenso a que se comentan errores. No es apropiado para esquemas que cambian con cierta frecuencia ya que todo el proceso de integración debería realizarse nuevamente.

2.9.3. Sistemas de Bases de Datos Federados

El propósito de la arquitectura de sistemas de bases de datos federados (FDBS) es eliminar la necesidad de construir un esquema global a la hora de integrar múltiples fuentes de datos. Esto permite a cada base de datos local tener más control sobre la información compartible. El control, por lo tanto, es descentralizado. La integración no tiene que ser completa, como en la integración en un esquema global, pero depende de la necesidad de los usuarios. Los FDBSs pueden ser fuerte o débilmente acoplados. Se los podría ubicar entre la no integración y la integración total. Una arquitectura típica de un FDBS debería considerar un modelo de datos y un lenguaje de implementación común.

En los FDBSs débilmente acoplados es responsabilidad de los usuarios mantener y crear el esquema federado. No se impone ningún control por el sistema federado o los administradores de la federación.

En los FDBSs fuertemente acoplados los administradores de la federación tienen total control sobre la creación y mantenimiento de los esquemas federados y acceso para exportar esquemas. La idea es proveer transparencia en la ubicación, replicación y distribución.

Capítulo 3 – DESARROLLOS SIMILARES Y HERRAMIENTAS UTILIZADAS

3.1. *Desarrollos similares*

En el mundo de hoy es claro ver la necesidad de integración de contenidos que genera la cantidad de información distribuida en diferentes repositorios y formatos. Esto es más notorio aún en el ambiente web, la dinámica que brinda el mismo permite una variedad inmensa de información estructurada, semiestructurada y nada estructurada. Desde empresas que buscan mejorar el resultados de sus negocios hasta organizaciones que desean crear fuentes de consulta completas o lo más amplias posible caen en la necesidad de integrar e interoperar.

El tema de integración esta fuertemente presente en diferentes áreas de Tecnología de la Información (TI). Debido al crecimiento de los volúmenes de información y al avance en las herramientas de hardware disponibles que aportan mayores velocidades y disponibilidad de recursos, junto a la necesidad de sacar provecho de la información que se posee, las organizaciones optan por realizar diferentes integraciones en múltiples niveles. Es común encontrar enfoques de integración que extraen información de diferentes fuentes, la integran de alguna manera creando una categorización propia que utilizarán al momento de la realización de búsquedas o como guía para acceder a los contenidos integrados.

A continuación se mencionan ejemplos encontrados y se da una breve descripción de los mismos. Para obtener mayor detalle sobre estos dirigirse al Anexo I – Desarrollos Similares.

Categorización según KMWorld

KMWorld es una autoridad reconocida en el campo del gerenciamiento de contenidos, documentos y conocimientos. Informa sobre los componentes, los procesos y las historias relacionadas al éxito que junto a las soluciones mejoran el funcionamiento de negocios. Ha realizado una selección y publicado en abril del 2004 una lista de las mejores empresas en gestión de contenidos y gestión del conocimiento [4].

Analizando los productos y herramientas presentadas por dichas empresas, se detectó que muchas de estas se ocupan de la integración de contenidos. El gran porcentaje de las mismas contempla contenidos en múltiples formatos. Se manejan: hojas de cálculo, documentos publicados en la web (formatos variados como HTML, XML y otros), documentos de texto, etc.

Presentan los resultados de diferentes formas, algunas brindan una interfaz grafica configurable por los propios usuarios finales de la herramienta, ofreciendo así facilidades para visualizar los resultados.

Se puede apreciar una orientación de estas herramientas hacia la parte comercial. Por lo analizado, la mayoría de éstas brindan utilidades a empresas con el fin de integrar aquella información que pueda ser útil a la hora de realizar estudios de posicionamiento de mercado o en la toma de decisiones.

Se ofrecen herramientas con diferentes enfoques, como ser el de mantener información centralizada sobre habilidades y conocimientos de empleados en empresas de desarrollo de software, otras que se basan en la recolección de información financiera que permita realizar análisis profundos con el fin de conocer el estado de la organización y de sus competidores, poder realizar proyecciones y tomar decisiones y otras múltiples orientaciones.

A continuación se listan algunos ejemplos analizados de esta extensa lista:

Anacubis

Ofrece la posibilidad de utilizar múltiples fuentes (páginas web, hojas electrónicas, bases de datos, etc.) para lograr integrar información de negocios de forma de realizar conciliaciones y análisis [5].

Arbortext

Utiliza XML (convierte todo a este formato) para integrar contenidos de diferentes tipos de fuentes. Permite tener centralizada gran cantidad de información y poder desplegar resultados de búsquedas en diferentes formatos [6].

TripleHop Technologies, Inc

Aporta una herramienta que permite realizar búsquedas sobre múltiples repositorios. Lo innovador que presentan es el tener en cuenta el contexto de la información utilizando análisis semánticos y estadísticos para dichas tareas [7].

Google

Google es una compañía que tiene como objetivo ofrecer la mejor experiencia de búsqueda en Internet ofreciendo acceso a información pertinente de cualquier parte del mundo. Google, como creador del motor de búsqueda más grande del mundo, ofrece la forma más rápida y sencilla de encontrar información en la web. Con acceso a más de 8.000 millones de páginas web, Google proporciona resultados relevantes a usuarios de todo el mundo normalmente en menos de medio segundo [8].

Conjuntamente con su sistema de búsquedas ofrece otros servicios dentro de los cuales encontramos Google News, un servicio automatizado que clasifica y publica noticias actualizadas constantemente a partir de 700 fuentes de información en

español, brindando así la posibilidad de buscar y explorar distintas perspectivas sobre una misma historia [9].

En Google News las noticias son extraídas automáticamente de miles de fuentes, clasificadas en categorías y relacionadas entre sí. Además se muestra una imagen (fotografía) asociada a cada noticia destacada. Las noticias provenientes de distintas fuentes que refieren al mismo tema aparecen agrupadas bajo el titular del medio más destacado. En ocasiones el número de fuentes que tratan el mismo tema puede ser de varios cientos. La información se renueva cada pocos minutos, generándose las portadas de las secciones sin intervención humana alguna.

Como característica interesante de Google News se puede resaltar la creación de trazas por fechas de todas las noticias que se han publicado sobre un mismo tema.

A la fecha 29 de mayo de 2004 Google News era todavía un producto en fase beta. A la fecha 13 de noviembre de 2004 ya ha dejado la fase beta y se promociona junto al buscador comúnmente utilizado. Al momento de realizar una búsqueda en el buscador de Google, si en Google News se obtiene algún resultado para la misma se encabeza la lista de resultados generales con el resultado de Google News.

Otro de los servicios que brinda Google es "El directorio web de Google" que organiza la Web por temas, de modo que se pueda acotar el espacio de búsqueda. Este servicio está disponible en diferentes idiomas [10].

3.2. Contraste del IC frente a estos desarrollos

En los ejemplos analizados se puede apreciar dos enfoques bastante diferentes. El primero de ellos se ocupa de la integración de noticias con el fin de ofrecer servicios de búsquedas y categorizaciones, mientras que el segundo enfoque se orienta hacia el lado de negocios, brindando herramientas que ayudan a realizar estudios de mercado y a la toma de decisiones según datos recopilados de múltiples formas.

La solución Integradora de contenidos (IC), producto del desarrollo de este proyecto, se pudo incluir en el primer grupo antes mencionado. A continuación se exponen las características principales del mismo que lo diferencian del resto de los desarrollos analizados.

El IC enfoca al tratamiento de artículos previamente categorizados por cada una de las fuentes que brinda dicha información y a su vez posee una categorización propia más genérica que le permite mantener un mapeo entre las categorías originales y las propias. Este mapeo se coordina entre administradores del IC y de las fuentes correspondientes. Esto permite utilizar el conocimiento de quienes desarrollan las noticias o artículos a la hora de categorizarlos en el IC y se evita así el tener que tratar con meta información o información que queda implícita en dichos contenidos.

Otra de las diferencias es que las noticias consideradas para la integración provienen de empresas dedicadas al periodismo que desean cooperar entre sí. Con esto se quiere decir que no se extrae la información desde internet simplemente sino que se sabe con certeza de donde proviene la misma y hay por detrás un vínculo de confianza hacia la fuente generadora de dicha información.

Algo que corresponde destacar es que en el IC hay gran importancia en la dimensión del tiempo, ya que se trata con noticias que llevan asociada una fecha de publicación. Esto brinda la posibilidad de diferenciar servicios aportados por IC en dos formas diferentes; la primera es la posibilidad de realizar una búsqueda sobre información actual (o sea publicada en el día de hoy) y la segunda es la de realizar búsquedas en todo el repositorio (que en este caso se asemeja a los buscadores Web conocidos).

Otra de los puntos que diferencian al IC es que este sistema realiza una integración dirigida no a usuarios finales sino a consumidores. Las diferentes fuentes registradas en el IC aportan información a ser integrada que será puesta a disposición de consumidores. A su vez, estos deberán realizar una tarea de distribución de acuerdo a los usuarios que poseen y sus preferencias.

3.3. Herramientas utilizadas – Tecnología .NET.

Microsoft .NET Framework es un nuevo componente de la familia de sistemas operativos, Microsoft Windows. Es el fundamento de la siguiente generación de aplicaciones basadas en Windows que son fáciles de construir, emplear e integrar con otros sistemas en red [11][12][13][14][15].

.NET Framework simplifica el desarrollo de software Windows. Proporciona a los desarrolladores un único enfoque para construir tanto las aplicaciones de escritorio como aplicaciones basadas en web. También permite a los desarrolladores utilizar las mismas herramientas y habilidades para desarrollar software para una variedad de sistemas que van desde teléfonos manuales avanzados hasta voluminosas instalaciones de servidores.

El software que se construye en .NET Framework puede ser más fácil de emplear y mantener que el software convencional. Se pueden diseñar aplicaciones para que se actualicen automáticamente a la versión más moderna. .NET Framework puede también minimizar los conflictos entre aplicaciones al ayudar a coexistir a componentes de software incompatibles.

.NET Framework busca desarrollar un entorno específicamente diseñado para el desarrollo y ejecución del software en forma de servicios que puedan ser tanto publicados como accedidos a través de Internet de forma independiente del lenguaje de programación, modelo de objetos, sistema operativo y hardware utilizados tanto para desarrollarlos como para publicarlos. Éste entorno es lo que

se denomina la plataforma.NET, y los servicios antes mencionados son los que se denomina servicios web.

El corazón de la plataforma.NET es el CLR (Common Language Runtime), que es una aplicación similar a una máquina virtual que se encarga de gestionar la ejecución de las aplicaciones para ella escritas. A estas aplicaciones les ofrece numerosos servicios que facilitan su desarrollo y mantenimiento y favorecen su fiabilidad y seguridad. Entre ellos los principales son:

- Modelo de programación consistente y sencillo, completamente orientado a objetos.
- Eliminación del temido problema de compatibilidad entre DLLs conocido como "infierno de las DLLs".
- Ejecución multiplataforma.
- Ejecución multilenguaje, hasta el punto de que es posible hacer cosas como capturar en un programa escrito en C# una excepción escrita en Visual Basic.NET que a su vez hereda de un tipo de excepción escrita en Cobol.NET.
- Recolección de basura.
- Aislamiento de memoria entre procesos y comprobaciones automáticas de seguridad de tipos en las conversiones.
- Soporte multihilo.
- Gestión del acceso a objetos remotos que permite el desarrollo de aplicaciones distribuidas de manera transparente a la ubicación real de cada uno de los objetos utilizados en las mismas.
- Seguridad avanzada, hasta el punto de que es posible limitar los permisos de ejecución del código en función de su procedencia (Internet, red local, CD-ROM, etc.), el usuario que lo ejecuta o la empresa que lo creó.
- Interoperabilidad con código preexistente, de manera que es posible utilizar con facilidad cualquier librería de funciones u objetos COM y COM+ creados con anterioridad a la aparición de la plataforma .NET.
- Adecuación automática de la eficiencia de las aplicaciones a las características concretas de cada máquina donde se vaya a ejecutar.

La plataforma .NET de Microsoft permite la generación, implementación, operación e integración de XML web services. Permite la implementación del sistema como la composición de servicios más simples dando como resultando nuevos servicios.

Los XML web services son entidades que proporcionan un elemento de funcionalidad determinado al que se puede tener acceso desde diversos sistemas potencialmente distintos mediante estándares de Internet ampliamente utilizados hoy en día como son XML y HTTP.

Los XML web services dependen en gran medida de la amplia aceptación de XML y otros estándares de Internet para crear una infraestructura que posibilita el funcionamiento de un conjunto de aplicaciones distribuidas.

Un XML web service puede ser utilizado internamente por una aplicación o bien ser expuesto de forma externa en Internet por varias aplicaciones. Dado que a través de una interfaz estándar es posible el acceso a un XML web service, éste permite el funcionamiento de una serie de sistemas distribuidos y heterogéneos como un conjunto integrado.

Los XML web service utilizan mensajería basada en XML como medio fundamental de comunicación de datos para contribuir a reducir las diferencias existentes entre entornos que utilizan distintos modelos de componentes, sistemas operativos y lenguajes de programación.

Para mas detalles sobre el funcionamiento y operación de las interfaces de .NET ver Anexo II – Tecnologías.

Capítulo 4 - DEFINICIONES Y ANÁLISIS DEL PROYECTO

4.1. Definición de alcance

Antes de comenzar el análisis de requerimientos se define el alcance del proyecto en función de los recursos disponibles y duración del mismo. A continuación se listan las principales tareas a realizar con el fin de lograr los objetivos planteados:

- Creación de una categorización genérica que permita encontrar una correspondencia entre esta y categorías pertenecientes a fuentes de información.
- Registro y administración de Consumidores y Fuentes participantes en el sistema IC.
- Consideración de idiomas en el diseño para integrar fuentes de diferentes países.
- Especificación de interfaces de comunicación que deberá cumplir una fuente en caso de solicitar ser integrada.
- Desarrollo necesario para la implementación de una fuente para el periódico ELPD.
- Desarrollo necesario para la implementación de un consumidor para el periódico ELPD. Para esto se plantea:
 - Manejo de perfiles de búsquedas de usuarios registrados.
 - Despliegue de resultados según diseño especificado por ELPD.
- Desarrollo necesario para la implementación de una fuente para el periódico La Nación de Argentina a partir de páginas RSS y HTML publicadas en su sitio web.

4.2. Análisis de requerimientos

Una vez que se ha definido el alcance se procede a realizar el análisis de requerimientos. Se documentan los requerimientos funcionales identificando dos grupos; los requerimientos del sistema IC y los del sistema a implementar para el consumidor de ELPD y luego los no funcionales.

4.2.1. Requerimientos funcionales

4.2.1.1. Sistema IC

- Crear una categorización estándar para el contenido a integrar independiente de la categorización utilizada por las fuentes registradas en el IC.
- El sistema a desarrollar debe permitir integrar contenidos de diferentes fuentes (periódicos) en formato digital.
Para esto se deberá:

- Desarrollar sistema de registro de fuentes de noticias desde las que se extraerá la información.
 - Permitir la inclusión y exclusión de fuentes al sistema de manera sencilla.
 - Disponer de métodos de control para la validación de una nueva fuente por parte de un administrador del IC.
 - Restringir el acceso de nuevas fuentes hasta que no tengan realizada la correspondencia entre sus propias categorías y las del IC.
 - Permitir la distribución de la categorización creada previamente para la clasificación de noticias. Esta categoría debe permitir evoluciones de versiones previas.
 - Desarrollar un mecanismo para centralizar los contenidos provenientes de las diferentes fuentes en forma periódica mediante la utilización de un catálogo.
 - Tener un mecanismo para el ingreso al sistema de IC con interfaces estándares.
 - Diseñar el sistema contemplando la posibilidad de manejar diferentes idiomas. Se considera suficiente la implementación de todas las funcionalidades en español, quedando la posibilidad de inclusión de otros idiomas a futuro.
- Ofrecer mecanismos para acceder a contenidos integrados a los consumidores del sistema. Se entiende por consumidor a un subsistema externo que administra un conjunto de usuarios y sus preferencias. Para esto se deberá:
 - Desarrollar un sistema de registro para nuevos consumidores.
 - Validar mapeos de correspondencias entre las preferencias de los consumidores y las categorías existentes en el IC.
 - Disponer de métodos de control para la habilitación de los consumidores que se van registrando.
 - Disponer de un sistema de autenticación de los consumidores de servicios aportado por el IC.
 - Tener un mecanismo para el ingreso al sistema de IC con interfaces estándares.
 - Desarrollar un sistema de notificación a los consumidores sobre cambios de versiones de categorización del IC.

4.2.1.2. Sistema Consumidor para ELPD

- Se deberá desarrollar el sistema consumidor para ELPD. Este consumirá servicios aportados por el IC. Para esto se deberá:
 - Utilizar el sistema de registro usuarios ya desarrollado por ELPD para el manejo de sus preferencias.
 - Disponer de un sistema de autenticación de los usuarios.

- Desarrollar métodos que permitan la recuperación de contenidos integrados según preferencias de los usuarios.
- La información que se presentará al usuario será título y avance de cada noticia según el perfil del usuario y un vínculo al desarrollo completo de la noticia en sitio Web de la fuente correspondiente.

4.2.2. Requerimientos no funcionales

- Desarrollar el sistema para que funcione con la tecnología y recursos disponibles en ELPD. Se deberá usar como sistema operativo Windows 2000, y motor de base de datos Microsoft SQL Server 2000.
- En principio no hay requerimientos de tiempo de respuesta, pero se considera que los tiempos de consultas no deberían superar los 50 segundos, que es un tiempo máximo estimado para que un usuario espere por la carga de una página (tiempo de latencia).
- El IC trabajará con el encabezado (título y avance) de cada noticia pero se considera la posibilidad de utilizar el contenido completo de las mismas para manejo interno del sistema.
- Se trabajara con el periódico La Nación de Argentina y para esto se deberá extraer la información en el formato presentado por ellos (RSS y HTML) disponiendo aquí del título y avance de cada noticia agrupado por categorías propias de dicho periódico.

4.3. Casos de Uso

A continuación se documentan los casos de uso (CU) más relevantes para la aplicación que involucran a los Sistemas de IC y Fuente.

4.3.1. Diagrama de Casos de Uso IC

Los CU son usados para definir la arquitectura de la aplicación y muestran la interacción de los diferentes actores del sistema. Los mismos se desprenden de los requerimientos documentados previamente y validados por ELPD, cada uno satisface al menos una funcionalidad que el sistema debe brindar.

A continuación se muestra un diagrama conteniendo los CU que se detallarán a continuación. Para obtener mayor detalle sobre estos dirigirse al Anexo IV – Casos de uso.

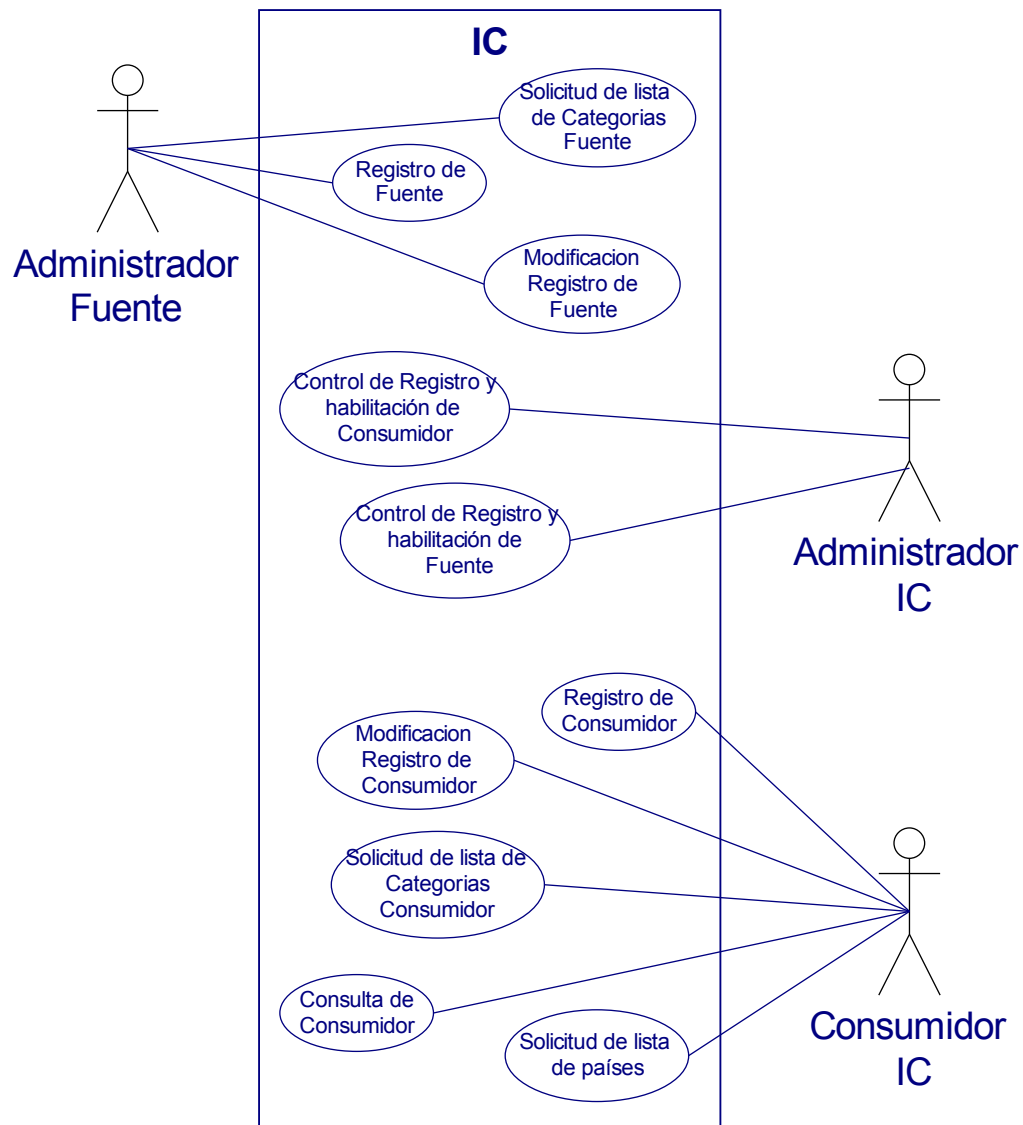


Figura 3: *Casos de uso IC.*

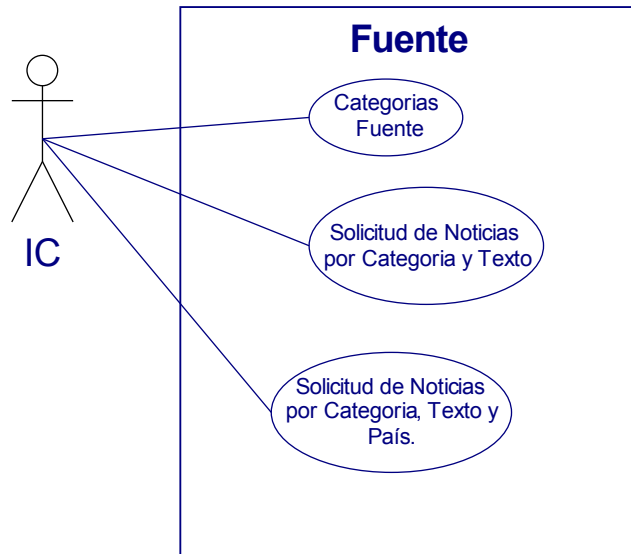


Figura 4: Casos de uso fuente.

4.3.2. Detalle de Casos de Uso

Casos de Uso:

Nombre: Solicitud de lista de categorías Fuente.

Identificador: CU1.

Actor: Fuente.

Sistema: IC.

Descripción: La fuente obtiene la categorización de noticias utilizada por el IC.

Precondiciones: La Fuente debe estar registrada en el IC.

Poscondiciones: Ninguna.

1-La Fuente se autentica y pide la categorización utilizada por el IC para la clasificación de noticias.	
	2- Si la autenticación se efectúa, el IC envía las categorías de noticias.
	Fin.

Nombre: Registro de Fuente.

Identificador: CU2.

Actor: Fuente.

Sistema: IC.

Descripción: La fuente hace un pedido de registro en el sistema IC. El mismo se hace en tres pasos fundamentales de los cuales se exige como mínimo completar correctamente el primero, debiendo ingresar posteriormente el resto de la información para poder ser habilitada. Primero la fuente coloca datos de identificación, autenticación y de contacto. Segundo el sistema IC pide la ubicación del servicio web requerido y una clave para utilizarse en futuras invocaciones al mismo como forma de autenticación. Tercero el sistema pide a la fuente que defina un mapeo entre las categorías propias y las del IC. Completando correctamente por lo menos la primera etapa, la fuente queda registrada en espera de un posterior control para su habilitación.

En todos los casos se manda un mail al administrador de la fuente (según datos proporcionados en registro) con estado del registro.

Precondiciones: Ninguna.

Poscondiciones: Ninguna.

1-El administrador de la Fuente pide al IC formulario para registro.	
	2- El IC envía el primero de los formularios de registro de fuente.
3- El administrador de la Fuente ingresa datos de identificación, autenticación y de contacto y envía.	
	4- Si se ingresan correctamente los datos, el IC envía segundo formulario de registro.
5- El administrador de la Fuente coloca dirección URL de servicio web que atiende pedidos del IC y clave a ser usada por el IC en futuras invocaciones.	
	6-El IC usando la dirección URL y clave del servicio web publicado por la fuente, realiza una validación de los métodos que se requiere que posea el XML web service. Usando uno de estos pide las categorías con que la fuente clasifica sus notas, y genera un formulario que permite establecer los mapeos a las categorías del IC.
9- El administrador de la Fuente establece el mapeo entre las categorías de la Fuente y las del IC y envía.	
	10- El IC almacena la información en su base de datos, informa el fin de la operación de registro y envía un mail al administrador de la fuente con el estado del registro.
	Fin.

Nombre: Control de registro y habilitación de fuente.

Identificador: CU3.

Actor: Administrador de IC.

Sistema: IC.

Descripción: El administrador IC pide al IC la lista de fuentes en proceso de ser aceptadas en el sistema. El IC entrega la lista y permite seleccionar la fuente a controlar. El IC despliega opciones para: 1- visualizar datos ingresados por la fuente en el momento de registro, 2- visualizar definición de mapeo entre categorías de la fuente y categorías del IC, 3- validar métodos del XML web service requerido, 4- realizar notificaciones a la fuente vía e-mail y 5- habilitar a una fuente para que el IC la considere en futuras integraciones.

El administrador realiza los chequeos que considera. Revisa la información de identificación y los mapeos entre categorías de la fuente y el IC. Se brinda la posibilidad de invocar a los métodos del XML web service dispuesto por la fuente y se despliegan resultados en formato web. El administrador decide habilitar o no a la fuente en el integrador. En caso de habilitación el sistema avisa por mail al administrador de la fuente el resultado de las operaciones. También queda disponible un sistema para realizar notificaciones vía e-mail a la fuente de parte del administrador del IC.

Precondiciones: Ninguna.

Poscondiciones: Ninguna.

1- El Administrador pide lista de fuentes registradas que todavía no están habilitadas.	
	2- El IC entrega lista de fuentes.
3- El Administrador selecciona la fuente que va a controlar.	
	4- El IC despliega opciones para: 1- visualizar datos ingresados por la fuente en el momento de registro, 2- visualizar definición de mapeo entre categorías de la fuente y categorías del IC, 3- validar métodos del XML web service requerido, 4- realizar notificaciones a la fuente vía e-mail y 5- habilitar a una fuente para que el IC la considere en futuras integraciones.
5- Si administrador elige 4.1:	
	6- IC despliega la información ingresada por la fuente en el momento de registro.
7- Si la información es incorrecta el administrador puede modificar lo que considere y/o realizar una notificación a la fuente del error vía e-mail mediante los métodos aportados por el IC (4.4). Pasa a: 8, 11, 14 o Fin.	
8- Si administrador elige 4.2	
	9- IC despliega la definición de mapeo entre las categorías de la fuente y las

	del IC realizada por la fuente.
10- Si la información es incoherente el administrador puede modificar lo que considere y/o realizar una notificación a la fuente del error vía e-mail mediante los métodos aportados por el IC (4.4). Pasa a: 5, 11, 14 o Fin.	
11- Si administrador elige 4.3	
	12- IC despliega opciones para invocar a los métodos del XML web service que se exige a la fuente.
13- Administrador invoca los métodos que considere. Si los resultados son incorrectos puede realizar una notificación a la fuente del error vía e-mail mediante los métodos aportados por el IC (4.4). Pasa a: 5, 8, 14 o Fin.	
14- Una vez realizados los controles necesarios el administrador puede activar la fuente (4.5) para que el IC la considere en futuras integraciones.	
	15- Si se habilita la fuente, se envía e-mail a la misma notificando de tal habilitación.
	Fin.

Nombre: Modificación de registro de Fuente.

Identificador: CU4.

Actor: Fuente.

Sistema: IC.

Descripción: La fuente hace un pedido de modificación de registro en el sistema IC. El IC solicita que la misma se autentique para poder acceder a dicha información. Se disponen de tres pasos de modificación, uno para la modificación de datos de la fuente (identificación, autenticación y de contacto), el segundo para la modificación de la URL y clave del XML web service requerido por el IC y el tercero para la modificación de la correspondencia entre categorías de la fuente y categorías del IC.

Cuando se pasa al tercer paso, se invoca al método del XML web service que obtiene las categorías de la fuente y se realiza un chequeo contra el mapeo que esta almacenado en la base de datos del IC. En caso de que exista una nueva categoría por parte de la fuente se notifica y se brinda la posibilidad de asignar correspondencia con una categoría del IC. En caso de que una categoría de la fuente almacenada en la base de datos del IC ya no sea listada por la fuente, se muestra dicha ocurrencia y se notifica que de actualizar el mapeo la misma será dada de baja del catálogo del IC. También se muestra el mapeo existente entre aquellas categorías devueltas por el XML web service y las categorías del IC con la

posibilidad de que el administrador de la fuente modifique dichas correspondencias.

Precondiciones: La Fuente debe estar registrada en el IC.

Poscondiciones: Ninguna.

1-El administrador de la Fuente pide al IC modificar su registro.	
	2- El IC solicita autenticación.
3- El administrador de la Fuente se autentica ingresando usuario y contraseña.	
	4- Si la autenticación es correcta, el IC envía el primero de los formularios de registro donde se pueden modificar datos de la fuente como ser: identificación, autenticación y datos de contacto. El IC despliega opciones para: 1-efectuar actualización de datos, 2-continuar para modificar la dirección URL y clave del XML web service.
5- Si administrador elige 4.1	
	6- IC valida información. Si es correcta efectúa modificación. Si hay alguna información incorrecta notifica al administrador del error y brinda opción de corregir y volver a intentar la modificación.
7- Administrador pasa a 8 o Fin	
8- Si administrador elige 4.2	
	9- El IC despliega opciones para: 1-modificar la dirección URL y clave del XML web service, 2- continuar para modificar la definición de mapeo de categorías entre fuente e IC
10- Si administrador elige 9.1	
	11- IC valida información. Si es correcta efectúa modificación. Si hay alguna información incorrecta notifica al administrador del error y brinda opción de corregir y volver a intentar la modificación.
12- Administrador pasa a 13 o Fin	
13- Si administrador elige 9.2	
	14- IC invoca al método del XML web service (utilizando la clave aportada por la fuente) que obtiene las categorías de la fuente y se realiza un chequeo contra el mapeo que esta almacenado en la base de datos del IC. En caso de que exista una nueva categoría por parte de

	la fuente se notifica y se brinda la posibilidad de asignar correspondencia con una categoría del IC. En caso de que una categoría de la fuente almacenada en la base de datos del IC ya no sea listada por la fuente, se muestra dicha ocurrencia y se notifica que de actualizar el mapeo la misma será dada de baja del catálogo del IC. También se muestra el mapeo existente entre aquellas categorías devueltas por el XML web service y las categorías del IC con la posibilidad de que el administrador de la fuente modifique dichas correspondencias.
15- Si hay nuevas categorías por parte de la fuente, el administrador asigna las correspondencias entre sus nuevas categorías y las categorías del IC. Si considera modifica un mapeo existente entre una categoría de la fuente y una del IC. Indica Actualizar.	
	16- Si hay nuevas categorías listadas por la fuente, da de alta las correspondencias con categorías del IC. Si hay categorías en la base de datos que la fuente ya no lista, da de baja las mismas. Actualiza correspondencias modificadas por el administrador de la fuente para aquellas categorías ya existentes.
	Fin.

Nombre: Solicitud de lista de categorías consumidor.

Identificador: CU5.

Actor: Consumidor.

Sistema: IC.

Descripción: El consumidor obtiene la categorización de noticias utilizada por el IC. Esta información será utilizada por el consumidor en el momento de realizar una consulta al IC.

Precondiciones: El Consumidor debe estar registrado en el IC.

Poscondiciones: Ninguna.

1-El Consumidor se autentica y pide la categorización utilizada por el IC para la clasificación de noticias.	
	2- Si la autenticación se efectúa, el IC envía las categorías de noticias.
	Fin.

Nombre: Solicitud de Lista de Países.

Identificador: CU6.

Actor: Consumidor.

Sistema: IC.

Descripción: El Consumidor obtiene la lista de países utilizada por el IC. Esta información será utilizada por el consumidor en el momento de realizar una consulta al IC.

Precondiciones: El Consumidor debe estar registrado en el IC.

Poscondiciones: Ninguna.

1-El Consumidor se autentica y pide la lista de países utilizada por el IC.	
	2- Si la autenticación se efectúa, el IC envía la lista de países utilizada.
	Fin.

Nombre: Registro de Consumidor

Identificador: CU7

Actor: Consumidor

Sistema: IC

Descripción: El consumidor hace un pedido de registro en el sistema IC. Este registro se hace en dos pasos de los cuales se exige como mínimo completar correctamente el primero, debiendo ingresar posteriormente el resto de la información para poder ser habilitado. Primero el consumidor coloca datos de identificación, autenticación y de contacto. Segundo el sistema IC despliega las posibles fuentes de las que puede consumir. El consumidor realiza su elección y queda registrado en espera de un posterior control para su habilitación.

En todos los casos se manda un mail al administrador del consumidor (según datos proporcionados en registro) con estado del registro.

Precondiciones: Ninguna.

Poscondiciones: Ninguna.

1-El administrador del Consumidor pide el IC formulario para registro.	
	2- El IC envía el primero de los formularios de registro de consumidor.
3- El administrador del Consumidor ingresa datos de identificación, autenticación y de contacto y envía.	
	4- Si se ingresan correctamente los datos, el IC lista las fuentes de las que puede consumir.

5- El administrador del Consumidor selecciona las fuentes de las que quiere consumir.	
	10- El IC almacena la información en su base de datos, informa el fin de la operación de registro y envía un mail al administrador del consumidor con el estado del registro.
	Fin.

Nombre: Modificación de registro de consumidor.

Identificador: CU8.

Actor: Consumidor.

Sistema: IC.

Descripción: El consumidor hace un pedido de modificación de registro en el sistema IC. El IC solicita que el mismo se autentique para poder acceder a dicha información. Se disponen de dos pasos de modificación, uno para la modificación de datos del consumidor (identificación, autenticación, de contacto) y el segundo para la modificación la elección de fuentes de las que desea consumir.

Precondiciones: El Consumidor debe estar registrado en el IC.

Poscondiciones: Ninguna.

1-El administrador del Consumidor pide al IC modificar su registro.	
	2- El IC solicita autenticación.
3- El administrador del Consumidor se autentica ingresando usuario y contraseña.	
	4- Si la autenticación es correcta, el IC envía el primero de los formularios de registro donde se pueden modificar datos del Consumidor como ser: identificación, autenticación y datos de contacto. El IC despliega opciones para: 1- efectuar actualización de datos, 2- continuar a la definición de fuentes de las que quiere consumir.
5- Si administrador elige 4.1	
	6- IC valida información. Si es correcta efectúa modificación. Si hay alguna información incorrecta notifica al administrador del error y brinda opción de corregir y volver a intentar la modificación.
7- Administrador pasa a 8 o Fin	
8- Si administrador elige 4.2	
	9- Se presentan las fuentes de las

	cuales el consumidor puede consumir y se marcan seleccionadas aquellas de las que el consumidor ya esta consumiendo.
10- El Administrador del Consumidor selecciona o deselecciona las fuentes de las que quiere o no quiere consumir respectivamente.	
	11- El IC actualiza selección del Consumidor.
	Fin.

Nombre: Consultas de consumidor.

Identificador: CU9.

Actor: Consumidor.

Sistema: IC.

Descripción: El Consumidor hace las consultas necesarias al IC para satisfacer la demanda de sus usuarios. La consulta se compone por cuatro partes: un texto a buscar, una categoría perteneciente al IC sobre las cual realizar la búsqueda, un país de donde se desean noticias y el idioma de la consulta. El IC devuelve resultados en un formato preestablecido.

Precondiciones: El Consumidor debe estar registrado en el IC.

Poscondiciones: Ninguna.

1- El Consumidor solicita búsqueda.	
	2- El IC solicita autenticación.
3- Se realiza la autenticación ingresando usuario y contraseña. Se pasa texto a buscar, categoría donde buscar, país desde donde quiere las noticias e idioma de la consulta.	
	4- Si el IC autentica al consumidor realiza búsquedas en su catálogo según parámetros recibidos. En caso de se necesario se invocan búsquedas a las fuentes de las cuales consume el consumidor. Devuelve los resultados de la búsqueda en un formato preestablecido.
	Fin.

4.3.3. Diagrama de Casos de Uso Consumidor IC para ELPD

A continuación se muestra un diagrama conteniendo los CU para implementar un consumidor para ELPD. Se detallan los mismos a continuación. Para obtener mayor detalle sobre estos dirigirse al Anexo IV – Casos de uso.

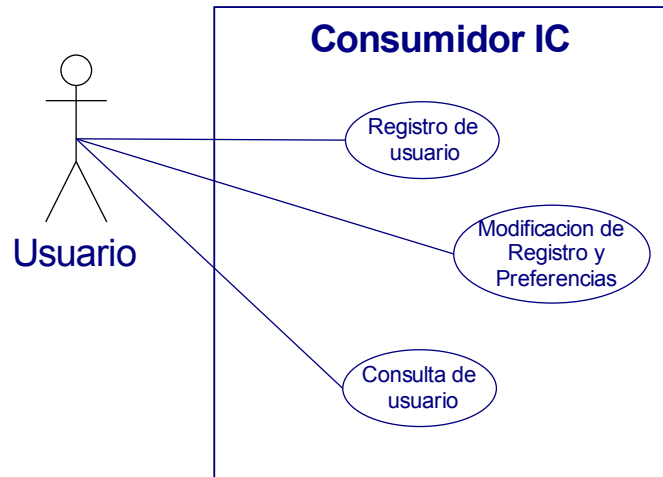


Figura 5: Casos de uso consumidor IC para ELPD.

4.3.4. Detalle de Casos de Uso

Nombre: Registro de Usuario.

Identificador: CU10.

Actor: Usuario.

Sistema: Consumidor ELPD.

Descripción: ELPD ya dispone de un sistema de registro de usuarios por lo que no fue necesario implementar este caso de uso.

Nombre: Modificar Registro y preferencias de Usuario.

Identificador: CU11.

Actor: Usuario.

Sistema: Consumidor ELPD.

Descripción: ELPD ya dispone de un sistema de modificación de registro de usuarios y su perfil de interés (preferencias) por lo que no fue necesario implementar este caso de uso.

Nombre: Consulta de Usuario.

Identificador: CU12.

Actor: Usuario.

Sistema: Consumidor ELPD.

Descripción: El Usuario pide consultar las noticias según sus preferencias. El Consumidor invoca una consulta sobre el IC (según CU9). Se despliegan resultados obtenidos en formato web.

Precondiciones: El usuario debe estar registrado en ELPD.

Poscondiciones: Ninguna.

1-El usuario ingresa al sitio de consumo de ELPD.	
	2- ELPD solicita autenticación.
3- El usuario se autentica y solicita una búsqueda según su perfil.	
	4- Si es autenticado el usuario, se consultan preferencias del mismo en base de datos local. Se invoca consulta al IC según la búsqueda a realizar. Se despliegan noticias al usuario
	Fin.

4.3.5. Flujo de información.

A continuación se presenta un diagrama de la arquitectura planteada para el sistema IC. Aquí se muestra gráficamente como fluye la información en el sistema. Se puede ver que las fuentes aportan contenidos al integrador quien luego de realizar su tarea pone a disposición la información integrada para aquellos consumidores habilitados.

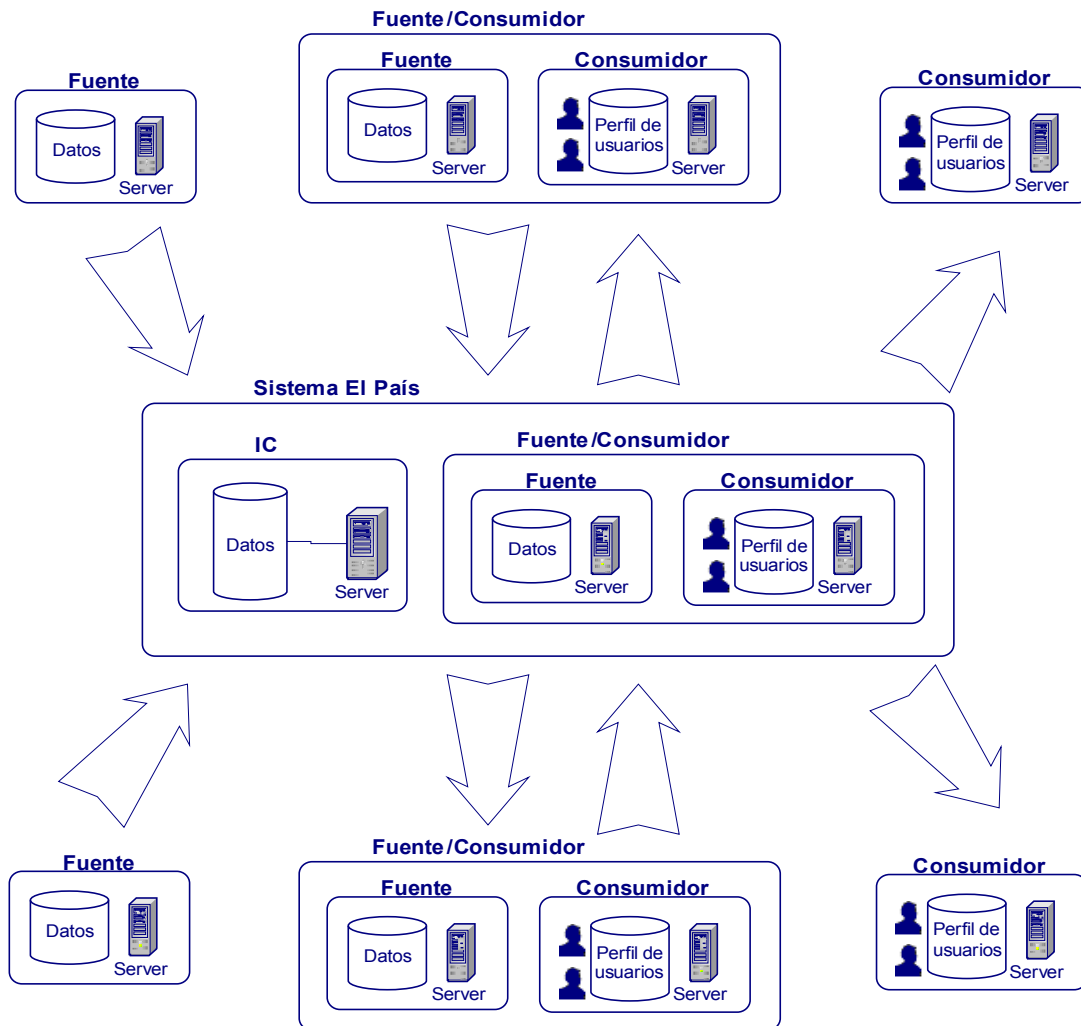


Figura 6: *Flujo de información del sistema IC.*

4.3.6. Diagramas de Secuencia

A continuación se presentan diagramas de interacción que describen cómo grupos de actores colaboran para conseguir algún fin en el sistema.

Estos diagramas muestran actores, así como los mensajes que se pasan entre ellos capturando así el comportamiento de los casos de uso presentados anteriormente. Se muestra las interacciones expresadas en función de secuencias temporales.

Por medio de los diagramas de las Figuras 7, 8 y 9 se presenta en forma general la interacción de los diferentes actores según casos de uso principales. Aquí se describen las acciones mínimas necesarias para llegar a un estado del sistema en el cual la información pueda fluir tal como se muestra en la Figura 6 de la sección anterior.

4.3.6.1. Secuencia de registro de fuente

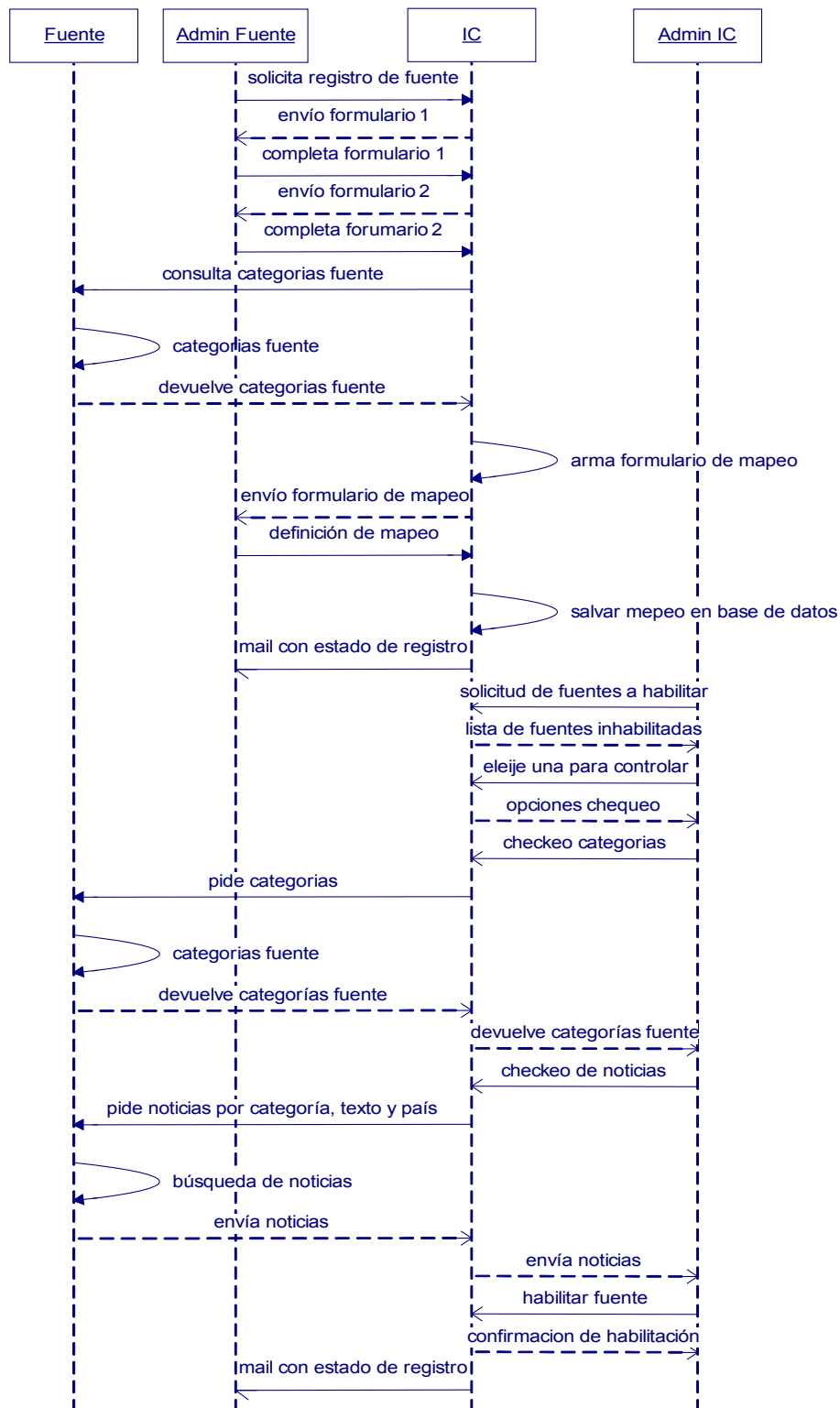


Figura 7: Diagrama de secuencia fuente-IC.

4.3.6.2. Secuencia de registro de consumidor

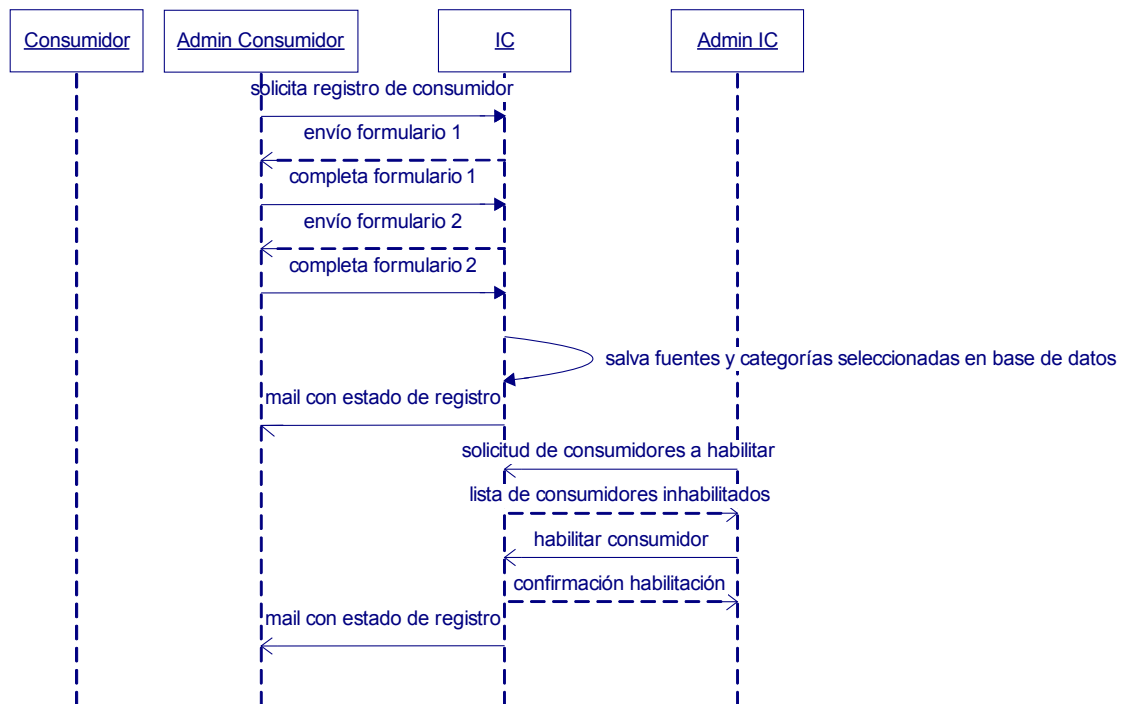


Figura 8: Diagrama de secuencia consumidor-IC.

4.3.6.3. Secuencia de consulta de usuario

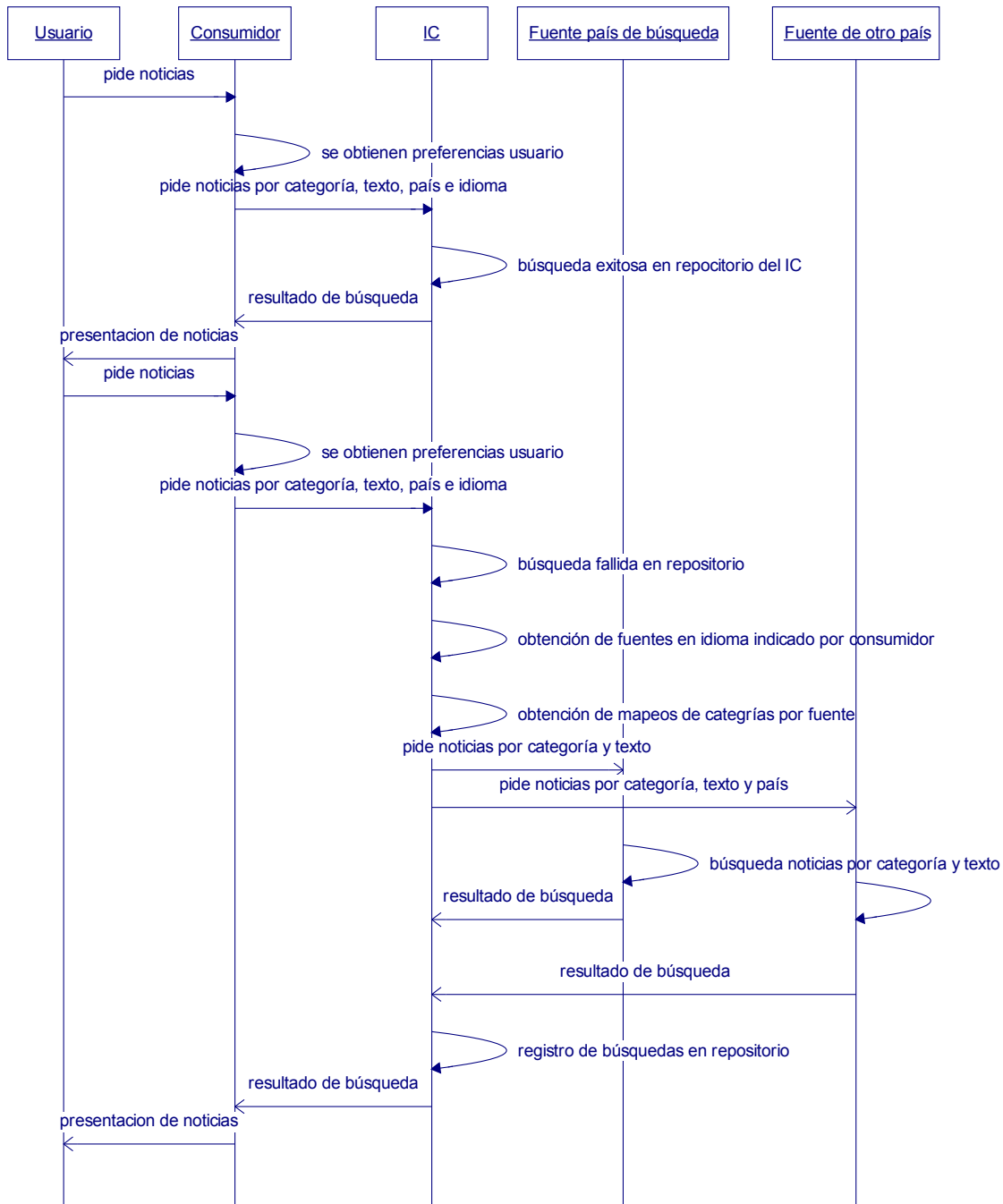


Figura 9: Diagrama de secuencia consulta de usuario-consumidor-IC-fuente.

Capítulo 5 – DISEÑO Y CONSTRUCCIÓN

Debido a las características del proyecto, la parte principal del mismo consiste en el desarrollo de aplicaciones que satisfaga los requerimientos especificados.

En este capítulo se tratan las decisiones tecnológicas, de diseño y de implementación tomadas para la construcción de dichas aplicaciones.

Se busca que la solución entregada cubra las necesidades planteadas por ELPD y posea las siguientes propiedades:

- confiabilidad,
- seguridad,
- disponibilidad,
- escalabilidad,
- facilidad de administración y
- rendimiento.

Para cubrir las características anteriores como primera opción se elige una arquitectura basada en dos ambientes, el web y el orientado a servicios.

5.1. Decisiones tecnológicas

Según las posibilidades planteadas por ELPD, como se indica en los requerimientos se debe trabajar en plataforma Microsoft. Dentro de este ambiente se decide la utilización de la tecnología .NET como ambiente de desarrollo.

Se elige la plataforma de desarrollo .NET debido a que permite integrar un ambiente web con uno orientado a servicios como lo son los XML web services. Una plataforma con estas características brinda la posibilidad de integrar estos dos ambientes de forma que cooperen entre sí para desarrollar un sistema genérico, escalable y extensible sin perder robustez ni confiabilidad.

Mediante la utilización de XML web services y un ambiente web es posible lograr una combinación interesante de sistemas interoperables. Se obtiene una interacción potente a nivel de datos y se brinda un acceso a usuarios del sistema en una forma estándar y ampliamente conocida como lo es el ambiente web.

5.2. Decisiones de Diseño

Formatos de fechas y horas

Se considera como regla general que todo el sistema deberá tomar como referencia la fecha y hora GMT. Para los intercambios de información el formato a utilizar es:

AAAA-MM-DD hh:mm

Donde:

AAAA: año con cuatro dígitos.

MM: mes con dos dígitos.

DD: día con dos dígitos.

hh: hora con dos dígitos (00-23).

mm: minutos con dos dígitos (00-59)

Criterio para saber si una nota de una fuente esta actualizada

Se considera una nota como actualizada si su fecha es mayor o igual que la fecha de última publicación de la fuente a la que pertenece.

Teniendo como dato previo $H_{PublicaciónFuente}$ la hora en que publica una fuente (ingresada en el momento de registro de la misma), la fecha de la última publicación de una fuente se obtiene de la siguiente forma:

- Se obtiene fecha actual: F_{Actual} .
- Se obtiene hora actual: H_{Actual} .
- Si $H_{Actual} > H_{PublicaciónFuente} \rightarrow D_{UltimaPublicaciónFuente} = Día(F_{Actual})$
- Sino $\rightarrow D_{UltimaPublicaciónFuente} = Día(F_{Actual}) - 1$

$$F_{UltimaPublicaciónFuente} = D_{UltimaPublicaciónFuente} \cdot H_{PublicaciónFuente}$$

Se utiliza nomenclatura AAAA-MM-DD hh:mm según se indica en "Formatos de fechas y horas" (decisión de diseño anterior).

Interfases de comunicación

Para todo tipo de interacción entre fuentes e IC o consumidor e IC se decide utilizar documentos estructurados en formato XML. Esto permite sencillamente estructurar la información, logrando que el receptor sepa la relación entre los datos, ya que el mismo documento XML describe el modelo relacional de los mismos.

Ambientes web y de servicios

Para la implementación de los sistemas requeridos se realiza un agrupamiento en dos ambientes diferentes, el web y el de servicios web como se comentó en "Decisiones tecnológicas".

Las funcionalidades que se presentan en ambiente web se agrupan en dos sitios. Uno de estos sitios, denominado *BackEnd*, se alojará en forma privada y bajo autenticación para fines administrativos del sistema y el otro denominado *FrontEnd*, se presentará en forma pública con un área restringida bajo autenticación para el acceso de fuentes y consumidores a la información de registro.

El ambiente de servicios web comprende la implementación de aquellos requerimientos que se basan en el intercambio de información. Aquí se utiliza un lenguaje estructurado como lo es el XML sobre protocolos ampliamente utilizado en internet (SOAP y HTTP).

Como método de autenticación en cada invocación de servicios web por parte de un consumidor o una fuente sobre el IC se deberán acompañar las solicitudes con el usuario y contraseña aportadas en el momento de registro.

Como método de autenticación a la hora de realizar la invocación de servicios web por parte del IC sobre las fuentes se deberán acompañar las solicitudes con la clave para el servicio web aportada por la fuente en el momento de registro.

Diagrama de base de datos

A continuación, por medio de la Figura 10, se presenta la estructura planteada para la base de datos central a ser utilizada por el sistema IC.

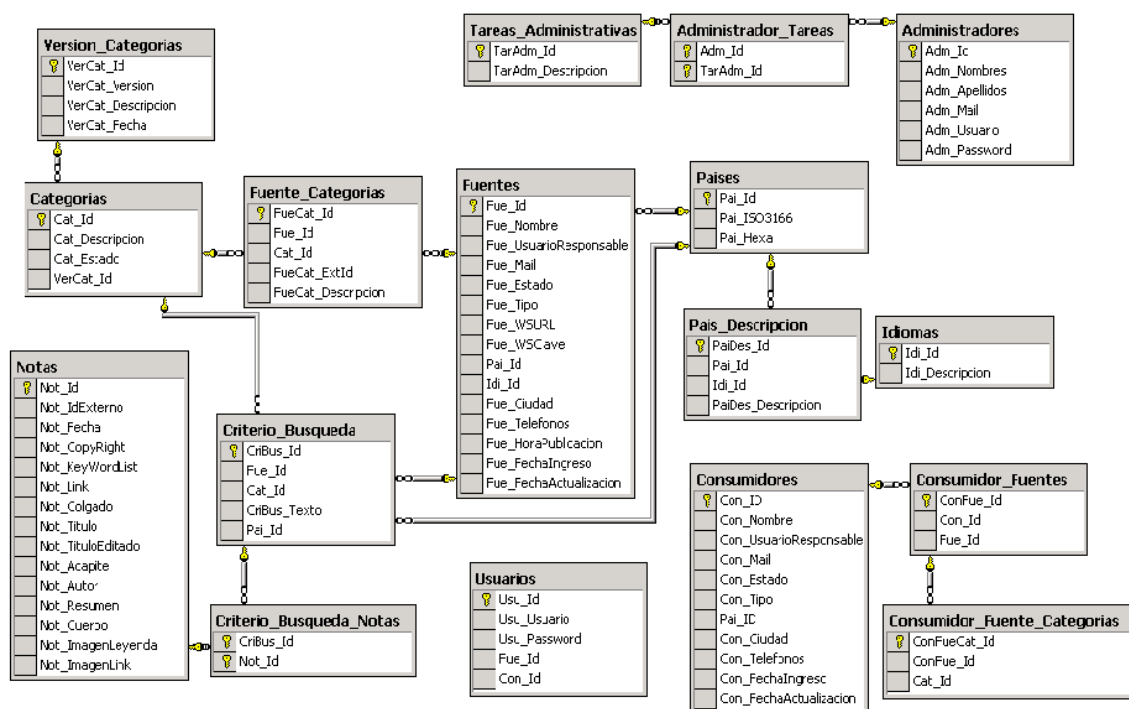


Figura 10: Diagrama de base de datos del sistema.

Categorización del IC

Al momento de integrar contenido de varias fuentes es natural encontrar diferentes criterios de categorización de noticias, siendo esto un inconveniente para afrontar por parte del IC. La solución propuesta para este problema es la creación de una categorización propia. Esta debe ser capaz de incluir cualquier categorización de una fuente que desee participar. Cada fuente que se registra debe definir una correspondencia entre las categorías que utiliza para agrupar contenidos y las categorías del IC. A esta correspondencia se denomina "mapeo" entre categorías.

Debido a que quien conoce la semántica de cada categoría es la propia fuente, se delega definición del mapeo a cada fuente. Esta tarea se puede realizar al momento de registro o posteriormente, pero una fuente no estará habilitada como tal para el IC hasta no haber definido el mapeo.

Previendo la posibilidad de que una categoría de una fuente no se corresponda con ninguna categoría del IC se crea una categoría llamada "Ninguna" para que la fuente asocie a las categorías que considere y se brinda la posibilidad de que una fuente sugiera la incorporación por parte del IC de una nueva categoría.

Consultas con problemas de referencia

Cuando se realizan consultas sobre fuentes que pertenecen a diferentes países, se presenta un problema de referencia, que se puede ver por medio del siguiente ejemplo:

Se necesita información sobre:

Frase: "transferencia"
Categoría: "Deportes"

Si se realiza una consulta por la frase "transferencia" en la categoría "Deportes" en las fuentes pertenecientes al país de donde se desean las noticias, los resultados obtenidos serían los que busca el usuario. Pero sin embargo si se realiza la misma consulta en una fuente perteneciente a otro país se obtendrían noticias que hacen referencia a "transferencia", categorizadas (en el IC) como noticias de "Deportes" pero no del país de interés del usuario.

La solución planteada en este caso es acompañar la consulta con el nombre del país (en el idioma correspondiente) desde el que se desean las noticias para aquellas fuentes que no pertenecen a este país.

En el ejemplo anterior, si se necesita información proveniente de Uruguay sobre:

Frase: "transferencia"

Categoría: "Deportes"

En las fuentes pertenecientes a Uruguay se buscará por noticias que hagan referencia al texto "transferencia" y estén categorizadas como "Deportes". Y en las fuentes pertenecientes a Argentina, se buscarán noticias que hagan referencia a los textos "transferencia" y "Uruguay" y estén categorizadas como "Deportes". La categorización aquí considerada es la perteneciente al IC pero cuando se consulta a una fuente, en realidad las categorías utilizadas serán aquellas que corresponden con la categoría "Deportes" según el mapeo definido por la fuente correspondiente.

Manejo de idiomas

El diseño de la base de datos se realiza de forma que sea posible incorporar la utilización de diferentes idiomas en el sistema.

Para esto se incorpora una tabla donde se almacenan los idiomas a utilizar. Así todos los componentes sensibles al idioma tendrán una referencia a esta tabla. Como ejemplo, el idioma de una fuente es un atributo de la misma y las consultas de los consumidores se realizan indicando en que idioma deben realizarse, permitiendo al IC detectar sobre que fuentes se debe realizar la búsqueda.

En caso de que el IC maneje varios idiomas se delega a los consumidores la tarea de realizar las consultas de sus usuarios en los idiomas correspondientes.

Para el manejo de países se dispone de la estructura mostrada en la Figura 11 que permite incorporar las descripciones de los mismos según los idiomas que se utilizarán en el sistema. Esto es importante porque dichas descripciones se utilizan en las consultas a las fuentes para solucionar "el problema de referencia" analizado anteriormente.



Figura 11: Casos de uso IC.

Criterios de búsqueda

Debido a que el IC dispone solamente de un avance y link de cada noticia, para la realización de búsquedas sobre contenidos integrados se crea un sistema basado en "criterios de búsqueda" como se explica a continuación. Se considera el país P de donde se requieren las noticias, una categoría C del conjunto de categorías

ofrecidas por el IC, una frase F a buscar y un idioma I en el cual se expresa la consulta. A esta cuaterna de valores se le denomina "criterio de búsqueda". Con esta información se realiza lo siguiente:

- 1- Se obtienen todas las fuentes para las cuales el consumidor tiene seleccionadas la categoría C y que tengan como atributo el idioma I.
- 2- Para cada fuente del grupo anterior se busca en el repositorio del IC si hay notas actuales obtenidas con este criterio. Se considera como criterio de actualidad el detallado en el capítulo 5.2.
- 3- En caso afirmativo se devuelven estas noticias como resultado de búsqueda según este criterio.
- 4- En caso negativo:
 - a. Se encuentra las categorías de la fuente que corresponde a la categoría C del IC.
 - b. Si la fuente pertenece al país P, se consulta por noticias pertenecientes a dichas categorías que contengan la frase F.
 - c. Si la fuente no pertenece al país P, se consulta por noticias pertenecientes a dichas categorías que contengan la frase F y la descripción país P (en el idioma de la fuente).
 - d. Se almacena en el repositorio las nuevas notas actualizadas y el criterio de búsqueda relacionado a estas. Cada nota posee como atributo un identificador externo que permite identificar su unicidad en la fuente a la que pertenece. Conociendo esta información, una nota se inserta únicamente en caso de no estar en el repositorio del IC.

Las noticias retornadas como resultado de búsqueda deben ser posteriores a una fecha indicada. Esta es calculada a partir de información que la fuente ingresa como hora de publicación en el momento de registro y pasada como parámetro a la fuente en el momento de la consulta.

A modo de ejemplo, supongamos que un usuario de un consumidor necesita información con el criterio de búsqueda:

Frase: "Tabaré"
Categoría: "Política"
País: "Uruguay"
Idioma: "Español"

Si el consumidor tiene seleccionada la categoría "Política" para las fuentes de idioma español "El País" de Uruguay y "La Nación" de Argentina:

En la fuente "El País" se busca en categoría "Política" la frase "Tabaré" y en la fuente "La Nación" se busca en categoría "Política" las frases "Tabaré" y "Uruguay".

Una vez obtenidos los resultados por parte del consumidor, este deberá implementar la distribución a sus usuarios por el método que el elija. Es

responsabilidad del consumidor la administración de sus usuarios y la invocación del servicio del IC en los diferentes idiomas según el perfil de cada uno de estos.

Se espera que la consulta realizada de esta forma reduzca el tiempo de respuesta, ya que se consulta una única vez por cada criterio a la fuente y en caso consultar nuevamente por le mismo criterio, se obtienen resultados desde el propio catalogo del IC.

5.3. Decisiones de implementación

Criterios de desarrollo

Para todo lo relacionado con nomenclatura y criterios a tener en cuenta a la hora de implementar se definió un documento donde se plantean los lineamientos a seguir. Para obtener mayor detalle sobre estos dirigirse al Anexo III – Criterios de Desarrollo.

Estructuras y lógica de base de datos

En la primera sección del anexo donde se detallan los criterios de desarrollo se detalla la nomenclatura y consideraciones a tener en cuenta a la hora de diseñar una base de datos. Para obtener mayor detalle sobre estos dirigirse al Anexo III – Criterios de Desarrollo.

Todo lo referente a manipulación de datos almacenados en el motor de SQL Server se implementa mediante “procedimientos almacenados” en el propio motor. El lenguaje Transact-SQL que utiliza Microsoft SQL Server permite no solo la realización de consultas SQL sino incorporar cierta lógica programática en estos procedimientos almacenados. Esto brinda la posibilidad de reutilizar códigos desde diferentes partes del sistema y mejora la performance ya que esta lógica queda compilada y optimizada.

Validación de XML

Como se detalla en la Sección 5.2 Decisiones de Diseño, todas las transferencias en el sistema se realizan en formato preestablecidos utilizando lenguaje XML. Debido a que es posible que estos XML sean generados con diferentes herramientas y en diferentes plataformas, es necesario disponer de una forma de validación. Para ello se opta por la utilización de XML Schemas dado a su mayor potencia frente a otras formas de validar.

Lenguaje de programación

Para desarrollar en la plataforma .NET se dispone del ambiente de desarrollo Microsoft Visual Studio .NET aportada por ELPD. Dentro de los lenguajes que esta herramienta permite utilizar se optó por C#. Todos los lenguajes en .NET tienen la misma capacidad de expresión pero se opta por C# dada que la sintaxis se asemeja a la de lenguajes C y C++ conocidos por quienes realizan la tarea de implementación.

Servicios web

Dentro del ambiente orientado a servicios se optó por la utilización de XML web services ya que la herramienta utilizada para desarrollo (Microsoft Visual Studio .NET) está orientada a la utilización de los mismos.

5.4. Subsistemas del IC

En la Figura 12 se representan gráficamente los principales componentes del sistema IC.

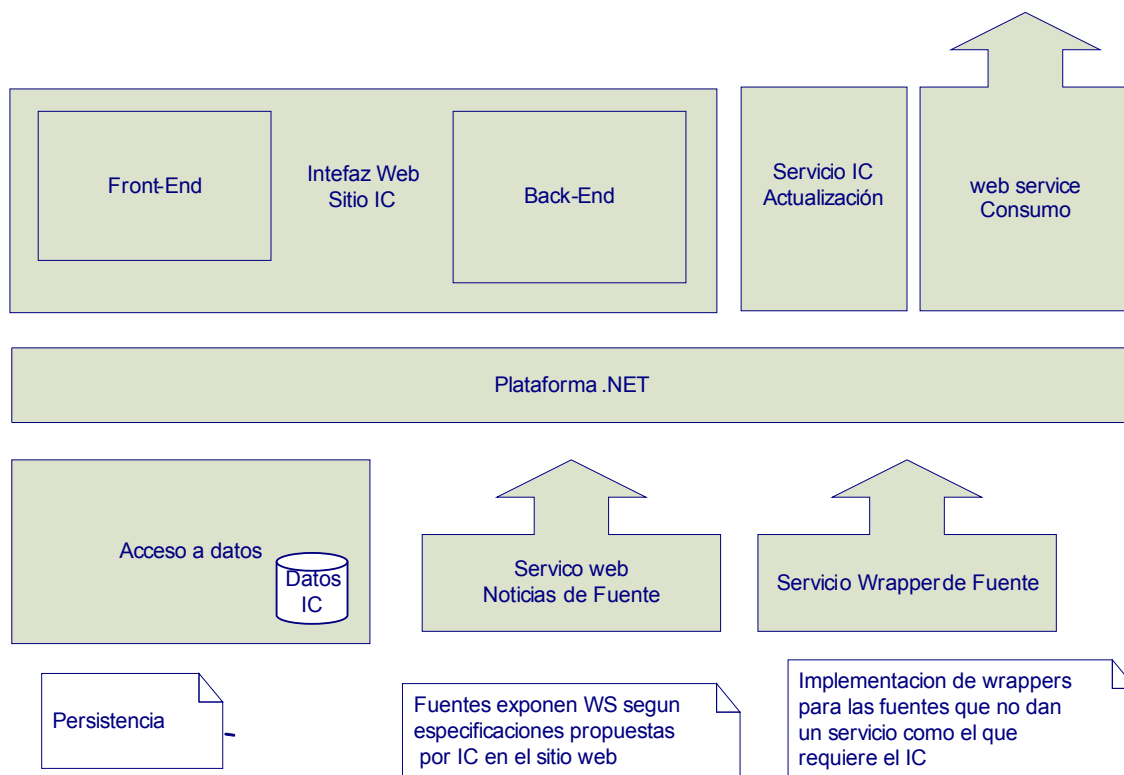


Figura 12: *Subsistemas del IC.*

Cada CU especificado para satisfacer un requerimiento, es implementado en forma de página web o en forma de servicio web según el ambiente al que mejor se adapte.

En el ambiente web se agrupan los CU correspondientes a:

- Suscripción y mantenimiento de registros de fuente y consumidores.
- Administración del sistema.

Esta elección permite centralizar de una manera ágil y cómoda todas las tareas que tienen relación directa con el repositorio central.

El ambiente de servicios web facilita la interacción entre distintos participantes, permitiendo a cada uno de ellos utilizar metodologías y lenguajes propios con la única restricción de respetar la interfaz de los servicios requeridos. Los CU que implican transferencia de información utilizan este mecanismo como forma de compartir información en un formato preestablecido.

La conjunción de estos ambientes, incluido el repositorio dan como resultado la funcionalidad principal de integración de noticias que el sistema debe tener. Así, rutinas de actualización de noticias y criterios de búsquedas junto a métodos de recuperación de los mismos, generan la funcionalidad de integración de contenidos de las diferentes fuentes y lo hacen disponible a sus consumidores.

La parte correspondiente a la interacción de fuentes con el IC se implementa por medio y web services ubicados en cada fuente. La arquitectura utilizada aquí es la de wrappers y mediadores explicada en el Capítulo 2 (Sección 2.9.1). Los wrappers se corresponden con las fuentes que realizan el trabajo de búsquedas y manejo de noticias conociendo su meta información, y el IC cumple función de mediador realizando una integración en una categorización homogénea a todas las fuentes.

5.5 Ambiente Web

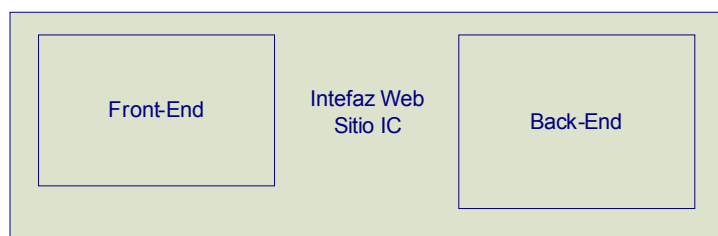


Figura 13: *Ambiente Web*

Como se muestra en la Figura 13, en lo que respecta al ambiente web, se pueden diferenciar dos sitios web, uno correspondiente a la parte de Front-End y otro para la parte de Back-End. A continuación se brinda un detalle de los mismos.

5.5.1 Front-End

El Front-End corresponde a un sitio web donde se agrupan todas las funcionalidades del IC a las que pueden acceder tanto fuentes como consumidores. Es un sitio público al que puede acceder cualquier navegante pero en ciertas áreas se requiere autenticación que demuestre que quien desea acceder corresponde a una fuente o consumidor registrado en el sistema.

En el Front-End se brinda la posibilidad de registro a fuentes y consumidores y existen además otras funcionalidades como la modificación y actualización de los datos de los mismos.

Dentro del registro de una fuente se solicitan datos de contacto de la misma y se dispone de un formulario donde la misma puede definir el mapeo de sus categorías con las categorías del IC. Esta información será usada en forma interna por el IC para realizar la correspondencia entre su categorización y la de cada fuente.

En el momento del registro de un nuevo consumidor, se le solicitan datos de contacto y que indique de que fuentes y que categorías de estas desea consumir. Estos datos quedan almacenados en el perfil del consumidor y se usan en futuras consultas de los consumidores por parte del IC.

Además de estas funcionalidades se ofrecen textos explicativos sobre requerimientos a cumplir por parte de fuentes y consumidores. En caso de considerarse necesario se aportan ejemplos y ayudas para las implementaciones requeridas.

5.5.2 Back-End

El Back-End corresponde a un sitio web restringido a ciertas personas donde se agrupan todas las funcionalidades del IC a las que pueden acceder administradores o usuarios autorizados del sistema. Aquí se ofrecen herramientas para la administración del sistema y se incluyen rutinas de testeo y control.

Las funcionalidades que se pueden encontrar aquí comprenden:

- Control y posibilidad de modificación de datos de registro de consumidores y fuentes.
- Métodos para validación web services que se exigen a las fuentes.
- Métodos para habilitación y deshabilitación de consumidores y fuentes.
- Sistema de envío de e-mails para la realización de notificaciones pertinentes a consumidores y fuentes.
- Administrador de categorías del IC y versionado de las mismas.
- Administración de administradores que tendrán acceso a este sitio y de los permisos correspondientes.

Toda fuente o consumidor que se registra por medio del FrontEnd queda en un estado inhabilitado en el sistema del IC aguardando la habilitación posterior de un administrador para comenzar a actuar en las tareas de integración o consumo.

Al momento de habilitar nuevas fuentes, que previamente se han registrado, hay un mínimo de verificaciones a realizar. Dentro de estos chequeos debe validarse el mapeo de sus categorías con las del integrador y para esto se dispone de herramientas que permiten solicitar noticias a la fuente según el mapeo definido.

Para las fuentes registradas, el IC pone a disposición a en este sitio web la posibilidad de listar la categorización que utiliza. Esto serviría en caso de que una fuente desee analizar en forma desconectada contra que categorías deberá realizar el mapeo de sus categorías.

También se disponen de métodos que permiten realizar una búsqueda sobre una fuente lo que permite al administrador que este haciendo el control corroborar el correcto funcionamiento de las búsquedas.

5.6 Ambiente de Servicios

Como se mencionó anteriormente, este ambiente comprende la implementación de aquellos requerimientos que se basan en el intercambio de información. En el sistema del IC se pueden diferenciar dos grupos de servicios. Los que son utilizados para intercambiar información entre el IC y las fuentes y los que son utilizados entre el IC y sus consumidores. A continuación se detallan características de los mismos.

5.6.1 Fuentes - IC

En la Figura 14 se muestra una representación gráfica de cómo se conforma una fuente una vez que ya esta registrada en el sistema.

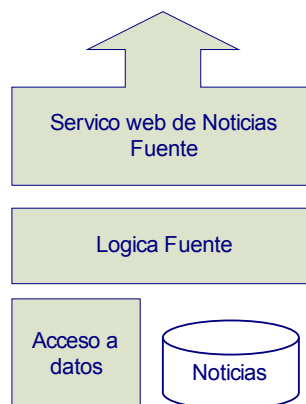


Figura 14: Fuente

Aquí se puede ver que una vez que la fuente ha sido habilitada en el IC, la interacción entre ambos se realiza mediante la invocación de un servicio web.

Para que una nueva fuente sea habilitada y considerada en futuras integraciones por el IC se exige que la misma disponga de un servicio web con los siguientes métodos:

Método	Parámetros	Descripción
NoticiasPorCategoria	pClaveServicio pCategoria pFecha pTextoABuscar pNombrePais	Permite invocar en la fuente una búsqueda de noticias pertenecientes a la categoría <i>pCategoria</i> con fecha mayor a <i>pFecha</i> , que contengan el texto <i>pTextoABuscar</i> . El parámetro <i>pNombrePais</i> es opcional y se utiliza para solucionar el problema de consultas con problema de referencias presentado en la Sección 5.2 del capítulo 5. <i>pClaveServicio</i> es utilizado para autenticar la invocación y contiene la clave suministrada por la fuente al registrarse.
CategoriasFuente	pClaveServicio	Retorna una lista de identificadores de las categorías con sus descripciones correspondiente. <i>pClaveServicio</i> es utilizado para autenticar la invocación y contiene la clave suministrada por la fuente al registrarse.

En el momento del registro, además de ingresar datos de identificación, autenticación y de contacto de la fuente, esta debe suministrar la URL donde se localiza el servicio web anterior y una clave a ser usada como forma de autenticación.

Invocando el método CategoriasFuente, que devuelve el listado de categorías propias de la fuente, se despliega un formulario donde es posible definir el mapeo entre estas categorías y las del IC.

La petición de noticias por parte del IC a cada fuente se hace a través de la invocación del método NoticiasPorCategoría según se indica en la tabla anterior.

Para obtener mayor detalle sobre interfaces utilizadas por los métodos anteriores dirigirse al Anexo V – Interfaces.

Hay casos particulares de fuentes que no brindan la posibilidad de una implementación de servicio del tipo requerido, por lo que para estos casos se plantea la implementación de un wrapper que extraiga al información de la fuente y entregue dicha información al IC y un XML web service que exponga los métodos requeridos según se describió anteriormente. En la Figura 15 se representa gráficamente de cómo estaría formato uno de estos wrappers. En primera instancia

se realiza la extracción de la información en el formato original de la fuente. Una vez que se posee dicha información se proceden a realizar conversiones que permitan publicar un servicio web que implemente los métodos requeridos por el IC.

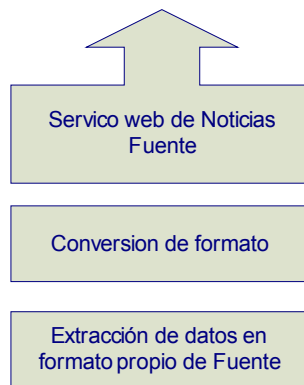


Figura 15: *Wrapper y XML web service*

Dentro del alcance definido para el proyecto se decidió incluir la utilización del periódico La Nación de Argentina como una fuente habilitada en el IC. Para no exigir la creación de un servicio web a dicha fuente se la considera en el caso mencionado anteriormente.

Se implementa para este caso específico un wrapper que extrae noticias en formato RSS y HTML provenientes del sitio web del periódico y se proporciona el XML web service para que el IC pueda realizar las invocaciones necesarias.

5.6.2 Consumidores - IC

A continuación se presenta gráficamente por medio de la Figura 16 como se estructura el consumidor desarrollado para ELPD.

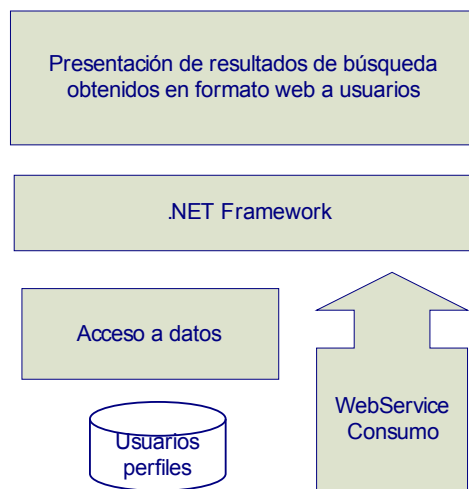


Figura 16: *Consumidor*

Aquí se representa como el consumidor para ELPD posee un repositorio donde tiene almacenado información de perfil de interés de sus usuarios, utiliza el servicio web de consumo aportado por el IC y realizando las transformaciones correspondientes pone a disposición de sus usuarios los resultados de búsquedas.

El IC dispone de un XML web service para los consumidores habilitados con los siguientes métodos:

Método	Parámetros	Descripción
ConsumoNoticias	pUsuarioConsumidor pClaveConsumidor pIdCategoriaIC pTexto pIdPais pIdIdioma	Permite invocar en el IC una búsqueda de noticias actuales del país <i>pIdPais</i> en el idioma <i>pIdIdioma</i> , pertenecientes a la categoría <i>pIdCategoriaIC</i> que contengan el texto <i>pTexto</i> . <i>pUsuarioConsumidor</i> y <i>pClaveConsumidor</i> son utilizados para autenticar la invocación y contienen el usuario y clave respectivamente suministrados por el consumidor al registrarse.
CategoriasICConsumidor	pUsuarioConsumidor pClaveConsumidor	Retorna una lista de identificadores de categorías habilitadas por el IC. Esta lista serviría como referencia a los consumidores para saber que valor utilizar en el parámetro <i>pIdCategoriaIC</i> del método <i>ConsumoNoticias</i> . Parámetros <i>pUsuarioConsumidor</i> y <i>pClaveConsumidor</i> son utilizados para autenticar la invocación y

		contienen el usuario y clave respectivamente suministrados por el consumidor al registrarse.
PaísesIC	pUsuarioConsumidor pClaveConsumidor	Retorna una lista de identificadores de países utilizados el IC. Esta lista serviría como referencia a los consumidores para saber que valor utilizar en el parámetro <i>pIdPais</i> del método <i>ConsumoNoticias</i> . Parámetros <i>pUsuarioConsumidor</i> y <i>pClaveConsumidor</i> son utilizados para autenticar la invocación y contienen el usuario y clave respectivamente suministrados por el consumidor al registrarse.
IdiomasIC	pUsuarioConsumidor pClaveConsumidor	Retorna una lista de identificadores de idiomas utilizados el IC. Esta lista serviría como referencia a los consumidores para saber que valor utilizar en el parámetro <i>pIdIdioma</i> del método <i>ConsumoNoticias</i> . Parámetros <i>pUsuarioConsumidor</i> y <i>pClaveConsumidor</i> son utilizados para autenticar la invocación y contienen el usuario y clave respectivamente suministrados por el consumidor al registrarse.

Los métodos *CategoriasICConsumidor*, *PaísesIC* e *IdiomasIC* brindan información a los consumidores para poder invocar el método principal *ConsumoNoticias* que es el que aporta los resultados de búsqueda.

El rango de búsqueda dentro de los contenidos integrados queda delimitado por las fuentes que se seleccionaron por parte del consumidor en el momento de registro del mismo.

Se considera que las noticias devueltas como resultado de búsquedas por parte de las fuentes consisten en un titular, un pequeño avance y un link al contenido completo en el sitio web de la fuente correspondiente. Estos datos son almacenados en el catalogo del IC y entregados a aquellos consumidores que realicen una búsqueda idéntica evitando así sobrecargar a las fuentes.

De la manera que se plantea la integración se trata a todas las fuentes de la misma manera y se delega el trabajo de nuevas búsquedas a las mismas, dejando para el IC la tarea de vincular las referencias de los contenidos en las fuentes con las preferencias de los consumidores.

Capítulo 6 – PRUEBAS Y RESULTADOS OBTENIDOS

6.1. Pruebas IC

Se realizan las pruebas de funcionalidad del IC según los CU detallados, considerando sus flujos alternativos y el resultado esperado para cada uno de estos. En el esquema de la Figura 17 se muestra la configuración utilizada para hacer las pruebas de funcionalidad del IC.

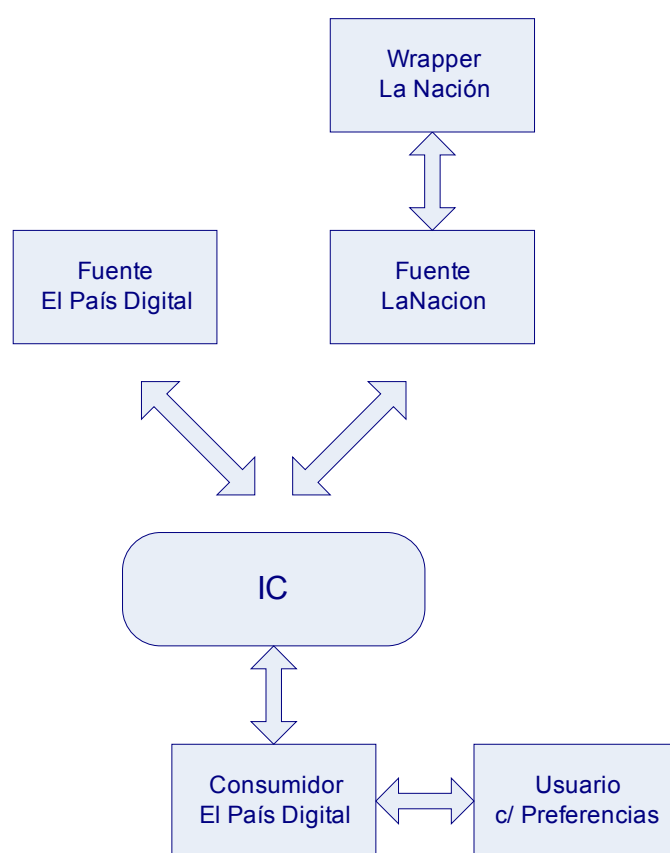


Figura 17: Ambiente de pruebas.

Se registran dos fuentes, una para ELPD de Uruguay y otra para La Nación de Argentina. Para esta última fue necesario el desarrollo de un wrapper que realice la extracción de noticias publicadas en su sitio web en formato RSS y HTML y convertir las mismas a un formato común para el IC. Para poder realizar pruebas de consultas sobre contenidos integrados se registra un consumidor para ELPD.

Esto nos brinda la posibilidad de probar los requerimientos analizados para el sistema consumidor que se plantea desarrollar para ELPD.

Para llegar a este ambiente de pruebas se siguen los CU de registro de fuente y consumidor. Se modifican registros como forma de testeo. Se controlan fuentes y consumidor desde el BackEnd del IC llegando a la habilitación de los mismos.

Para la autenticación de los usuarios del consumidor se utiliza la información que posee ELPD de sus usuarios registrados.

Se realizó un relevamiento entre un grupo de usuarios de ELPD con el fin de obtener perfiles de preferencias para las pruebas. En este relevamiento se recopiló información de preferencias para 276 usuarios diferentes con un total de 1710 criterios de búsqueda.

6.2. Pruebas de transferencias e integración

Las pruebas se realizan en el ambiente anterior teniendo las fuentes y consumidores ya registrados en el IC. Para la realización de las mismas se toma una publicación al azar en cada periódico.

Para testear tiempos de transferencias de noticias entre fuente e IC se consideran dos ambientes de prueba:

- A1: Ambas fuentes ubicadas en la misma computadora donde está instalado el sistema IC.
- A2: Una fuente ubicada en la misma computadora donde está instalado el sistema IC y otra en forma remota utilizando una conexión con enlace de 128kbps.

De acuerdo a lo especificado en la Sección 4.2.2 Requerimientos no funcionales, se utilizan computadoras con sistema operativo Microsoft Windows 2000, motor de bases de datos Microsoft SQL Server 2000 y demás herramientas necesarias para poder ejecutar aplicaciones desarrolladas con la tecnología .NET. Las características de hardware de los equipos utilizados son:

- para alojar el sistema IC y ambas fuentes utilizadas en el ambiente de prueba A1: procesador AMD Sempron 2400+, 512Mb de memoria ram.
- para alojar la fuente desarrollada para La Nación en el ambiente de prueba A2: procesador Intel Pentium 4, 256Mb de memoria ram.

Reutilización de consultas

Las preferencias recolectadas para los usuarios de ELPD se transforman en un conjunto de criterios de búsqueda a ser utilizados para realizar las consultas en el IC según se explica en la Sección 5.2 Decisiones de Diseño.

A continuación se presenta en las Figuras 18 y 19 la relación encontrada entre criterios de búsquedas utilizados y consultas reales efectuadas sobre una fuente.

En la Figura 18 se puede ver como a medida que se van invocando búsquedas sobre el IC disminuye la cantidad de consultas realizadas a la fuente (color oscuro). Esto se debe a la reutilización de criterios de búsquedas que va almacenando el IC en su repositorio. Se incorpora a esta gráfica la recta (color blanco), que representaría el comportamiento en caso de no aplicarse el diseño considerado. Aquí el comportamiento es lineal ya que cada criterio corresponde a una consulta.

Mediante la comparación anterior se puede apreciar que el diseño realizado produce mejoras con respecto a las consultas realizadas.

En la Figura 19 se muestran este resultado en forma de la reutilización de consultas.

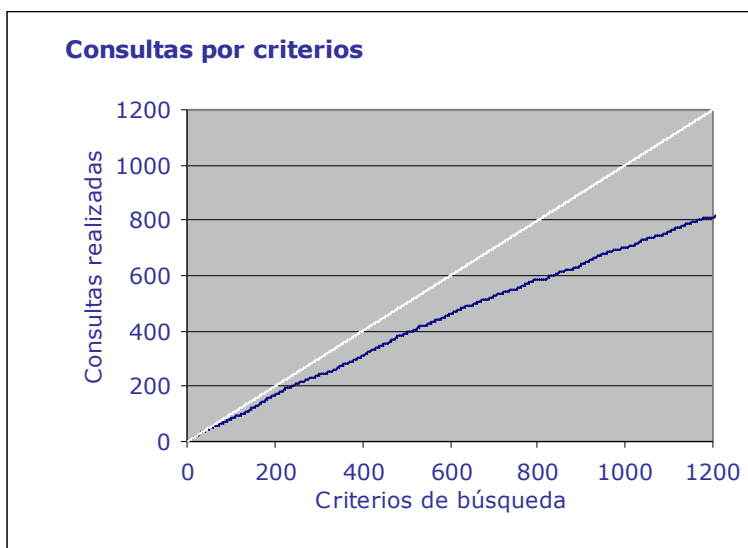


Figura 18: *Criterios de búsqueda vs consultas.*

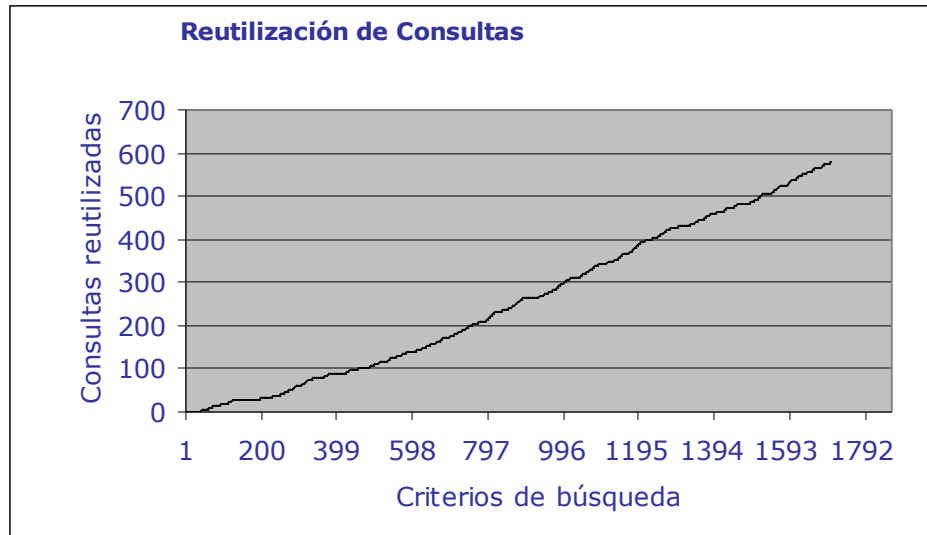


Figura 19: *Reutilización de consultas.*

La reutilización de consultas crece a medida que aumenta el número de criterios de búsquedas y es de esperar que cuantos más usuarios utilicen el sistema, mayor sea la repetición de criterios, mejorando así tiempos de consulta y respuesta para estos usuarios.

Tiempos de respuesta

Para evaluar los tiempos de respuesta del IC se realizan consultas según los diferentes criterios recolectados para los usuarios del ELPD. Las consultas son ejecutadas por el consumidor desarrollado para el ELPD (con una pequeña variación que permite realizar las consultas de todos los usuarios en forma consecutiva) ubicado en el mismo equipo donde se encuentra el IC.

En la Figura 20 se muestran los tiempos de respuesta de las consultas al IC en función de la cantidad de notas procesadas y retornadas que cumplen con los criterios de búsquedas utilizados. Se muestran dos casos, se entiende como local la prueba realizada en el ambiente de prueba A1 y por remota la realizada en el ambiente de prueba A2 detallados al comienzo de esta sección.

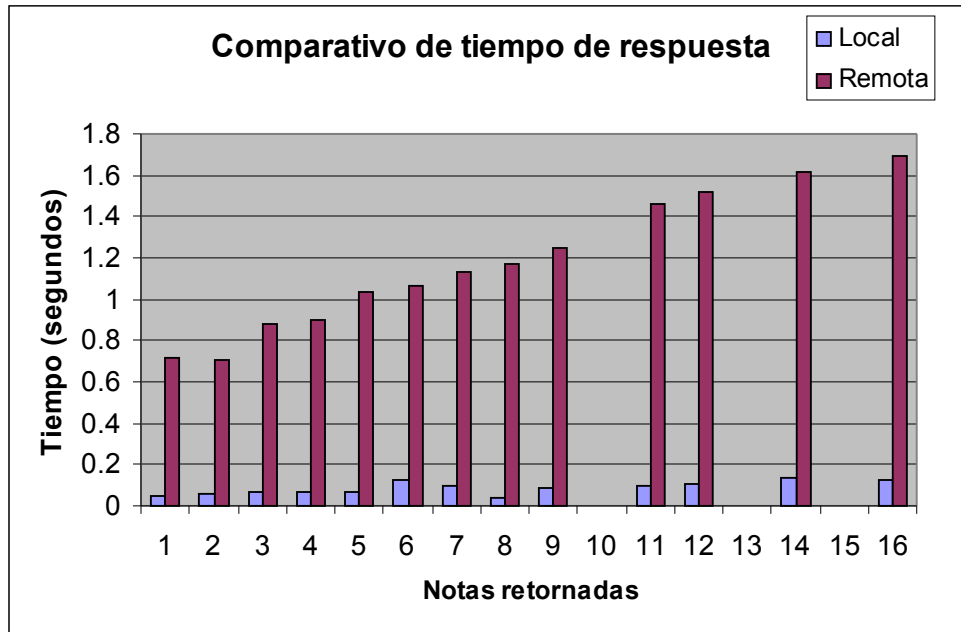


Figura 20: *Transferencias.*

En la gráfica de la Figura 20 se ve que en la invocación local los tiempos de respuesta permanecen no varían en forma notoria, mientras que en la invocación remota este tiempo aumenta en forma proporcional a la cantidad de noticias retornadas. Por lo que se puede ver, la demora más significativa se encuentra en la transferencia de contenidos y esto se debe a que las búsquedas en las fuentes se realizan en una cantidad reducidas de noticias (las pertenecientes a la publicación actual del periódico correspondiente).

Capítulo 7 – CONCLUSIONES Y PASOS A SEGUIR

7.1. Conclusiones

Consideramos que se cumplieron los objetivos planteados:

Se desarrolló una herramienta (sistema IC) que permite realizar la integración de contenidos provenientes de periódicos. Para esto se realizó lo siguiente:

- Creación de una categorización genérica que permite encontrar una correspondencia entre categorías del IC y categorías pertenecientes a las fuentes. Además, se brinda la posibilidad de sugerir incorporación de nuevas categorías por parte de la fuente.
- Se desarrolló un sitio web (FrontEnd) donde consumidores y fuentes pueden registrarse en el sistema IC.
- Se desarrolló un sitio web (BackEnd) donde administradores del IC pueden verificar la correctitud de la información ingresada por fuentes y consumidores. En este sitio también se ponen a disposición métodos que permiten realizar pruebas de las funcionalidades requeridas.
- Se realizó un diseño teniendo en cuenta la incorporación de idiomas en el sistema.
- Se especificaron interfaces de comunicación en formato XML que deberán utilizar fuentes y consumidores.
- Se implementaron los sistemas necesarios para las fuentes correspondientes a los periódicos ELPD de Uruguay y La Nación de Argentina.

Adicionalmente se desarrolló un sistema capaz de ofrecer a los usuarios de ELPD un resumen personalizado de noticias de acuerdo al interés de los mismos, desplegando los resultados en formato web según las especificaciones de la empresa.

Según lo analizado en la Sección 6.2 Pruebas de transferencia e integración, se capta que los tiempos de respuesta del sistema IC dependen en gran parte de la cantidad de noticias transferidas entre las fuentes y el IC. Se espera que el diseño utilizado para manejar criterios de búsqueda según se explica en la Sección 5.2 Decisiones de Diseño, ataque directamente este problema reduciendo la cantidad de consultas a cada fuente.

Las pruebas realizadas muestran que se logró el objetivo de integración planteado por la empresa ELPD cumpliendo requerimientos funcionales y no funcionales.

7.2. Pasos a seguir

En el transcurso del proyecto se fueron analizando alternativas interesantes a ser incluidas en los sistemas desarrollados pero que escapan al alcance planteado originalmente. A continuación se detallan las más significativas:

Optimización de consultas sobre fuentes

Se plantea una variación en la forma que se realizan las transferencias de noticias a la hora de consultar a las fuentes.

Como está implementado el sistema, cada vez que se hace una consulta sobre una fuente esta devuelve un XML en formato especificado conteniendo detalles básicos de las noticias (título, avance, link, identificador de la nota en la fuente, etc.) que verifican el criterio de búsqueda especificado. Se plantea una mejora de esta metodología realizando una invocación posterior a la hora de publicación de cada fuente, donde se traerían los contenidos básicos de todas las noticias publicadas. Cuando se desea consultar a una fuente se envía el criterio o lista de criterios deseados, la fuente realizaría la búsqueda correspondiente y devolvería un listado de identificadores de noticias asociados a los criterios indicados.

Como las noticias fueron previamente recuperadas y almacenadas en el repositorio del IC, la tarea que restaría sería encontrar aquellas notas que correspondan con los identificadores devueltos por la fuente y devolver al consumidor el resultado final.

Con esto se busca ahorrar tiempo de transferencia entre las fuentes y el IC evitando realizar transferencias de los mismos contenidos en forma repetida.

Integración de contenidos completos

En el caso de que una fuente proporcione el contenido completo de todas sus noticias al IC, se plantea la implementación de la siguiente funcionalidad:

Se propone realizar una invocación posterior a la hora de publicación de esta fuente donde se traerían los contenidos completos de todas las noticias publicadas. Posteriormente se indexarían estos nuevos contenidos utilizando herramientas de indexado de SQL Server.

Cada vez que un consumidor realiza una consulta con un criterio de búsqueda, se sigue la misma metodología de búsqueda aplicada actualmente, pero realizando la misma en forma local para aquellas fuentes de las que se dispone contenidos completos. Se esperaría de esta manera bajar los tiempos de transferencia a costo de mayor proceso local para la indexación y búsquedas.

Actualización periódica

Las fuentes de noticias son periódicos o prensa escrita por lo tanto realizan una publicación diaria. Si bien algunas tienen avances de noticias durante el transcurso de la jornada se considera como publicación la primera que realiza la fuente en el día.

Se plantea como una mejora en este sentido la implementación de un módulo que regularmente realice las consultas por nuevas noticias a las fuentes según los criterios de búsqueda más frecuentemente usados con el fin de mantener actualizado el repositorio del IC. Se mejoraría el tiempo de respuesta del IC a sus consumidores.

Consultas a fuentes en formato XML

Las consultas que se puede invocar a una fuente con el sistema actual son búsquedas simples, exactas y de un único texto, o combinación de dos textos con operador *AND* de SQL. Se plantea como mejora expresar en un formato XML a determinar consultas sobre las fuentes.

Con esto se podría aumentar la capacidad de consulta al permitir la combinación de operadores de SQL y el uso de diferentes tipos de búsqueda como prefijos, semejanza, etc.

Capítulo 8 – GLOSARIO Y REFERENCIAS

8.1. Glosario

<i>Consumidor</i>	Sistema al que el IC permite consultar contenidos integrados con el fin de brindar a sus usuarios información de acuerdo a sus preferencias.
<i>CLR</i>	Common Language Runtime: Núcleo de la plataforma .NET. Es el motor encargado de gestionar la ejecución de las aplicaciones para ella desarrolladas y a las que ofrece numerosos servicios que simplifican su desarrollo y favorecen su fiabilidad y seguridad.
<i>COM</i>	Component Object Model: Modelo utilizado para desarrollar componente de software reutilizables.
<i>COM+</i>	Integración de la arquitectura COM y MTS.
<i>Criterio de Búsqueda</i>	Cuaterna de valores (país de donde se requieren noticias, categoría del conjunto de las ofrecidas por el IC, frase a buscar e idioma en el cual se expresa la consulta) utilizado por el IC para representar una búsqueda sobre contenidos integrados.
<i>El País Digital</i>	Sección de El País S.A. que presenta el proyecto.
<i>ELPD</i>	Abreviatura para El País Digital, sección de El País S.A. que presenta el proyecto.
<i>Fuente</i>	Sistema que aporta contenidos a integrar.
<i>GMT</i>	Greenwich Meridian Time. Tiempo local al meridiano (0º) que pasa y esta marcado en una placa en el observatorio de Grennwich en Inglaterra para unificara el tiempo en todo el mundo.
<i>HTML</i>	Hypertext Markup Language: Lenguaje de programación utilizado para crear páginas web a ser interpretadas por los navegadores.
<i>HTTP</i>	Hypertext Transfer Protocol: Protocolo que gobierna la transferencia de hipertexto entre dos o más computadoras.
<i>IC</i>	Sistema integrador de contenidos.

<i>Mapeo de categorías</i>	Correspondencia entre las categorías que utiliza una fuente para agrupar contenidos y las categorías del IC.
<i>MTS</i>	Microsoft Transation Server: permite a componentes de software ser distribuídos por la red y que se ejecuten en otras computadoras con sistema operativo Windows NT Server.
<i>RSS</i>	Really Simple Syndication: Es una forma de facilitar contenidos desde cualquier sitio en la red para su inserción fácil en una página web o para ser consumidor por un software de escritorio.
<i>SOAP</i>	Simple Object Access Protocol: Protocolo simple basado en XML que permite a las aplicaciones intercambiar información sobre HTTP
<i>TI</i>	Tecnología de la información.
<i>Usuario</i>	Persona que se registra ante un consumidor con el fin de obtener información de acuerdo a su perfil de interés.
<i>WS</i>	Los Servicios Web (Web Service) permiten a las aplicaciones compartir información e incluso invocar funciones de otras aplicaciones independientes, sin importar el sistema operativo, plataforma en la que se ejecuten o incluso el dispositivo que se utilice para acceder a ellos.
<i>XMLws</i>	XML web service: Son aplicaciones empresariales basadas en web que usan estándares abiertos de XML y protocolos de transporte estándar para el intercambio de datos con los clientes que lo invocan

8.2. Referencias

- [1] DBAID. Sistema Experto para el diseño de Bases de Datos. Herman E. Dolder, Esteban Lubomirsky. Primera Edición, 1988.
- [2] An Overview of Multidatabase Systems: Past and Present. Athman Bouguettaya, Boualem Benatallah, Ahmed Elmagarmid. 1998.
- [3] Diccionario enciclopédico ESPASA. 2000.

- [4] KM World
Web <http://www.doc6.es/secciones/kmworld.asp>
Local [Referencias/Anexos/AnexoI/KMWorld.pdf](#)
Ultima visita realizada 30/10/2004.
- Web http://www.kmworld.com/publications/magazine/index.cfm?action=readarticle&Article_ID=1741&Publication_ID=109
Local [Referencias/Anexos/AnexoI/KMWorld-Seleccion2004.pdf](#)
Ultima visita realizada 30/10/2004.
- [5] Anacubis
Web <http://www.anacubis.com/>
Local [Referencias/Anexos/AnexoI/KMWorld-Anacubis.pdf](#)
Ultima visita realizada 30/10/2004.
- [6] Arbortext
Web <http://www.arbortext.com/>
Local [Referencias/Anexos/AnexoI/KMWorld-Arbortext.pdf](#)
Ultima visita realizada 30/10/2004.
- [7] TripleHop Technologies, Inc
Web <http://www.triplehop.com/>
Local [Referencias/Anexos/AnexoI/KMWorld-Triplehop.pdf](#)
Ultima visita realizada 01/11/2004.
- Web <http://www.triplehop.com/ie/solutions/knowledge.html>
Local [Referencias/Anexos/AnexoI/KMWorld-Triplehop-EnterpriseSearch.pdf](#)
Ultima visita realizada 01/11/2004.
- [8] Google
Web <http://www.google.com.uy>
Local [Referencias/Anexos/AnexoI/Google.pdf](#)
Ultima visita realizada 13/11/2004.
- Web <http://www.google.com.uy/intl/es/options.html>
Local [Referencias/Anexos/Anexo1/Google2.pdf](#)
Ultima visita realizada 13/11/2004.
- [9] Google News
Web <http://news.google.es/>
Local [Referencias/Anexos/Anexo1/GoogleNewsEspana.pdf](#)
Ultima visita realizada 13/11/2004.
- [10] Directorio web de Google
Web <http://www.google.com.uy/dirhp?hl=es>

Local [Referencias/Anexos/Anexo1/GoogleDirectorioWeb.pdf](#)
Ultima visita realizada 13/11/2004.

- [11] Microsoft Developer Networks
<http://msdn.microsoft.com/>
Ultima visita realizada 30/01/2005.
- [12] Curso de Microsoft: MSC024
Contenido: Developing Microsoft ASP .NET Web Applications Using Visual Studio .NET
Developing XML web services Using Microsoft ASP .NET.
- [13] Building Web Solutions with ASP.NET And ADO.NET
Microsoft. Dino Esposito.
- [14] ASP.NET Unleashed, Second Edition. Stphen Walther. 18/07/2003
- [15] Professional ASP.NET 1.0 Special Edition. Richard Anderson, Brian Francis, Alex Homer, Rob Howard, David Sussman, Karli Watson. 2002.