

Hacia la Retroalimentación del Procesamiento de Flujos de Datos Sustentado en Memoria Organizacional

Diván, Mario José^{1,2}, Martín, María de los Ángeles¹ & Olsina, Luis¹

¹*Facultad de Ingeniería, Universidad Nacional de La Pampa*

²*Facultad de Ciencias Económicas y Jurídicas, Universidad Nacional de La Pampa*

Abstract

Este trabajo evoluciona el conjunto de datos de entrenamiento del enfoque integrado de procesamiento de flujos de datos centrado en metadatos de mediciones (EIPFDcMM), llevándolo desde un repositorio relacional a una memoria organizacional con razonamiento basada en casos (MOBC). Esto permite complementar las alarmas pro activas del EIPFDcMM con un tratamiento recomendado sobre la entidad bajo análisis ante una situación dada. Dicho tratamiento es obtenido mediante razonamiento basado en casos desde la MOBC. Se empleó el esquema CINCAMI/MIS como medio para portar la información entre la MOBC y el EIPFDcMM, requiriendo organizar la MOBC mediante una estrategia clave – valor para gestionar información con diferentes niveles de estructuración. Esto representa un avance importante, por cuanto permite no solo establecer un mecanismo de retroalimentación a partir del EIPFDcMM, sino que permite a partir de la MOBC complementar las alarmas con cursos de acción basados en la experiencia previa. Con el fin de ilustrar la estrategia, se utiliza una prueba de concepto.

Palabras Clave

Flujos de Datos. Memoria Organizacional. Medición y Evaluación. C-INCAMI.

1. Introducción

Existen aplicaciones que procesan un conjunto de datos en forma continua, y ante cada arribo [1]. Dentro de dicho tipo de aplicaciones, se enmarca el enfoque integrado de procesamiento de flujos de datos centrado en metadatos de mediciones (EIPFDcMM)[2], el cual está especializado para flujos de datos asociados con proyectos de medición y evaluación (M&E). A diferencia de las aplicaciones de procesamiento de flujos de datos tradicionales, el EIPFDcMM se sustenta en C-INCAMI (*Context-Information Need, Concept model, Attribute, Metric and Indicator*)[3,4], como marco conceptual de

un proyecto de M&E, a los efectos de incorporar consistencia en el procesamiento, permitiendo la comparabilidad de las medidas en el tiempo.

Si bien el EIPFDcMM incorpora un comportamiento preventivo mediante clasificadores incrementales, o bien, en el peor de los casos, detectivo a través de diversos análisis estadísticos, no cuenta actualmente con la posibilidad de incorporar lecciones aprendidas o conocimientos previos.

Por otro lado, los sistemas de gestión de conocimiento (KMS) tienen como objetivo administrar el conocimiento organizacional, de modo que después pueda ser empleado para aprender, resolver problemas y como apoyo a la toma de decisiones[5,6]. Los sistemas actuales gestionan el conocimiento en forma no estructurada, demandando demasiados recursos y esfuerzo para su gestión [15]. En tal sentido, una solución alternativa a dicho problema, es gestionar el conocimiento en forma estructurada, lo que se denomina Memoria Organizacional (MO)[6].

De este modo, el presente trabajo plantea como contribuciones específicas, *i) una estrategia para retroalimentar la memoria organizacional* estructurada como <clave,valor>, a partir de las alarmas y las clasificaciones surgidas desde el tomador de decisiones del EIPFDcMM, y *ii) una estrategia para aplicar la experiencia adquirida y almacenada en la memoria organizacional al procesamiento de flujos de datos*, para mejorar progresivamente la toma de decisiones en base a las lecciones aprendidas. Esto representa un importante avance con respecto a [7], ya que ahora se

plantea la posibilidad de modelar e integrar el conocimiento de una organización, al procesamiento de flujos de datos y a la toma de decisiones preventiva.

El presente artículo se organiza en seis secciones. La sección 2 resume el procesamiento de flujos de datos centrado en mediciones. La sección 3 sintetiza el marco de memoria organizacional basada en casos. La sección 4 ilustra una prueba de concepto, basado en un sistema de recomendación de diagnóstico médico simplificado, con el objetivo de esquematizar la integración y aplicación de la memoria organizacional al procesamiento de flujos de datos, y viceversa. La sección 5 discute los trabajos relacionados, y por último, en la sección 6, se resumen las conclusiones y trabajos a futuro.

2. Procesamiento de flujos de datos centrado en metadatos de mediciones

El Enfoque Integrado de Flujos de Datos Centrado en Metadatos de Mediciones (EIPFDcMM)[2] es una estrategia para el procesamiento de flujos de datos a partir de fuentes de datos heterogéneas, especializado en proyectos de Medición y Evaluación (M&E), e implementada en un prototipo de software. A los efectos de garantizar la comparabilidad de las medidas en el tiempo, la estrategia se sustenta en el marco conceptual de medición y evaluación C-INCAMI[4,3], el cual define los módulos, conceptos y relaciones que intervienen en el área de M&E.

Básicamente, el EIPFDcMM procesa las medidas provenientes de las diversas fuentes, efectuando diversos análisis estadísticos en línea, y aplicando clasificación incremental sobre los flujos de datos, con el objetivo de implementar un comportamiento preventivo sobre la entidad bajo análisis, o bien, detectivo en el peor de los casos[8]. De este modo, la posibilidad de incorporar el conocimiento proveniente de una memoria organizacional basada en casos, enriquecería el comportamiento preventivo del EIPFDcMM, disminuyendo

la incertidumbre al momento de la generación de los distintos modelos.

Así, la presente sección se divide en dos sub-secciones. En la primera sub-sección se da un panorama de C-INCAMI, como fundamento para la definición de un proyecto de M&E; mientras que en la segunda sub-sección, se brinda una síntesis del EIPFDcMM, como herramienta de procesamiento de los flujos de datos.

2.1 Panorama de C-INCAMI

C-INCAMI es un marco conceptual [4,3] que define los módulos, conceptos y relaciones que intervienen en el área de M&E. Se basa en un enfoque en el cual la especificación de requerimientos, la medición y evaluación de entidades y la posterior interpretación de los resultados están orientadas a satisfacer una necesidad de información particular. Está integrado por los siguientes componentes principales: 1) Gestión de Proyectos de M&E; 2) Especificación de Requerimientos no Funcionales; 3) Especificación del Contexto del Proyecto; 4) Diseño y Ejecución de la Medición; y 5) Diseño y Ejecución de la Evaluación. La mayoría de los componentes están soportados por los términos ontológicos definidos en [3]. En la Figura 1 se muestra un diagrama con los principales conceptos y relaciones para los componentes de requerimientos, contexto y medición. Los conceptos de la Figura 1, definen la estructura de los metadatos que deberán ser persistentes en el EIPFDcMM, para poder llevar adelante el proceso de M&E de una o más entidades bajo análisis (C-INCAMI DB, en Figura 2).

Los flujos de medidas que se informan desde las fuentes de datos al prototipo, se estructuran incorporando a las medidas, metadatos basados en C-INCAMI tales como la métrica a la que corresponde, el grupo de seguimiento asociado, el atributo de la entidad que se mide, entre otros. Dentro del flujo, se etiquetan conjuntamente con cada medida asociada al atributo, las medidas asociadas a cada propiedad de contexto. Gracias a la

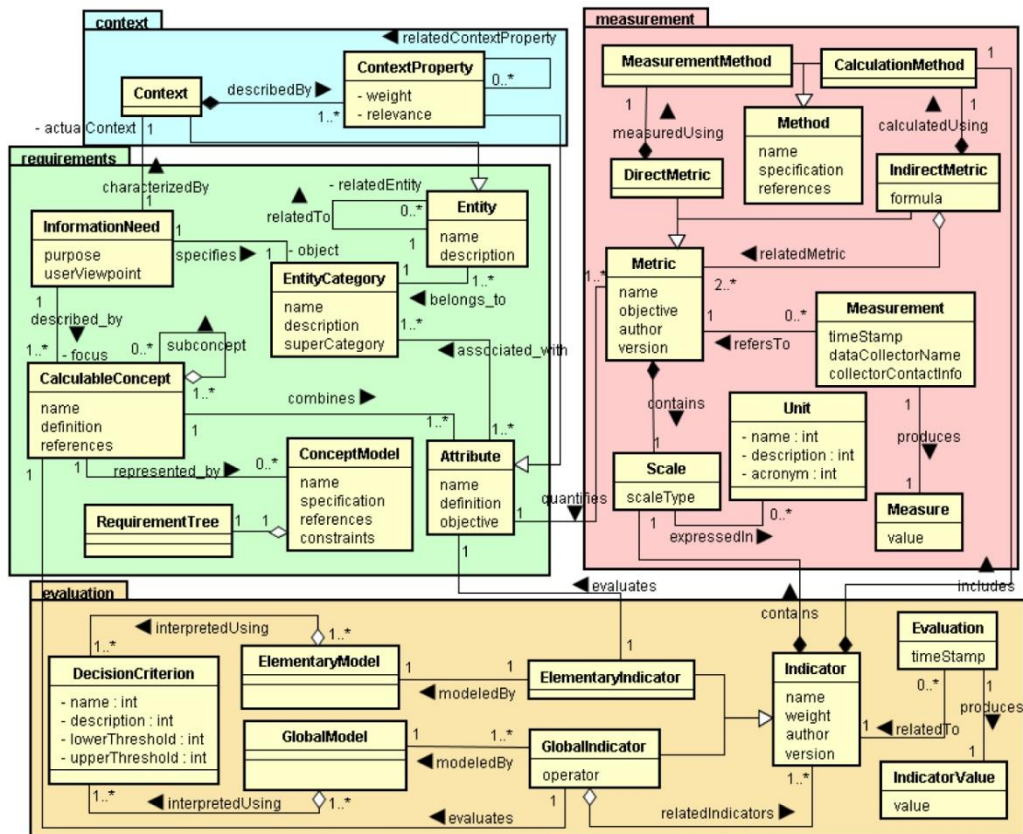


Figura 1. Principales conceptos y relaciones de los componentes Especificación de Requerimientos no Funcionales, Especificación del Contexto y de la Medición.

formalización del proyecto de M&E en base a C-INCAMI, el hecho de procesar el flujo etiquetado, permite la estructuración del contenido de un modo consistente y alineado con el objetivo del proyecto de M&E. Esta estructuración de las mediciones dentro del prototipo, mantiene el concepto con el que se asocia cada medida; por ejemplo, si es una medida de atributo o bien de propiedad contextual. De este modo, se enriquece el análisis estadístico dado que es posible en forma directa, verificar la consistencia formal y sintáctica de cada medida contra su definición formal previo a avanzar con técnicas estadísticas de análisis apropiadas. En [8] se presentó un escenario de aplicación del EIPFDcMM referido al monitoreo de pacientes trasplantados, el cual definía el proyecto de M&E en base a C-INCAMI a los efectos de automatizar el proceso de recolección, análisis y evaluación de las mediciones. Así, en forma sintética y a modo de ejemplo, la necesidad de información era “monitorear los

principales signos vitales en un paciente trasplantado al momento en que se le da el alta desde el centro médico” (ver Figura 1, el componente *Requirements*). La entidad bajo análisis era representada por el paciente trasplantado ambulatorio. Según los expertos, la *temperatura corporal*, la *presión arterial sistólica* (máxima), la *presión arterial diastólica* (mínima) y la *frecuencia cardiaca* representaban los atributos de los signos vitales relevantes a monitorear en ese tipo de pacientes. Además, los expertos señalaban que era necesario monitorear la *temperatura ambiental*, la *presión ambiental*, la *humedad* y la *posición del paciente* (latitud y longitud) como parte de las propiedades de contexto (ver Figura 1, *Context*). La cuantificación de los atributos se realiza por medio de las métricas (ver Figura 1, *Measurement*). De este modo, para el monitoreo se definieron métricas que cuantificasen a cada uno de los atributos citados, a saber: la presión arterial sistólica, presión arterial diastólica, temperatura

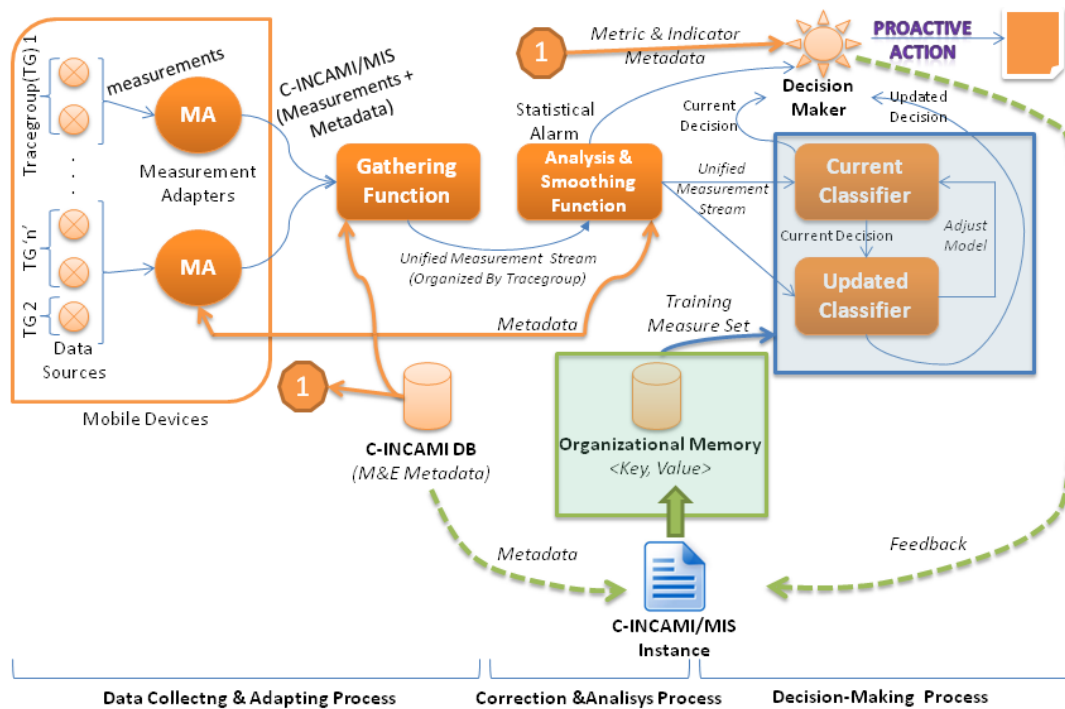


Figura 2. Esquema Conceptual del EIPFDcMM

corporal y frecuencia cardiaca. En cuanto a las propiedades de contexto, se deseaba contar con un monitoreo sobre la temperatura ambiental, la presión ambiental, la humedad y la posición del paciente. Definidas las métricas, se definieron los indicadores elementales (ver Figura 1, *Evaluation*), a los efectos de sentar la base para la interpretación de los atributos y conceptos calculables. De este modo y como puede apreciarse en la Figura 1, el marco conceptual permite establecer los módulos, conceptos y relaciones necesarias para automatizar el proceso de recolección, análisis, y evaluación de las diferentes medidas

2.2 Panorama de EIPFDcMM

El *EIPFDcMM*[2] es una estrategia orientada al procesamiento de flujos de datos originados en fuentes de datos heterogéneas. El prototipo recibe flujos de datos estructurados según un esquema basado en XML denominado CINCAMI/MIS por *Measurement Interchange Schema*. En dicho esquema, se incorporan los conceptos subyacentes a C-INCAMI con el objetivo de estructurar y relacionar los valores sintácticos en función

del rol que cada concepto desempeña dentro del proceso de M&E. Así, cada flujo de datos, además de informar las medidas, informará también sus metadatos asociados, lo que permitirá obtener su definición formal, posibilitando de este modo análisis consistentes y comparables desde el punto de vista estadístico.

Como puede apreciarse en la Figura 2, las mediciones se generan en las fuentes de datos heterogéneas, las cuales abastecen a un módulo denominado adaptador de mediciones (MA en Figura 2) generalmente embebido en dispositivos móviles por una cuestión de portabilidad y practicidad, aunque podría embeberse en cualquier dispositivo de cómputo con asociación a fuentes de datos. MA incorpora junto a los valores medidos, los metadatos de medición y los informa a una función de reunión central (Gathering Function –GF). GF incorpora los flujos de mediciones estructurados en base a CINCAMI/MIS en un buffer organizado por grupos de seguimiento –modo dinámico de agrupar a las fuentes de datos definido por el director del proyecto de M&E-, con el objeto de permitir análisis estadísticos consistentes a nivel de grupo de seguimiento o bien por

región geográfica donde se localicen las fuentes de datos, sin que ello implique una carga adicional de procesamiento. Adicionalmente, GF incorpora técnicas de load shedding [9] que permiten gestionar la cola de servicios asociada a las mediciones, mitigando los riesgos de desborde independientemente el modo en que se agrupen.

Una vez que las mediciones se encuentran organizadas en el buffer, se aplica análisis descriptivo, de correlación y componentes principales (Analysis & Smoothing Function –ASF-) guiados por sus propios metadatos, a los efectos de detectar situaciones inconsistentes con respecto a su definición formal, tendencias, correlaciones y/o identificar las componentes del sistema que más aportan en términos de variabilidad. De detectarse alguna situación en ASF, se dispara una alarma estadística al tomador de decisiones (Decision Maker -DM) para que evalúe si corresponde o no disparar la alarma externa (vía, e-mail, SMS, etc) que informe al personal responsable de monitoreo sobre la situación, y eventualmente, se fundamenta aquí la posibilidad de retroalimentar una memoria organizacional a partir de la nueva decisión y-o experiencia. En paralelo los nuevos flujos de mediciones son comunicados al clasificador vigente (Current Classifier –CC-), quien deberá clasificar las nuevas mediciones si corresponden o no a una situación de riesgo e informar dicha decisión al DM. Simultáneamente, se reconstruye el CC incorporando las nuevas mediciones al conjunto de entrenamiento y produciendo con ellas un nuevo modelo (Updated Classifier -UC). El UC clasificará las nuevas mediciones y producirá una decisión actualizada que también será comunicada al DM. Así y a los efectos de inicializar los clasificadores, en [2] se emplea un conjunto de entrenamiento relacional, por otro lado aquí se fundamenta el hecho de evolucionar tal repositorio hacia la forma de una memoria organizacional estructurada bajo el esquema <Clave, Valor> (Ver Figura 2),

lo cual permitiría no solo entrenar al clasificador a los efectos de su inicialización, sino también capitalizaría la experiencia previa y retroalimentaría la memoria a partir de las nuevas decisiones y-o experiencias formuladas por el tomador de decisiones del EIPFDcMM.

El DM determina si las decisiones indicadas por los clasificadores (CC y UC) corresponden a una situación de riesgo y en cuyo caso con qué probabilidad de ocurrencia, actuando en consecuencia según lo definido en el umbral mínimo de probabilidad de ocurrencia definido por el director del proyecto de M&E. Finalmente, independientemente de las decisiones adoptadas, el UC se torna en CC sustituyendo al anterior, en la medida que exista una mejora en su capacidad de clasificación según el modelo de ajuste basado en curvas ROC (Receiver Operating Characteristic)[10].

La Figura 2 expone sólo un esquema conceptual del funcionamiento del EIPFDcMM, útil para su introducción general. Ahora bien, la estrategia cuenta con una definición formal de sus procesos en base a SPEM (Software & Systems Process Engineering Metamodel)[11].

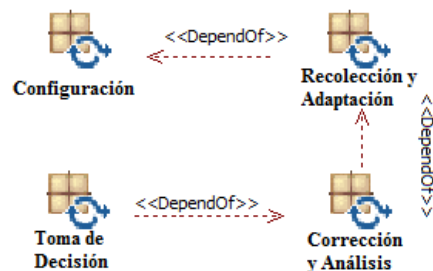


Figura 3. Dependencia entre procesos del EIPFDcMM usando SPEM

En la definición formal de la estrategia[7], pueden observarse cuatro procesos: a) Configuración, b) Recolección y Adaptación, c) Corrección y Análisis y d) Toma de Decisión. Las dependencias entre los mencionados procesos pueden observarse en la Figura 3.

El proceso de configuración, para un proyecto de M&E dado, tiene por objetivo establecer la correspondencia entre las fuentes de datos asociadas con un MA, y las métricas respectivas vinculadas con un

atributo de la entidad bajo análisis, o bien, con sus propiedades contextuales. Este proceso se da una única vez, al momento de inicializar un dispositivo de medición con respecto a EIPFDcMM. Los procesos de Recolección y Adaptación, Corrección y Análisis y Toma de Decisión, se dan en forma posterior a la configuración, y su alcance funcional ha sido introducido junto con la Figura 2, en donde puede observarse incluso el seccionamiento inferior de la misma en base a los mencionados procesos.

3. Memoria Organizacional basada en Casos

Los sistemas de administración del conocimiento (KMS) tienen como objetivo administrar y almacenar el conocimiento organizacional, de manera que después pueda ser utilizado para aprender, resolver problemas y como apoyo en la toma de decisiones [5,6]. Una estrategia para ello, es almacenar el conocimiento en forma estructurada en lo que se denomina Memoria Organizacional Basada en Casos (MOBC), que permita contestar preguntas como: ¿Porqué se hizo esto de cierta manera?, ¿Este problema ha sido resuelto antes?, ¿Qué aprendimos la última vez que sucedió ese problema?. En tal sentido, aquí se fundamenta la sustitución de un repositorio relacional con datos de entrenamiento para el EIPFDcMM, por una memoria organizacional que permita entrenar a sus clasificadores, y adicionalmente, retroalimentarse a partir de nuevas experiencias (Ver Figura 2).

Un caso es una pieza contextualizada de conocimiento que representa una experiencia. Contiene la lección pasada que es el contenido del caso y el contexto en el cual la lección puede ser utilizada [12]. Típicamente, un caso comprende:

- El *problema* que describe el estado del mundo cuando ocurrió el caso.
- La *solución* que describe cómo se resuelve el problema, y/o
- El resultado que describe el resultado obtenido como consecuencia de la solución del problema.

El proceso de razonamiento basado en casos consiste en asignar valores a las variables de características del problema (caracterizar el problema), y encontrar los valores adecuados para las instancias de la solución, a través de criterios de evaluación de similitud de casos.

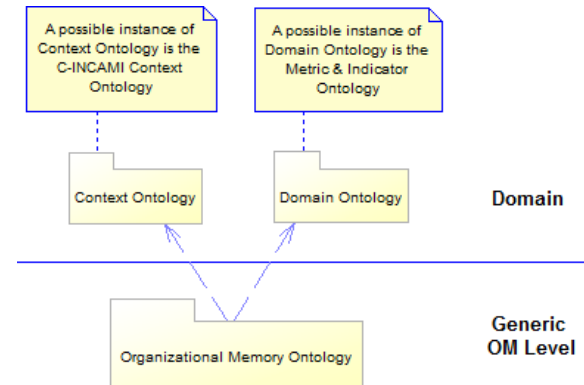


Figura 4. Los dos niveles de ontologías del marco conceptual

Tradicionalmente, hay varios tipos de métodos para representar casos, que van desde representaciones no estructuradas a totalmente formales y automáticamente procesables[13]. Estos últimos, están basados en representaciones totalmente estructuradas, y consisten en aplicar, por ejemplo, técnicas orientadas a objeto centradas en el uso de metadatos.

El marco de memoria organizacional propuesto está basado en ontologías[14] que operan en dos niveles distintos de abstracción (Ver Figura 4): Por un lado, en el nivel de memoria organizacional genérico, se define la ontología de memoria organizacional en sí (esta ontología será resumida brevemente en la sub-sección siguiente); y por otro lado, para caracterizar los casos de acuerdo al dominio de conocimiento y teniendo en cuenta su contexto[4], se necesita proveer al marco con ontologías de dominio y contexto respectivamente (ontologías de nivel de dominio).

3.1 Ontología de Memoria Organizacional Basada en Casos

En esta sección se describen los principales conceptos de la ontología de memoria organizacional basada en casos [15], que se ilustran en el diagrama UML de la Figura 5.

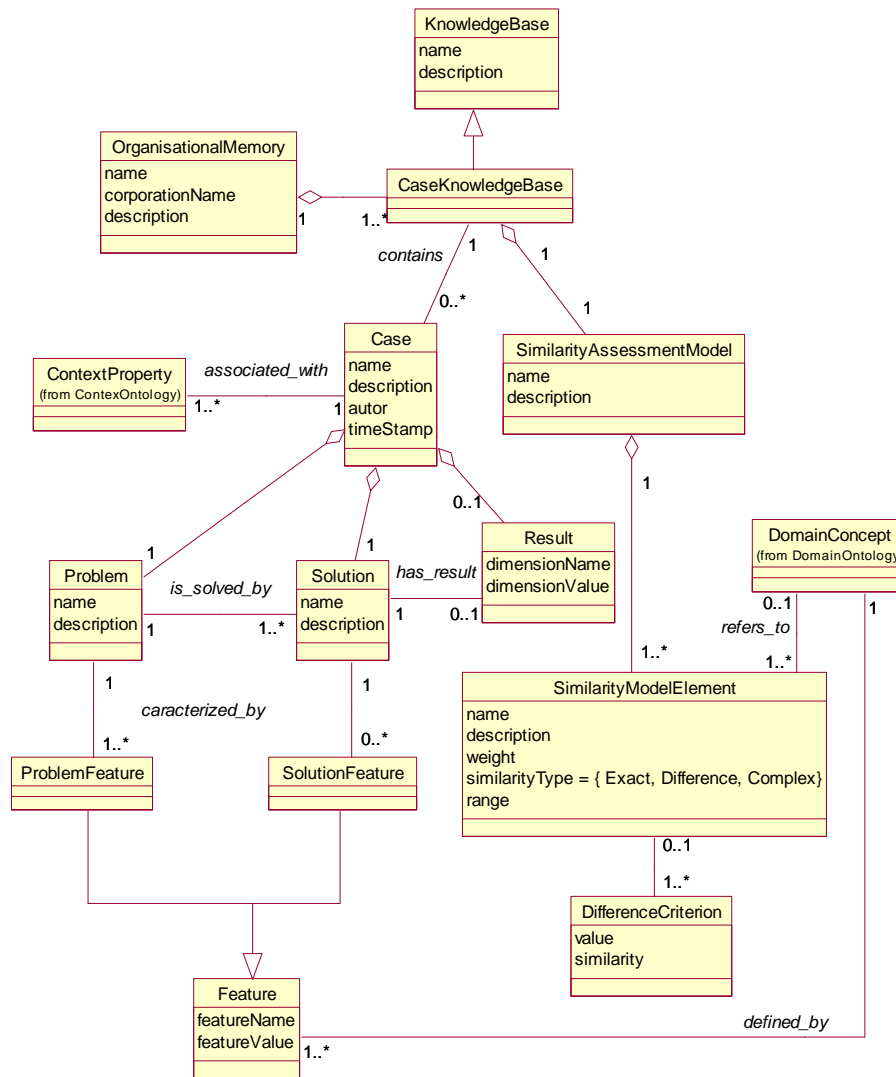


Figura 5. Modelo conceptual de la ontología de memoria organizacional basada en casos

Una *memoria organizacional* basada en casos, es un repositorio que almacena el conocimiento adquirido en experiencias pasadas como son lecciones aprendidas, buenas prácticas, heurísticas, etc., es decir, almacena *casos*. Para una mejor organización y búsqueda de dichas experiencias, la memoria organizacional se compone de varias *bases de conocimientos basada en casos (CaseKnowledgeBase)*, que agrupan los casos por conocimientos de distintas áreas.

Un caso es una pieza *contextualizada* de conocimiento que representa una experiencia, por lo que es fundamental en toda memoria organizacional guarda la información del contexto donde ocurre cada caso. Por lo tanto, a cada caso se le asocian

las propiedades de contexto (*Context Properties*) correspondientes al dominio de aplicación, definidas en la ontología de contexto.

La representación del conocimiento a través de casos, facilita el reúso del conocimiento adquirido en situaciones de problemas similares pasados para ser aplicado a un nuevo problema[16]. En una definición formal un caso es un par ordenado <P, S>, donde: P es el espacio del *problema*; y S es el espacio de la *solución*.

Los problemas y las soluciones se describen a través de variables de características del problema (*ProblemFeature*) y variables de características de la solución (*Solution Feature*) respectivamente. El proceso de razonamiento basado en casos consiste en

asignar valores a las variables de características del problema, y encontrar los valores adecuados para las instancias de la solución, a través de criterios de evaluación de similitud de casos (en la siguiente subsección se muestra una representación para calcular la similitud de casos). Por lo tanto, para cada tipo de conocimiento se debe especificar un modelo de similitud (*SimilarityAssessmentModel*).

Para que una memoria organizacional pueda ser implementada en la Web semántica, y pueda ser procesada automáticamente, necesita tener asociada una ontología de dominio[14], la cual proporciona la terminología (*Domain Concept*) que provee los tipos de las variables que caracterizan al problema y a la solución.

3.2 Representación del Modelo de Similitud

Para que un sistema CBR (*Case-Based Reasoning*) sea útil a una organización, debería ajustarse a las principales fuentes de conocimiento de la empresa, y por lo tanto necesitan funciones de similitud apropiadas a cada base de casos[12]. En esta sección se propone un modelo que permite definir la estructura de un caso indicando las *features* que lo caracterizan y su modelo de similitud.

Por ejemplo, para el dominio de medicina, una base de casos podría guardar conocimiento relacionado a diagnósticos de enfermedades y otra al tratamiento realizado, que sirva como base para recomendar tratamiento a nuevos pacientes con síntomas similares. La forma en que se caracterizan y se evalúan la similitud de los casos de diagnóstico, es completamente distinta a como se lo hace para un tratamiento, siendo necesario, por lo tanto, definir la estructura del caso y el modelo de similitud apropiado a cada base de casos.

Como se observa en el modelo de la Figura 5, a cada base de conocimiento basado en casos se le asocia un modelo de similitud (*SimilarityAssessmentModel*), que

se compone de varios elementos de similitud (*SimilarityModelElement*), uno para cada característica constituyente del caso. De este modo, el modelo conceptual expuesto en la Figura 5, define la estructura de la memoria organizacional a partir de la cual se entrenarán los clasificadores del EIPFDcMM (Organizational Memory en Figura 2), como así también se procederá a retroalimentar la estrategia mediante las decisiones generadas por el tomador de decisiones del mismo.

3.2.1 Medida de Similitud.

Tradicionalmente, la similitud entre un caso recuperado R y un nuevo caso C , se define como la suma de las similitudes entre los valores de sus características constituyentes multiplicados por sus pesos de relevancia relativa:

$$Similitud(R, C) = \sum_{f \in F} w_f \cdot sim_f(f_R, f_C)$$

En donde w_f es el peso de relevancia de la característica f , y sim_f es la función de medida de similitud de una característica específica f , perteneciente al conjunto F de todas las características disponibles.

Por lo tanto, para proveer una representación adecuada de la similitud, es necesario representar tanto los pesos de relevancia como la descripción de la función de similitud para una característica específica. Los pesos se representan como un atributo dentro de cada elemento de similitud, y la función de similitud se restringe a tres tipos generales de funciones de similitud: *Exact*, *Difference* y *Complex* [16,17].

- La función de similitud *Exact*, devuelve 1 si los valores de característica son iguales, y 0 en otro caso.
- La función de similitud *Difference*, es inversamente proporcional a la diferencia entre los valores de las características. Esta función solamente se puede aplicar cuando es posible definir la diferencia entre los valores.

- La función de similitud *Complex*, resuelve la similitud para todas aquellas situaciones donde las dos funciones de similitud anteriores no son aplicables. En nuestro modelo, estos parámetros están representados en la clase *DifferenceCriterion*.

4. Un Caso Práctico: Sistema de Recomendación de diagnóstico médico.

A continuación se ilustra la aplicación de la MOBC en el diseño de un sistema de monitoreo de pacientes y diagnóstico médico. De este modo, el proyecto de M&E se define en términos del monitoreo de pacientes (C-INCAMI DB en Figura 2), a los efectos de que el EIPFDcMM informe continuamente sus medidas y en base a la memoria organizacional, sea posible recomendar posibles diagnósticos asociados.

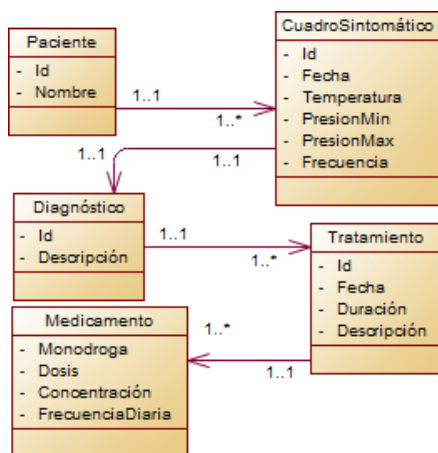


Figura 6. Modelo conceptual simplificado del dominio de diagnóstico

Para diseñar el sistema de recomendación basado en la MOBC, se debe personalizar la Memoria Organizacional, definiendo la ontología de dominio de acuerdo al tipo de conocimiento que se desea almacenar (Ver Figura 6). Dicha ontología de dominio provee los conceptos (atributos) que caracterizarán el caso (o pieza de conocimiento), representados en la Figura 5 por la clase *DomainConcept*.

En el presente ejemplo el sistema deberá administrar conocimiento relacionado al "diagnóstico a partir de la evaluación de síntomas". Dicha ontología facilita la

representación del conocimiento sobre problemas de diagnóstico en forma de casos, su almacenamiento y posterior reúso en el diagnóstico de nuevos pacientes, basado en experiencias previas similares.

Un problema (posible enfermedad) se puede caracterizar por sus síntomas. La Figura 6 presenta un modelo conceptual UML de una posible ontología de dominio, para representar Diagnósticos. A los fines de simplificar nuestro ejemplo, y para este estudio, no se desarrolló la ontología completa. Consideramos que la temperatura corporal, la presión arterial sistólica (máxima), la presión arterial diastólica (mínima) y la frecuencia cardíaca representan los atributos relevantes que caracterizan al problema en el paciente, pero esta lista debe ser completada y adaptada a cada institución de salud, de acuerdo a sus necesidades. Cada cuadro sintomático, puede tener asociado un diagnóstico, y cada diagnóstico, puede tener distintos tratamientos. Asociando esta ontología de dominio al marco MOBC presentado en la sección 3, es posible representar las "lecciones aprendidas" (o casos de diagnóstico), con la estructura problema-solución, definida en dicha sección, representando cada atributo de clase una instancia de la clase *DomainConcept* (Ver Figura 5).

En la

El almacenamiento y tratamiento de los diagnósticos mediante una memoria organizacional basada en casos, permite su procesamiento automático, y facilita la difusión y reúso del conocimiento, como por ejemplo, en el entrenamiento del clasificador del EIPFDcMM.

Tabla 1 se muestra un ejemplo de un caso que representa el diagnóstico "Gripe", y su respectiva solución (o acción correctiva) basado en el modelo conceptual de la Figura 6.

El almacenamiento y tratamiento de los diagnósticos mediante una memoria organizacional basada en casos, permite su procesamiento automático, y facilita la difusión y reúso del conocimiento, como

por ejemplo, en el entrenamiento del clasificador del EIPFDcMM.

Tabla 1. Ejemplo de un caso de diagnóstico almacenado en la MOBC

Caso 1: Gripe
Problema:
Cuadro.id=21584, Cuadro.fecha= 22/07/2013, Cuadro.temperatura= 38.9, Cuadro.PresionMin= 8, Cuadro.PresionMax= 12, Cuadro.frecuencia=95”
Solución:
Medicamento.monodroga= Amoxicilina, Medicamento.concentracion= 500mg, Medicamento.dosis= 1 comprimido, Medicamento.frecuenciadiaria= 4 veces

Además, la memoria organizacional deberá soportar funcionalidades de apoyo a todas las actividades que comprende un proceso de administración del conocimiento, entre las que se encuentran:

- **Captura y creación:** La captura y creación del conocimiento puede realizarse en forma manual, a través del ingreso de diagnósticos informados por el médico, o bien, en forma automática mediante el monitoreo y procesamiento de flujo de datos mediante EIPFDcMM. En este último caso, ante cada alarma emitida desde EIPFDcMM, se puede catalogar la misma a partir del CBR de la MO bajo un cuadro sintomático, asociándole un diagnóstico y tratamiento.
- **Búsqueda y acceso:** La utilidad de la memoria organizacional basada en casos reside en que el conocimiento almacenado puede ser reusado en la solución de problemas de diagnósticos similares. Para fortalecer este reuso, la memoria provee un motor de razonamiento basado en casos, que recupera soluciones, teniendo en cuenta casos similares, y el contexto donde se aplica.
- **Recomendación:** a través del razonamiento basado en casos, el sistema recomendará tratamientos. Así,

cuando EIPFDcMM informa una nueva experiencia (Ver Figura 2), se utiliza el CBR para localizar un diagnóstico y tratamiento desde la MO, que puedan complementar la notificación a la entidad bajo monitoreo.

Para ilustrar como se puede entrenar el clasificador del EIPFDcMM ante el arribo de nuevas medidas provenientes del monitoreo, se muestra un ejemplo en el que la MO almacena entre sus datos dos casos: el caso Gripe (Ver Tabla 1) y el caso hipertensión, (Ver Tabla 2). Ante el arribo de nuevos datos el clasificador generará un nuevo caso (datos + metadatos + decisión mediante C-INCAMI/MIS) e intentará establecer (si corresponde) una alarma interactuando con la MO a través de la función de recomendación.

Tabla 2. Ejemplo 2 de un caso de diagnóstico almacenado en la MOBC

Caso 2: Hipertensión
Problema:
Cuadro.id=21638, Cuadro.fecha= 20/09/2013, Cuadro.temperatura= 35.7, Cuadro.PresionMin= 13, Cuadro.PresionMax= 17, Cuadro.frecuencia=103”
Solución:
Medicamento.monodroga= Carvedilol, Medicamento.concentracion= 12,5, Medicamento.dosis= 1 comprimido, Medicamento.frecuenciadiaria= 1 vez

Las medidas informadas desde el EIPFDcMM a la MO, sugieren una temperatura de 36.9, una presión arterial diastólica de 11, una presión arterial sistólica de 20 y una frecuencia cardíaca de 98. El EIPFDcMM indica la decisión debido a que ha detectado un incremento progresivo en la presión, pero desconoce cómo proceder en términos de tratamiento médico, y allí es donde recurre a la memoria organizacional. El motor de CBR de la MO buscará un caso similar. Para nuestro ejemplo aplicamos la función similitud *Difference*, que valora la similitud igual a 1 si ambas características son iguales y $1/|f_c - f_n|$ en otro caso; además

aplicamos igual peso (0,25) a cada una de las cuatro características. Para calcular la similitud global de cada caso pre-existente con respecto al nuevo caso (NC), se aplica la fórmula de la sección 3.2.1, reemplazando las ponderaciones y los valores de similitud para cada característica (Ver Tablas 1 y 2), como se expone a continuación:

$$\begin{aligned} \text{Similitud}(\text{Caso1}, \text{NC}) = & 0,25 \cdot \frac{1}{|38,9-36,9|} + 0,25 \cdot \frac{1}{|8-11|} + \\ & 0,25 \cdot \frac{1}{|12-20|} + 0,25 \cdot \frac{1}{|95-98|} = \mathbf{0,32} \end{aligned}$$

$$\begin{aligned} \text{Similitud}(\text{Caso2}, \text{NC}) = & 0,25 \cdot 0,83 + \\ & 0,25 \cdot 0,5 + 0,25 \cdot 0,33 + 0,25 \cdot 0,2 = \mathbf{0,47} \end{aligned}$$

De este modo, el caso tiene mayor similitud a un caso de hipertensión, lo que permitirá disparar en forma conjunta con la alarma, el tratamiento recomendado (Carvedilol concentración 12,5, 1 comprimido por día) para la situación detectada

5. Trabajos Relacionados

Existen diversos trabajos asociados con el procesamiento de flujos de datos[18,1,19], al igual que la utilización de algoritmos de minería de datos incrementales, es decir, aquellos que se actualizan ante el arribo de cada dato[20,21]. No obstante y a diferencia de tales propuestas, nuestra estrategia se sustenta en un marco de medición y evaluación con base ontológica, lo que permite incrementar la consistencia en el procesamiento y conseguir la comparabilidad de las medidas en el tiempo. Adicionalmente, nuestra propuesta incorpora una base de conocimiento con razonamiento basado en casos, lo que permite retroalimentar la estrategia de procesamiento, incorporando las lecciones aprendidas y capitalizando tal conocimiento.

Por otro lado y en cuanto al área específica de la Gestión del Conocimiento, se han desarrollado numerosas propuestas como por ejemplo [6,22,23]. El inconveniente es que la mayoría de ellas, capturan y almacenan el conocimiento en repositorios

con estrategias no estructuradas o semi-estructuradas. Esto último, adolece de mecanismos potentes de procesamiento semántico y automático de dicho conocimiento.

6. Conclusiones y Trabajo Futuro

En el presente artículo, se han discutido las bases sobre memoria organizacional con razonamiento basado en casos, como estrategia para el reúso del conocimiento estructurado como <clave, valor>, en el entrenamiento de los clasificadores del EIPFDcMM, como así también como estrategia para complementar las alarmas con cursos de acción basados en experiencia previa. Esto último constituye un avance importante, por cuanto el EIPFDcMM simplemente partía de un conjunto de entrenamiento relacional, sin capacidades de recomendación con respecto a la experiencia previa. Dado que un clasificador es un método supervisado que requiere de entrenamiento para mejorar su efectividad, el hecho de establecer las bases para una estrategia integrada capaz de capitalizar las experiencias aprendidas, estructuradas y cuantificadas representa otro importante avance con lo planteado originalmente en EIPFDcMM.

Por otro lado, también se ha discutido una estrategia de integración de la EIPFDcMM con respecto a MOBC, pero desde la perspectiva del productor de conocimiento, lo cual es un avance con respecto a [15], ya que no se contemplaban mecanismos automáticos de retroalimentación.

De este modo, este trabajo ha avanzado sobre dos estrategias diferenciadas y complementarias entre el EIPFDcMM y la MOBC, a saber: a) la MOBC como proveedor de experiencias aprendidas y estructuradas, lo cual permite mejorar la efectividad del clasificador en el procesamiento en línea, como así también sugerir cursos de acción que complementen las alarmas, y b) el EIPFDcMM como proveedor de nuevas experiencias, derivadas de la aplicación del conocimiento al procesamiento en línea, posibilitando la

retroalimentación de la MOBC. Adicionalmente, para ilustrar esta situación, se ha planteado una prueba de concepto. Como trabajo a futuro, se profundizarán las bases de ambas estrategias, a los efectos de lograr un modelo integrado de procesamiento, con procesos adecuados y formalizados.

Referencias

1. Chakravarthy, S., Jiang, Q. (2009) "Stream Data Processing: A Quality of Service Perspective". Springer.
2. Diván, M. (2011) "Enfoque Integrado de Procesamiento de Flujos de Datos centrado en Metadatos de Mediciones". Tesis Doctoral Universidad Nacional de La Plata, La Plata, Buenos Aires (Argentina).
3. Olsina, L., Papa, F., Molina, H. (2007) "How to Measure and Evaluate Web Applications in a Consistent Way". In Rossi, G., Pastor, O., Schwabe, D., Olsina, L., eds. Ch. 13 in *Web Engineering*. Springer. pp. 385-420.
4. Molina, H., Olsina, L. (2007) "Towards the Support of Contextual Information to a Measurement and Evaluation Framework". In *International Conference on the Quality of Information and Communications Technology (QUATIC)*. Lisboa (Portugal), pp.154-163.
5. Dogson, M. (1993) "Organizational Learning: A Review of Some Literatures". Pp. 375-394.
6. Conklin, J. (1996) "Designing Organizational Memory: Preserving Intellectual Assets in a Knowledge Economy". <http://www.gdss.com/DOM.htm>, Group Decision Support Systems.
7. Diván, M., Olsina, L. (2013) "Vista de Proceso del Enfoque Integrado de Procesamiento de Flujos de Datos centrado en Metadatos de Mediciones". In *Argentine Symposium on Software Engineering. 42 Jornadas Argentinas de Informática (JAIIO)*, Córdoba (Argentina).
8. Diván, M., Olsina, L., Gordillo, S. (2011) "Strategy for Data Stream Processing based on Measurement Metadata: An Outpatient Monitoring Scenario". *Journal of Software Engineering and Applications* 2011(4), pp. 653-665.
9. Rundensteiner, W., Mani, M., Wei, M. (2008) "Utility-driven Load Shedding for XML Stream Processing". In *International World Wide Web (WWW) Conference*, Beijing (China), pp.855-864.
10. Duin, R., Tortorella, F., Marrocco, C. (2008) "Maximizing the area under the ROC curve by pairwise feature combination". *Journal of Pattern Recognition* 41(6), pp. 1961-1974.
11. Object Management Group (2008) "SPEM: Software Process Engineering Meta-Model Specification", Object Management Group (OMG).
12. Kolodner, J. (1993) "Case-based Reasoning". Morgan Kaufmann.
13. Chen, H., Wu, Z. (2003) "On Case-Based Knowledge Sharing in Semantic Web". In *XV International Conference on Tools with Artificial Intelligence*, California, pp.200-207.
14. Martín, M., Olsina, L. (2009) "Added Value of Ontologies for Modeling an Organizational Memory". In Girard, J., (ed.) *Building Organizational Memories: Will You Know What You Knew?* IGI Global, USA. pp.127-147.
15. Martín, M. (2010) "Memoria Organizacional Basada en Ontologías y Casos para un Sistema de Recomendación en Aseguramiento de la Calidad". PhD Thesis, Facultad de Informática, Universidad Nacional de La Plata, La Plata, Buenos Aires (Argentina).
16. Aamodt, A., Plaza, E. (1994) "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches". *Artificial Intelligence Communications* 7(1), pp.39-59.
17. Coyle, L., Doyle, D., Cunningham, P. (2004) "Similarity for Case-Based Reasoning". Technical Report TCD-DS-2004-25, Trinity College, Dublin.
18. Tatbul, N. (2003) "Load Shedding in a Data Stream Manager". In *VLDB*, Berlin (Germany).
19. Namit, J. (2008) "Towards a Streaming SQL Standard". *Proceedings of the VLDB Endowment* 1(2), pp. 1379-1390.
20. Bifet, A., Holmes, G., Pfahringer, B., Kirkby, R., Gavaldà, R. (2009) "New Ensemble Methods For Evolving Data Streams". In *ACM Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD). International Conference on Knowledge Discovery and Data Mining*, Paris (France), pp.139-148.
21. Bifet, A., Holmes, G., Kirkby, R., Pfahringer, B. (2010) "MOA: Massive Online Analysis". *Journal of Machine Learning Research* XI, pp. 1601-1604.
22. Natali, A., Falbo, R. (2002) "Knowledge Management in Software Engineering Environments". In *16th Brazilian Symposium on Software Engineering*, Gramado - RS (Brasil), pp.238-253.
23. Ale, M., Gerarduzzi, C., Chiotti, O., Galli, M. (2007) "Onto-Dom: A Question-Answering Ontology-Based Strategy For Heterogeneous Knowledge Sources". In *VI Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento*, Lima (Perú), pp.79-86.