

Uso de un modelo dual en una historia clínica electrónica federada

J.A. Maldonado Segura, M. Robles Viejo, C. Ángulo Fernández, D. Moner Cano, P. Crespo Molina.

Área de Informática Médica del Grupo BET, Universidad Politécnica de Valencia, Valencia España,
jamaldo@upvnet.upv.es

Resumen

Un de los principales retos del uso de las tecnologías de la información y las comunicaciones en sanidad se encuentra en facilitar a cualquier profesional sanitario involucrado en la atención sanitaria de un paciente la información clínica que necesita para realizar su trabajo. La normalización de la arquitectura de la historia clínica es crucial si los datos clínicos son compartidos entre varios profesionales o instituciones. Por ello, que dentro de los trabajos de la red de investigación cooperativa de nuevos servicios de salud basados en telemedicina se ha considerado importante que el servicio de historia clínica electrónica de la plataforma tecnológica desarrollada sea compatible con la norma europea EN13606 del Comité Europeo de Normalización (CEN/TC251). Dentro de este servicio se ha considerado el desarrollo de un servicio que permita a los repositorios de datos no normalizados servir sus datos en formatos compatibles con la norma EN13606. Este artículo pretende dar una visión de la formalización empleada para la construcción de la vista global e integrada de la información clínica distribuida en el sistema informático PANGEA desarrollado por el Grupo de Informática Médica de la Universidad Politécnica de Valencia para este propósito.

1. Motivación

Una historia clínica electrónica federada (HCEF) es una historia clínica electrónica virtual. Por virtual se entiende que no se encuentra almacenada en un sistema de base de datos sino que se construye al vuelo y bajo demanda a partir de información distribuida en varios sistemas informáticos, posiblemente heterogéneos entre sí, pertenecientes a una o varias organizaciones. La HCEF debe conservar la autonomía de las fuentes de datos. Cada fuente de datos puede permitir el acceso a toda o parte de la información que contiene, son los responsables de los datos los que controlan qué información se comparte con el resto de sistemas que forman la federación.

Uno de los aspectos clave en cualquier sistema federado es la construcción de la vista global (esquema federado) de la información ofrecida por el sistema. En un contexto de HCE, este esquema federado debe ser lo suficientemente genérico y potente para representar adecuadamente cualquier información que se pueda extraer de las fuentes de datos. De todo esto se desprende que para el desarrollo de un sistema de HCEF es crucial el uso de un modelo de información consensuado para la representación de la información clínica, es decir, una

arquitectura de información para la comunicación de la historia clínica (AHCE).

En el ámbito de las AHCE actualmente existe una tendencia hacia el uso de modelos duales como CEN/TC251 EN13606[1] u OpenEHR[2]. El uso de un modelo dual de AHCE en un sistema de historia clínica electrónica federada implica que en el esquema federado se debe plasmar la a separación entre datos y conocimiento. Por un lado, los extractos de HCE que gestiona el sistema deben ser instancias del modelo de referencia. Por otro, el esquema federado debe estar compuesto por un conjunto de conceptos clínicos definidos como arquetipos, esto es, como restricciones sobre el modelo referencia.

Los esquemas de las fuentes de datos que componen la federación describen la información clínica que se debe utilizar para generar las instancias de los arquetipos. Es necesario, por tanto, definir algún tipo de correspondencias entre los arquetipos y los esquemas de las fuentes de datos que describan cómo utilizar la información contenida en las fuentes de datos para crear las instancias de los arquetipos. El objetivo de este artículo es describir brevemente el modelo de datos utilizado para la representación de cualquier extracto de HCE compatible con la norma europea EN13606 y por otro desarrollar un sistema de tipos con la misma capacidad expresiva que el cADL del lenguaje de descripción de arquetipos (ADL) [3] que sirva de “puente” entre los esquemas y los arquetipos. Esto supone una base para desarrollar sistemas de historias clínicas federadas donde las entidades constitutivas del esquema federado sean los arquetipos con todas las ventajas que esto puede suponer para alcanzar una alta interoperabilidad semántica.

2. Modelo de datos para HCE

Un modelo de datos es un conjunto de conceptos que permite describir la estructura de una base de datos. Un esquema federado, al igual que el esquema de una base de datos tradicional, se debe expresar por medio de un modelo de datos que recibe el nombre de modelo de datos común o canónico (MDC). Los esquemas de las fuentes de datos participantes son traducidos al MDC, lo cual permite la representación de los datos en un formato común y facilita la integración de los esquemas. En

nuestro caso, el MDC elegido debe ser adecuado para representar instancias (extractos de HCE EN13606) y que los esquemas utilizados para el MDC deben ser capaces de describir la parte estructural (restricciones) de los arquetipos EN13606.

A continuación se describe el modelo de datos propuesto para la representación de extractos de historias clínicas electrónicas compatibles con la norma europea EN13606. Aunque el modelo propuesto está pensado para su uso con EN13606 gracias a su flexibilidad puede utilizarse para el modelado de datos provenientes de otros modelos de referencia. El modelo de datos se basa en árboles etiquetados con referencias donde los datos están asociados a los nodos. Es parecido a otros modelos de datos semiestructurados como [4] [5]. En concreto, es muy similar al presentado en [4], aunque al contrario que en éste, se contempla la posibilidad de que los nodos no estén ordenados.

Para su formalización asumimos la existencia de los siguientes conjuntos: un conjunto infinito de identificadores de objetos (nodos) O , un alfabeto de etiquetas Σ que contiene las etiquetas que pueden estar asociadas a los nodos del árbol y un conjunto finito de tipos primitivos b_1, \dots, b_n , que pueden tener dominios infinitos. Denotamos por $\text{dom}(b_i)$ el dominio de b_i y por Dom el dominio formado por la unión de todos los dominios de los tipos primitivos. Las referencias se modelan como un tipo especial de nodo, que está etiquetado con un símbolo especial “&”, que pertenece a Σ y que tiene exactamente un hijo.

La relación de estos conjuntos con el modelo de referencia es directa. Los identificadores de objetos se utilizan para identificar unívocamente a los objetos (nodos) que forman el árbol. Como los extractos de datos a representar son instancias del modelo de referencia los árboles de datos deben utilizar los nombres de las entidades que aparecen en él. Es por esto que el alfabeto Σ está compuesto por los nombre de atributos, relaciones y clases del modelo de referencia. Los nodos internos representan atributos tanto simples como complejos y cuyas etiquetas provienen de Σ , mientras que las hojas contienen los datos y por tanto están etiquetadas con valores pertenecientes a los tipos primitivos. Existe un nodo especial denominado nodo raíz y denotado por Δ , el cual constituye la raíz del árbol de datos, además no tiene etiqueta y no puede ser referenciado por otro nodo. Por ser la raíz del árbol, desde él debe existir un camino a cualquier otro nodo.

El modelo propuesto permite que los nodos hijos de cualquier nodo puedan estar o no ordenados. El orden es necesario para poder representar estructuras como listas o documentos donde el orden los elementos es relevante. El modelo propuesto es compatible tanto con XML como con dADL (sintaxis para descripción de datos del ADL), ya que subyacente a ambos se encuentra un modelo de datos basado en árboles etiquetados.

Definición 1: Árbol de datos con referencias. Un árbol de datos con referencias (en adelante árbol de datos) es una estructura $A = (\Delta, O_A, \lambda_A, S_A)$, donde:

1. Δ es el nodo raíz del árbol de datos
2. $O_A \subset O$
3. $\lambda_A : O_A \rightarrow \Sigma \cup Dom$ es una función de etiquetado que asigna un valor perteneciente a Σ a los nodos interiores del árbol y un valor proveniente de $\Sigma \cup Dom$ a los nodos hoja.
4. S_A es una función que asigna a todos los nodos (incluyendo el nodo raíz) una secuencia de identificadores que puede estar ordenada o no. Si $\lambda_A(o) = \&$ entonces $S_A(o) \in O_A$. Si la secuencia está ordenada la representamos encerrada entre los símbolos “<” y “>”, mientras que si no está ordenada la encerramos entre llaves “{“ y “}”.
5. La estructura resultante al considerar únicamente los hijos de nodos no etiquetados con & es un árbol.

Una instancia del modelo de referencia (conjunto de objetos) se describe por una secuencia no ordenada de árboles (si no contamos las referencias) que tienen como padre el nodo raíz (Δ). Cada árbol representa un objeto, el nodo de más alto nivel está etiquetado con el nombre de la clase (tipo). La estructura del árbol refleja la estructura del objeto. Resumiendo, las etiquetas interiores describen atributos, tanto simples (sólo tienen un hijo) como complejos (tienen más de un hijo) mientras que los nodos hojas están etiquetados con datos.

3. Formalización de arquetipos

3.1. Arquetipos: tipos sobre datos semiestructurados

La solución propuesta para la especificación del esquema global se basa en el desarrollo de un sistema de tipos para el modelo de datos descrito anteriormente. Se propone la utilización de un metalenguaje para la definición de conjuntos de árboles de datos (formalización de los extractos de HCE compatibles con EN13606) con características comunes en cuanto a su etiquetado, estructura y significado clínico. Denominaremos a cada uno de estos conjuntos un tipo de árbol de datos. Cabe destacar que este sistema de tipos está centrado exclusivamente en el modelado de la parte estructural de los arquetipos, es decir, el conjunto de restricciones que definen estructuras de datos válidas basadas en el modelo de referencia. El sistema de tipos también permite la definición del esquema federado. En efecto, el esquema federado no es más que un conjunto de tipos.

Para su formalización asumimos la existencia de los siguientes conjuntos: un conjunto finito de tipos primitivos denotado por C , un conjunto infinito de variables de tipos T , donde T y C son disjuntos y un conjunto P de predicados unarios de etiquetas.

Denominaremos a la unión de los tipos primitivos y variables de tipos como nombres de tipos.

Definición 2. Definición de tipo. Una definición de tipo es una regla que puede tomar una de las dos siguientes formas:

1. Tipo $t := l \langle r \rangle$
2. Tipo $t := l \{ r \}$

Donde:

- $t \in T$ y $l \in P$
- r es una expresión regular sobre $T \cup C$
- $l = \{ \& \}$ ó $\& \notin l$
- si $l = \{ \& \}$ entonces $r = t_1 \dots t_m, \forall j \in \{ 1..m \} t_j \in T$

Se permite el uso de dos clases de reglas para definir tipos. Las reglas que siguen la estructura $t := l \langle r \rangle$, y las que siguen la estructura $t := l \{ r \}$. Las primeras describen un conjunto t de árboles de datos donde el nodo raíz está etiquetado con una etiqueta que satisface l , los hijos del nodo raíz están ordenados y r describe la secuencia permitida. Las segundas describen un conjunto t de árboles de datos, donde el nodo raíz de todos ellos está etiquetado con una etiqueta que satisface l , los hijos del nodo raíz no están ordenados y alguna secuencia de éstos satisface r .

Definición 3. Esquema. Un esquema E es una tupla $(N \cup \Delta, D)$ donde:

- $N \subset T$ es un conjunto finito de nombres de tipos.
- Δ es el nombre de un tipo especial que denominaremos tipo raíz.
- D es un conjunto de definiciones de tipos que contiene precisamente una regla por cada uno de los nombres de tipos en $N \cup \Delta$, además, en estas reglas se utilizan únicamente nombres de tipos de N .

La representación de un esquema orientado a objetos, y por ende, de un modelo de referencia es también sencilla. Así por ejemplo, cada clase no abstracta del modelo de referencia se representa por un tipo. Los tipos de datos atómicos del modelo de referencia constituyen el conjunto de tipos primitivos. Por otro lado, las variables de tipos servirán para nombrar el resto de clases del modelo de referencia. Cada uno de estos tipos tiene un predicado de etiqueta que acepta únicamente el nombre de la clase. Los atributos se representan con un tipo cuyo predicado de etiqueta únicamente es satisfecho por el nombre del atributo. La Figura 1 muestra la representación de la clase *Folder* del modelo de referencia de EN13606.

Falta dotar de una semántica a los esquemas, es decir, definir el conjunto de árboles de datos (extractos de HCE) que denotan. Intuitivamente un árbol de datos A es conforme a un esquema E si es posible asignar a cada uno de los nodos un tipo T definido en E . Esto se define de manera formal a continuación.

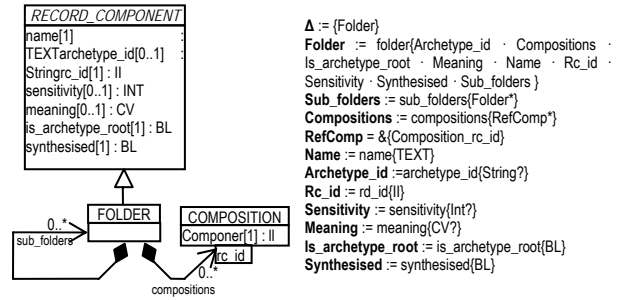


Figura 1. Representación de la clase Folder de EN13606

Definición 4. Sea $E=(N, D)$ un esquema y A un árbol de datos. Decimos que A es conforme a E bajo la asignación de tipo F y lo denotamos por $A :_F E$, si y solo si existe una función de asignación de tipo F para todos su nodos que tiene la forma $F: O_A \cup \Delta \rightarrow N \cup \Delta$ y que cumple lo siguiente:

1. $F(\Delta) = \Delta$
2. $\forall O_i \in O_A, \text{Etiqueta}_E(F(O_i)) \models \lambda(O_i)$
3. $\forall O_i \in O_A \cup \Delta$ Si O_i es ordenado y s_1, \dots, s_n son sus hijos entonces $F(O_i)$ es el nombre de un tipo ordenado y $F(s_1), \dots, F(s_n) \in L(\text{Expreg}_E(F(O_i)))$
4. $\forall O_i \in O_A \cup \Delta$ Si O_i no es ordenado y s_1, \dots, s_n son sus hijos entonces $F(O_i)$ es el nombre de un tipo no ordenado y existe una permutación P de $F(s_1), \dots, F(s_n)$ tal que $P \in L(\text{Expreg}_E(F(O_i)))$

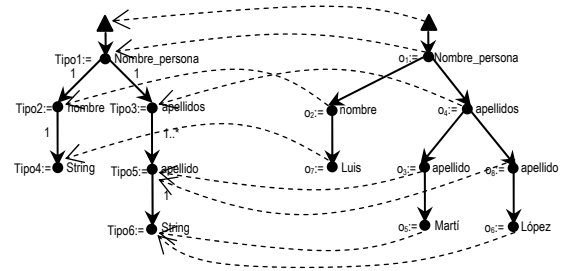


Figura 2. Ejemplo de función de asignación de tipos

La primera condición obliga a que el nodo raíz del árbol de datos sea una instancia del tipo raíz. La segunda condición indica que la etiqueta de cualquier nodo del árbol de datos, excepto el nodo raíz, debe satisfacer el predicado de etiqueta del tipo asignado. La tercera condición asegura que los tipos y orden de los hijos de un nodo ordenado del árbol de datos son válidos con respecto al orden y tipos descritos por la expresión regular de su tipo. La última condición es similar a la anterior pero se refiere a los nodos no ordenados y por tanto la condición anterior se relaja a que exista al menos una permutación de los hijos del nodo del árbol de datos que sea válida con respecto a la expresión regular del tipo padre. La Figura 2 muestra un ejemplo de función de asignación de tipos.

3.2. Especialización de arquetipos

De acuerdo a la última versión de ADL [3] los arquetipos se pueden especializar. La regla principal que controla la especialización es que las instancias de datos creadas de

acuerdo al arquetipo hijo (más especializado) son también instancias del arquetipo padre. Se propone como mecanismo para modelar la especialización de arquetipos el concepto de subsumición de tipos. La subsumición consiste en establecer un conjunto de relaciones estructurales entre los tipos de forma que la existencia de estas relaciones asegura que todas las instancias del tipo subsumido (más específico) son instancias del tipo que subsume (más general).

Definición 5. Subsumición de esquemas. Sean $E=(T_E, D_E)$ y $E'=(T_{E'}, D_{E'})$ dos esquemas. Decimos que el E' subsume a E bajo la función de subsumición θ , si y solo si θ es una función $\theta: T_E \cup \{\Delta\} \rightarrow T_{E'} \cup \{\Delta\}$ que cumple las siguientes condiciones:

1. $\theta(r) = \Delta \leftrightarrow r = \Delta$
2. $\forall r \in T_E \text{ etiqueta}_E(r) \subseteq \text{etiqueta}_{E'}(\theta(r))$
3. $\forall r \in T_E \cup \Delta$ si r es ordenado entonces $\theta(r)$ también lo es y $L(h_\theta(\text{expreg}_E(r))) \subseteq L(\text{expreg}_{E'}(\theta(r)))$
4. $\forall r \in T_E \cup \Delta$ si r no es ordenado entonces $\theta(r)$ tampoco lo es y $\bigcup_{x \in L(h_\theta(\text{expreg}_E(r)))} \text{per}(x) \subseteq \bigcup_{y \in L(\text{expreg}_{E'}(\theta(r)))} \text{per}(y)$

Donde:

- $h_\theta: \text{ExpReg}(T_E) \rightarrow \text{ExpReg}(T_{E'})$ es la extensión de θ a las expresiones regulares.
- $\text{per}(x)$, siendo x una cadena, es el conjunto formado por todas las permutaciones posibles de x . Por ejemplo, $\text{per}(a \cdot b) = \{a \cdot b, b \cdot a\}$.

Un arquetipo se puede especializar restringiendo el dominio de las clases, relaciones y atributos que contiene, más concretamente:

1. Atributos atómicos. Un atributo atómico se puede especializar restringiendo su dominio.
2. Atributos y relaciones. Endureciendo las restricciones de existencia o cardinalidad en relación a las existentes en el arquetipo padre.
3. Referencias a otros arquetipos. Restringiendo el conjunto de arquetipos que pueden ser referenciados en relación al arquetipo padre.
4. Referencias internas. Reduciendo el conjunto de nodos que pueden ser referenciados en relación al arquetipo padre.
5. Nodos objetos. Un nodo que expresa restricciones sobre instancias de alguna clase del modelo de referencia se puede restringir endureciendo su restricción de ocurrencia en relación a la existente en el arquetipo padre o dividiendo el nodo objeto en múltiples nodos de forma que cada uno de ellos es una especialización del nodo padre y sustituimos el nodo objeto por una alternativa.

Todas estas formas de especialización de arquetipos dan lugar a que exista una relación de subsumición entre el esquema del arquetipo padre y el esquema del arquetipo hijo, la demostración se puede encontrar en [6].

4. Conclusiones

En el presente artículo se ha descrito el modelo de datos y sistema de tipos propuesto para la formalización de los extractos de HCE y la parte de definición de extractos de HCE de los arquetipos. En lo referente al modelo de datos, se ha descrito su formalización y cómo representar instancias de un modelo de referencia. En relación al sistema de tipos, se ha descrito su formalización, cómo representar un modelo de referencia y la parte estructural de un arquetipo y la semántica de la especialización entre arquetipos. La forma de modelar las restricciones contenidas en un arquetipo difiere del ADL. En ADL solo aparece aquello que se restringe en relación al modelo de referencia. En el sistema de tipos propuesto esto no es así, sino que en la formalización aparecen todas las entidades del modelo de referencia independientemente de que se restrinjan o no. Esto es necesario para poder definir las correspondencias con las fuentes de datos que contienen la información clínica necesaria para instanciar los arquetipos.

El sistema de tipos es lo suficientemente potente y flexible para representar tanto el modelo de referencia como los arquetipos. La relación de subsumición propuesta permite modelar tanto las relaciones instancia-arquetipo como las arquetipo_padre-arquetipo_hijo y arquetipo-modelo de referencia.

Agradecimientos

El trabajo presentado está financiado por la red temática de investigación de nuevos servicios de salud basados en telemedicina y por el Ministerio de Educación y Ciencia, proyecto con referencia TSI2004-06475-C02-01.

Referencias

- [1] CEN TC/251. *Health Informatics-Electronic Healthcare Record Communication*. <http://www.cen251.org>
- [2] <http://www.openehr.org>
- [3] OpenEHR. *The Archetype Description Language*. Disponible en Internet: <http://svn.openehr.org/specification/TRUNK/publishing/architecture/am/adl.pdf> (consultado Sep-2005).
- [4] Abiteboul, S., Cluet, S., Milo, T. Correspondence and translation for heterogeneous data. *Theoretical Computer Science* 275(1-2), pp. 179-213, 2002.
- [5] Kuper, G. M., Simeon, J. (2001). Subsumption for XML types. *Proceedings of the 8th International Conference on Database Theory (ICDT'01)*, pp. 331-345, 2001.
- [6] José Alberto Maldonado (2005); *Historia Clínica Electrónica Federada Basada en la Norma Europea CEN/TC251 EN13606*; Tesis Doctoral, Universidad Politécnica de Valencia.