



Representación del Conocimiento mediante Ontologías : Ensayo sobre Citología Tiroidea.

Javier ESQUIVIAS*, Jose Manuel ZURITA**, Mercedes GOMEZ***, Eduardo VILLAR****, Jose Manuel SANCHEZ**, Jose ALONSO***

* HOSPITAL CLINICO UNIVERSITARIO SAN CECILIO-GRANADA ESPAÑA

** ETS INFORMATICA-GRANADA ESPAÑA

*** Hospital Clínico San Cecilio-GRANADA ESPAÑA

**** Hospital Virgen de las Nieves-GRANADA ESPAÑA

Resumen

El conocimiento que un patólogo posee sobre un diagnóstico es difícilmente transferible. Por dos motivos: primero porque su conocimiento está basado, en gran medida en la experiencia propia y en la que no están nitidamente definidos los parámetros que identifican el diagnóstico; y en segundo lugar, la forma en que el patólogo representa su conocimiento para explicarlo a los demás es en muchos casos poco homogénea. Esta falta de homogeneidad de formas de representación suponen un claro inconveniente a la hora de utilizar esa información, para que otros patólogos interesados pudieran trabajar con los mismos estándares para poder compartir y comparar sus conocimientos.

En este trabajo proponemos representar el conocimiento utilizando ontologías. El concepto de ontología es bastante reciente, puede definirse (Borst 1997) como *una especificación explícita y formal sobre una conceptualización consensuada*. El término *explícita* se refiere a que las ontologías definen sus conceptos, propiedades, relaciones, funciones en algún lenguaje de implementación capaz de contener ese conocimiento. El término *formal* se refiere a que la ontología debe implementarse en algún lenguaje inteligible y computable por el ordenador. La *conceptualización* se refiere a una visión abstracta y simplificada de algún fenómeno del mundo real, identificando sus conceptos relevantes. *Consensuada* refleja la noción de que el conocimiento que capturan las ontologías no es conocimiento privado de un único individuo, sino que deberá aceptarse por un grupo de sistemas que deberán comprometerse con el vocabulario utilizado en la ontología.

La utilización de las ontologías permiten hacer esa información fácilmente manejable en Internet, ya que su lenguaje de representación facilita la tarea de su almacenaje y búsqueda con protocolos estándares. De ahí que este conocimiento representado mediante ontologías pueda estar disponible en la red para que otros investigadores puedan hacer uso de él mediante la *reusabilidad* del mismo. Esto quiere decir, que podemos obtener dicho conocimiento y modificarlo, ampliándolo o eliminando conceptos con lo cual poco a poco se iría construyendo una mejor aproximación a dicho conocimiento con la aportación de cada profesional. De hecho, hay multitud de ontologías médicas disponibles en la red a las que uno puede acceder mediante su dirección URL.

Pretendemos, a modo de ensayo, usando ontologías, representar el conocimiento de los patólogos respecto a la citología de tres diagnósticos relacionados con la patología tiroidea (Carcinoma papilar, Nódulo coloide y Nódulo adenomatoso) con el objetivo de dejar accesible a través de una dirección URL dicha información al resto de la comunidad científica. Somos conscientes de que este conocimiento plasmado mediante ontologías no es completo, debido a la complejidad del mismo. Por tanto, centramos fundamentalmente el interés del trabajo en el hecho de facilitar al resto de patólogos la posibilidad de acceder al conocimiento de citología de tiroides a través de Internet y en un formato homogéneo que pueda ser utilizado por quien lo desee aportando a la vez su conocimiento.

Introducción

El diagnóstico morfológico (en este caso citológico) se basa en interpretar las variaciones que muestran las células con respecto a las células consideradas normales.

La patología de la glándula tiroides es muy variada, comprendiendo lesiones neoplásicas, lesiones inflamatorias y lesiones metabólicas.

El conocimiento de dicha patología requiere adquirir conceptos claros sobre la morfología y la fisopatología de la glándula. En todo conocimiento podemos distinguir cuatro elementos: El sujeto que conoce, El objeto conocido, La operación misma de conocer y El resultado obtenido que es la **información** recabada acerca del objeto.

Dicho de otra manera: el sujeto se pone en contacto con el objeto y obtiene una **información** acerca del mismo. Cuando existe congruencia o adecuación entre el objeto y la representación interna correspondiente, decimos que estamos en posesión de una verdad.

Los mapas conceptuales constituyen una aproximación a la representación del conocimiento, que en el caso de las ontologías, además de establecer una clasificación taxonómica del conocimiento, permiten establecer definiciones propiamente dichas del concepto en sí utilizando los métodos propios de la lógica. Esto implica que la representación del conocimiento que se realice mediante ontologías será factible de una comprobación automatizada de la consistencia de los datos que se definen.

La consistencia interna de la información juega un papel fundamental en la construcción de cualquier sistema automatizado inteligente que maneje conocimiento, puesto que la garantía de un correcto funcionamiento del mismo viene necesariamente marcada por la consistencia de la información que procesa.

Hemos tomado como modelo la Citología por punción de la glándula tiroides, para llevar a cabo un ensayo sobre la forma de representar el conocimiento usando ontologías.

Objetivos:

1. Construir una ontología sobre los rasgos citológicos representativos de la citología del tiroides con el objetivo de dejar accesible a través de una dirección URL dicha información al resto de la comunidad científica.
2. Representar los rasgos diagnósticos de tres patologías determinadas (Nódulo coloide, nódulo adenomatoso y carcinoma papilar).
3. Mostrar un método de representación de conocimiento basado en ontologías que permite la reusabilidad del mismo por otros investigadores mediante la modificación y/o complementación del conocimiento propuesto en este estudio.

Material y Métodos

Se ha usado la herramienta Protégé, de libre distribución, <http://protege.stanford.edu/>.

Protégé es un editor para construir ontologías y un marco general para representar el conocimiento. Con la ayuda de este software se pueden construir aplicaciones basadas en el conocimiento que representen la información de forma ontológica. La característica de este tipo de herramientas es que utilizan un lenguaje muy flexible para ser utilizado en la Web, por eso se ha acuñado el término de Web Semántica, es decir, que facilitan la tarea no sólo de ser manejable a través de la Web, sino que hacen una descripción semántica de la información, con lo que todo el sistema es menos rígido y por tanto se hace más flexible y potente.

Protégé reconoce por tanto, Frames, XML Schema, RDF Schema y OWL, que son lenguajes semánticos utilizados en la Web, en contraposición a la rigidez del HTML. El lenguaje con más perspectiva de futuro en la representación del conocimiento usando ontologías es el OWL. Nosotros hemos utilizado en consecuencia dicho lenguaje en nuestra investigación. Protégé permite realizar programas en OWL con una gran facilidad, debido a su entorno gráfico, tan sólo con el uso de otro programa asociado a él que se denomina Protégé-OWL Plugin (Horridge y demás, 2004). Este es una especie de anexo a Protégé para poder trabajar con el lenguaje OWL. Se encuentra disponible en la dirección, <http://protege.stanford.edu/plugins/>.

Con respecto a la adquisición del conocimiento de la Citología de la glándula tiroides, se ha seguido el mismo procedimiento explicado en otra comunicación de este Congreso Virtual (Nº 212: *El diagnóstico en citología: Extracción de rasgos y cuantificación de los*

misimos. Ensayo sobre citología tiroidea).

Se han segregado los Rasgos en Componente, Calificador y Característica. Cada uno de estos elementos ocupa su lugar en el mapa conceptual, y la interrelación entre ellos determina un rasgo. La relación de los rasgos con los diagnósticos elegidos constituirá la representación del conocimiento del aspecto citológico de cada diagnóstico.

Resultados

El resultado de la aplicación queda reflejado en la ontología ubicada en la dirección **URL:**

<http://decsai.ugr.es/~jmsa/OntologiaDeTiroides.pprj>

<http://decsai.ugr.es/~jmsa/OntologiaDeTiroides.owl>

El primer fichero OntologiaDeTiroides.pprj contiene el proyecto global de la ontología que deberá cargarse en el programa Protégé. El fichero OntologiaDeTiroides.owl corresponde a la Ontología descrita en el lenguaje OWL. Ambos son necesarios para Protégé.

En la Figura 1 observamos la clasificación taxonómica de los conceptos que agrupan el conocimiento de la citología de tiroides, así como su diagnóstico.

En la Figura 2 se observa la representación gráfica que conlleva la clasificación taxonómica de los conceptos representados en la ontología. Se aprecia la relación *is-a* (es un), que indica un grado de subordinación del concepto del que parte la flecha al concepto al que llega la flecha.

En la Figura 3 se hace una descripción mediante la Lógica del diagnóstico Nódulo Adenomatoso utilizando los rasgos definidos previamente en la Ontología. La interpretación semántica de lo que se dice es la siguiente: Es necesario en la descripción del Nódulo Adenomatoso que se verifique que,

1. Es un tipo de Diagnóstico.
2. Tiene un tipo de calificador de celularidad variable.
3. Tiene un tipo de calificador de coloide escaso.
4. Tiene un tipo de característica de celularidad monocapa.
5. Tiene como característica de celularidad un núcleo normal.
6. Tiene como característica de celularidad una agrupación macrofolicular y/o microfolicular con o sin coloide en su interior.
7. No puede tener un tipo de característica de celularidad con pseudoinclusiones nucleares o con hendiduras nucleares.

Evidentemente, esta descripción para la identificación del nódulo adenomatoso puede y debe ajustarse más mediante la obtención de nuevos rasgos y la generación de fórmulas de la Lógica de Descripción, similares a las descritas aquí, que se aproximen aún más a la definición de este diagnóstico.

Discusión

Este ensayo constituye una primera aproximación al manejo de las ontologías.

La forma de organizar el conocimiento puede ser extraordinariamente variada. Se ha elegido la presente, por varias razones: Primera, porque ya se había desarrollado previamente el trabajo de extraer el conocimiento sobre este tema para la construcción de un sistema experto. En segundo lugar, porque estimamos que la expresión de los rasgos morfológicos se presta

menos a confusión que otros conceptos en medicina, aunque la imprecisión está siempre presente en el manejo de la terminología, aún en el campo de la citología diagnóstica. En tercer lugar, la utilización de ontologías permite que el conocimiento adquirido esté representado de tal forma que puede ser fácilmente accesible en cuanto a su complementación o modificación por otros investigadores en el área.

Uno de los mayores inconvenientes con que nos hemos encontrado es el de definir claramente los RASGOS. Entendiendo a estos como "conceptos" que hacen referencia a un "objeto" (diagnóstico), estamos de acuerdo con la opinión de Chantal Pérez de que un concepto lleva normalmente asignada una representación física concreta, y como constructo que es, posee ciertas propiedades distintivas de los demás conceptos, con los que guarda diversos tipos de relaciones. Tanto sus propiedades intrínsecas como sus relaciones con los demás conceptos deben ser evidentes, y por tanto susceptibles de ser especificados de forma explícita. En el caso que nos ocupa, hemos de admitir que la propuesta de RASGOS que se ha formulado es susceptible de mejora en muchos casos, ya que se presta a confusión. O al menos debería explicarse mejor, de forma consensuada, el significado. Por ejemplo, COLOIDE ABUNDANTE o CELULARIDAD ESCASA son términos subjetivos. En estos casos proponemos, basándonos en las condiciones de la Clasificación de Bethesda sobre citología Cervico-vaginal la siguiente nomenclatura:

	Celularidad	Coloide
Abundante	> 15 grupos o más de 40 células por c.g.a.	> 50% de superficie ocupada
Escasa	< 9 grupos o menos de 15 células por c.g.a.	<5% de superficie ocupada

Otro aspecto que es susceptible de discusión se refiere a la segregación de los Rasgos en Componente, Calificador y Característica. Se ha hecho así por acuerdo, para conformar la Base de Datos del Sistema Experto confeccionado por los autores. Probablemente se requiere mayor discusión y acuerdo sobre el significado de los términos empleados y aclarar de forma explícita las propiedades que las diferencien de los otros calificadores y características.

También consideramos que la clasificación taxonómica y definición de los conceptos y diagnósticos empleados es susceptible de estudios paralelos o más específicos que ayuden a profundizar aún más en el conocimiento subyacente a la citología de tiroides. Esto se puede hacer fácilmente como ya hemos comentado gracias a la flexibilidad de los lenguajes que subyacen en la generación de ontologías. Por tanto este trabajo constituye un punto de partida sobre el que ampliar la representación del conocimiento y se verá enriquecido con la aportación de quien lo desee, para lo cual cuenta con la ventaja de su ubicuidad.

Conclusiones

En este artículo se ha representado el conocimiento sobre la citología de tiroides mediante ontologías. El uso de ontologías supone un marco de trabajo en el cual otros investigadores pueden ver, modificar o cambiar el conocimiento representado; todo ello debido a la flexibilidad de los lenguajes utilizados.

En este trabajo, se ha representado el conocimiento implícito a los diagnósticos: nódulo coloideo, nódulo adenomatoideo y carcinoma papilar.

Como herramienta hemos utilizado el Protégé y el Protégé-OWL Plugin. Este último es un software que se instala junto al primero para poder trabajar con el lenguaje OWL en Protégé, que es uno de los lenguajes semánticos para la Web más potentes.

Como trabajos futuros nos planteamos el análisis de los rasgos utilizados en la citología de tiroides, por si pueden ser complementados o bien buscar otros rasgos diferentes que aproximen mejor el conocimiento. También estudiaremos con más detalle las definiciones dadas de los tres diagnósticos representados con el objeto de ser lo más exactos posibles en su descripción. Por último, ampliaremos a un mayor número de diagnósticos este estudio para hacer más completo el conocimiento sobre la citología de tiroides.

Agradecimientos

Esta investigación ha sido financiada parcialmente por el Ministerio de Educación y Ciencia a través del Proyecto de Investigación: TIN2004-07236.

Bibliografía

Borst, W.N. Construction of engineering ontologies for knowledge sharing and reuse. Doctoral Thesis. Centre for Telematics and Information Technology, Universiteit Twente.

de la Morena A. XVIII Congreso Nacional de la SEAP. Benalmádena. 1997

Horridge, M., Knublauch, H., Rector, A., Stevens, R., Wroe, C. A practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools Edition 1.0. University of Manchester. 2004.

Chantal Pérez . *Estudios de Lingüística Española (ELiEs)* Volumen 18 (2002) ISSN: 1139-8736. Depósito Legal: B -39120-2002. Copyright: © Chantal Pérez.

Pérez Barrios A. Punción aspiración con aguja fina del tiroides. XIV Curso de citopatología clínica. Madrid. Mayo 2000.

Young JA. Fine Needle Aspiration Cytopathology. Blackwell Scientific Publications. Oxford. 1993.

Bibbo M. Comprehensive Cytopathology. W.B. Saunders Co. Philadelphia. 1997

Miralles T.G. ; Buesa JM. Punción-Aspiración con aguja fina. Salvat. 1988

Oertel Y.C. : Citología por Punción –Aspiración con aguja fina (PAAF) de mama y tiroides. Curso. Madrid, 1982.

Clark D. P. Thyroid Cytopathology. ESSENTIALS IN CYTOPATHOLOGY. Foreword by Edmund S. Cibas, M.D. Series Editor Dorothy L. Rosenthal. Springer 2005