

# Capítulo 18

## Diseño y Desarrollo de Herramientas para el Aprendizaje de Relaciones Semánticas presentes en el Lenguaje Natural

Antonio Vaquero Sánchez, Francisco Alvarez Montero y Fernando Sáenz Pérez

*Title*—Design and Development of Tools for Learning Semantic Relations present in Natural Language

*Abstract*—In our previous work, we used a methodology based on software engineering principles to develop tools for building and querying electronic dictionaries based on a DAG-shaped taxonomy, with language learning purposes. However, the tools do not enforce any kind of control over the use of the single semantic relation that is used during the DAG-shaped taxonomy construction. In order to teach specific semantic relations (e.g., is-a, member-of, component-of) the tools must control their appropriate usage, preventing learners from relying in their intuition to construct the taxonomy, and thus forfending them, from making inappropriate and inconsistent modelling choices. We intend to enforce such a control by providing relations with a set of properties defining their intension. Then, using these properties as an aide, the authoring tools could assist the author in determining, if between two concepts there can be a given semantic relation, by asking meaningful questions about its intension. This paper represents a first step towards that goal.

*Keywords*—Algebraic Properties, Control, Electronic Dictionaries, Intrinsic Properties, Methodology, Semantic Relations

*Resumen*—En trabajos anteriores, usamos una metodología basada en Ingeniería del Software para el desarrollo de herramientas para la creación y consulta de diccionarios electrónicos con una taxonomía en forma de grafo acíclico, con el propósito de potenciar el aprendizaje de lenguas. Sin embargo, las herramientas no imponen ningún tipo de control sobre el uso de la única relación semántica utilizada en la construcción de la

Este trabajo fue presentado originalmente al VIII Congreso Iberoamericano de Informática Educativa, RIBIE2006, en San José, Costa Rica.

Antonio Vaquero Sánchez is Director of the Institute of Knowledge Technologies (Instituto de Tecnologías del Conocimiento), of the Complutense University of Madrid (UCM), Spain (phone: +34-913947622; fax: +34-913947529; e-mail: vaquero@ sip.ucm.es).

Francisco Alvarez Montero is with the Institute of Knowledge Technologies of the Complutense University of Madrid (UCM), Spain (e-mail: fjalvare@fdi.ucm.es).

Fernando Sáenz Pérez is with the Institute of Knowledge Technologies of the Complutense University of Madrid (UCM), Spain (e-mail: fernan@sip.ucm.es).

taxonomía. Para poder enseñar relaciones semánticas específicas las herramientas deben ser capaces de controlar el uso apropiado de dichas relaciones, impidiendo de que los usuarios se basen en su intuición para construir la taxonomía, evitándoles así tomar decisiones de modelado incorrectas e inapropiadas. Para imponer tal control nuestra idea es dotar a las relaciones semánticas con un conjunto de propiedades que definan su intension. Entonces, usando estas propiedades como una ayuda, las herramientas de creación podrían asistir al usuario a la hora de que éste trate de enlazar dos conceptos utilizando una relación específica, haciéndole preguntas acerca de la intension de la relación. Este artículo representa un primer paso hacia esa meta.

*Palabras clave*— Control, Diccionarios Electrónicos, Metodología, Propiedades Algebraicas, Propiedades Intrínsecas, Relaciones Semánticas

### I. INTRODUCTION

EN artículos anteriores [1], [2], hemos declarado que la lengua es un valor importante, pero que dentro del aula, es sin embargo, un dominio del conocimiento bastante débil, debido a las dificultades técnicas que presenta tanto a los profesores como a los estudiantes. Se puede argumentar que esta deficiencia de conocimiento lingüístico puede ser mitigada usando recursos léxicos clásicos en formato de papel, como diccionarios, o utilizando sus versiones electrónicas.

Sin embargo, como señala [3], estos recursos presentan enormes restricciones en términos de acceso a las palabras, y a sus homólogos conceptuales: los significados. Por tanto, se debe disponer de entornos adecuados para el aprendizaje de lenguas, si se quiere paliar este tipo deficiencias [4]. Además, las interfaces de estos entornos deben ser diseñadas para proveer una manera intuitiva, coherente y efectiva de acceder a la información deseada dentro de la base de datos.

Hemos desarrollado varias herramientas [1], [2] que superan las deficiencias de la mayoría de los diccionarios electrónicos, al proveer un acceso a la información basado tanto en la forma como en el significado de las palabras [3], [5].

Las herramientas están basadas en una teoría descomposicional del significado de las palabras, que tiene

dos niveles de representación: un nivel conceptual-semántico, y otro sintáctico-semántico.

El primero está representado por una taxonomía conceptual u ontología que expresa el significado de las palabras, el segundo por palabras agrupadas en conjuntos de sinónimos (sysnets) donde información léxica-sintáctica puede ser registrada.

Además, las herramientas siguen un modelo constructivista [6], es decir, proveen medios para la reunión, representación (“externalización” en palabras de [7]), estructuración y creación de objetos a través de los cuales se puede navegar [8].

Este tipo de enfoque tiene varias ventajas educativas. Primero, la ontología sirve como una herramienta cognitiva útil para extender la memoria y facilitar el procesamiento de información, al permitir a un estudiante expresar significados y relaciones de forma directa [9].

Segundo, el agrupamiento de palabras (en synsets) es también cognitivamente útil, dado que el agrupamiento deja claro las similitudes y diferencias entre grupos [9].

Tercero, las relaciones léxicas y semánticas son un aspecto importante en el aprendizaje del lenguaje natural, porque al contrario de los diccionarios basados en papel, las personas no almacenan las palabras en orden alfabético, en vez de eso, lo hacen por significados y relaciones (enlaces/asociaciones entre palabras/conceptos) [3].

Además, las herramientas de autoría controlan la corrección de lo que se está construyendo, impidiendo a los usuarios aprender hechos erróneos.

Este enfoque basado en ontologías se está haciendo popular también en el área de “e-learning”, donde los “topic maps” son usados para la clasificación, navegación o exploración de conceptos, instancias, relaciones y recursos dentro de áreas temáticas específicas [10].

No obstante, el desarrollo de ontologías o taxonomías para propósitos educativos sufre de los mismos problemas de sus homólogos en Inteligencia Artificial: principalmente el mal uso y confusión de las relaciones semánticas [11], debido a la falta de mecanismos de validación de consistencia.

Aunque la comunidad de Inteligencia Artificial ya ha empezado a reconocer la importancia de tales mecanismos [12], [13], este tema ha sido ignorado por la comunidad de Informática Educativa [14].

Por ejemplo, en [10], los autores declararon que su herramienta para representar y estructurar contenidos educativos usa varias relaciones ontológicas (v.g., part-of, superclass-subclass, etc). Sin embargo, no describen ningún mecanismo de verificación para controlar su uso, y asegurar que el significado (cualquiera que este sea) de dichas relaciones sea respetado.

Nuestras herramientas sufren de esta misma falta de mecanismos, ya que sólo permiten representar una taxonomía basada en una sola relación implícita, cuyo significado preciso no está claro. Consecuentemente, el proceso de enlazar dos conceptos con esta relación siempre será subjetivo, es decir, dependiente del usuario y no de un objetivo determinado.

Sería relativamente fácil tomar nuestras herramientas, e integrar mecanismos de validación de relaciones como OntoClean [12] o Archonte [13]. Sin embargo, por razones que aduciremos después, estos enfoques no son del todo funcionales, ya que dejan el objetivo para el cual la ontología está siendo construida fuera de la ecuación, y se enfocan en los conceptos, no en las relaciones, para implementar sus métodos de validación de consistencia.

Además, hemos creado nuestras herramientas siguiendo nuestra propia metodología basada en el desarrollo de bases de datos relacionales y en principios de ingeniería del software [15], y pretendemos seguir usándola.

Por tanto, nuestra meta es ampliar la metodología para: a) ser capaces de representar más de una relación semántica en la construcción de la ontología; b) establecer claramente la intención de cada relación; y c) controlar el uso de estas relaciones, y así evitar que los usuarios tomen decisiones de modelado incorrectas.

El resto del artículo está organizado de la siguiente manera. En la sección 2, señalamos los problemas taxonómicos comunes de diversos diccionarios electrónicos. En la sección 3, explicamos las diferentes metodologías para la estructuración de taxonomías que se encuentran disponibles. En la sección 4, describimos un conjunto de ideas que pretenden ayudar a los desarrolladores a representar más de una relación semántica, a especificar en más detalles el significado de cada una, y a controlar su utilización. En la sección 5, presentamos un modelo conceptual que integra las ideas antes mencionadas. Finalmente, en la sección 6, presentamos nuestras conclusiones y el trabajo futuro.

## II. RELACIONES SEMÁNTICAS EN DICCIONARIOS ELECTRÓNICOS

En sus nuevos formatos electrónicos, los diccionarios tienen un tremendo potencial, suponiendo que se construyan de tal manera que se puedan no sólo por experto o máquinas, sino también por usuarios no expertos y estudiantes. Sin embargo, a pesar del enorme interés en diccionarios electrónicos en general, poca atención se la ha dado a este último tipo de usuarios [5], [16].

Una de las características más importantes que está presente en cualquier diccionario (electrónico o no) es la presencia de relaciones semánticas, que en conjunto forman una red semántica. Sin embargo, estas redes o jerarquías presentan defectos.

Por ejemplo, aunque las versiones electrónicas de los diccionarios de papel (también llamados en Inglés “machine readable dictionaries”) son citadas como recursos útiles para el aprendizaje de idiomas [16], las relaciones en estos recursos son implícitas, y difíciles de detectar por el usuario.

Además, la información que poseen estos recursos está mezclada de tal manera, que las redes relacionales de estos diccionarios contienen definiciones circulares, así como circularidades y rupturas en la representación de conocimiento (v.g., una espátula es un contenedor/a spatula is a container) que pueden obstaculizar el proceso de aprendizaje de

vocabulario. La Fig. 1, tomada de [17], ilustra una ruptura en la representación de conocimiento en uno de estos diccionarios electrónicos.

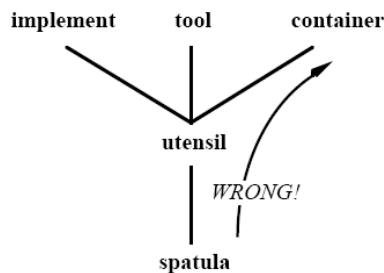


Fig. 1. Una espátula es un contenedor

En WordNet [18], un diccionario ampliamente citado [5], [19] por la comunidad de Aprendizaje de Idiomas Asistido por Ordenador (ALAO o CALL por sus siglas en Inglés), las relaciones “is-a” y “part-of” entre synsets (la estructura que usa WordNet para representar conceptos) no son usadas de una manera consistente.

Por ejemplo, Burgun y Bondenreinder [20] informan que de acuerdo a las relaciones taxonomicas entre los hiperónimos de fiebre/fever en WordNet 1.6, fiebre termina siendo categorizada como una característica psicológica o “psychological feature”. La Fig. 2 muestra este error.

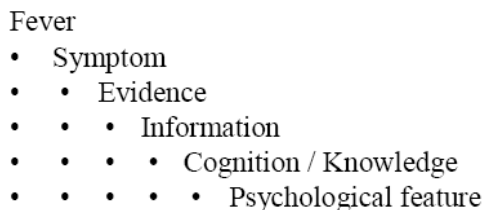


Fig. 2. Fiebre es una característica psicológica

Otro diccionario que presenta problemas similares, con respecto al uso de relaciones semánticas es SNOMED-RT [21]. En este diccionario, podemos encontrar errores como el de ambos testículos son un testículo o “both testes is-a testicle” [22]. La Fig. 3 ilustra este error.

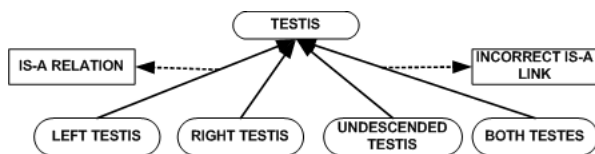


Fig. 3. “Both testes is-a testis”

Los errores antes citados claramente señalan que inclusive mnemónicos aparentemente inofensivos como “is-a” deben manejarse con cuidado mientras se construye una red u ontología, para de esta manera, minimizar el efecto de decisiones de modelado incorrectas y estructurar coherentemente la ontología

Si una herramienta ha de permitir a un estudiante crear una ontología de manera constructiva, dicha herramienta debe ejercer un control sobre el uso que haga el usuario de las

distintas relaciones semánticas, a través de mecanismos de verificación, de manera que la red semántica no sea creada (como pasó con los diccionarios electrónicos, WordNet y SNOMED-RT) dependiendo en la intuición de su(s) creador(es).

Sin embargo, aunque el detectar errores en las entradas del usuario, en términos de la utilización de relaciones es un aspecto clave en cualquier esfuerzo de construcción de un modelo conceptual u ontología, las iniciativas para proveer mecanismos de control para este tipo de proceso han sido marginales: a saber las metodologías OntoClean [12] y Archonte [13].

Estas dos metodologías son similares, pero al mismo tiempo, como veremos abajo, siguen enfoques opuestos para lograr la misma meta.

### III. METODOLOGÍAS PARA ESTRUCTURAR TAXONOMÍAS

#### A. OntoClean

OntoClean está fundada en las ideas filosóficas y metafísicas del esencialismo [23], que establecen que para cualquier tipo específico de entidad (v.g., un tigre), es teóricamente posible especificar una lista finita de propiedades (v.g., las metapropiedades de rigidez, identidad y unidad de OntoClean) las cuales una entidad debe tener para pertenecer a un grupo específico o clase natural (v.g., ver la tabla de propiedades y la taxonomía de clases en [12]).

También está influenciada por las ideas del esencialismo psicológico [24], que enuncia que el mundo está dividido en esencias de las cuales se pueden inferir propiedades preestablecidas (v.g., las metapropiedades antes mencionadas), y que estas propiedades juegan un papel clave en nuestras tareas diarias de razonamiento y categorización, al dar soporte a nuestras inferencias acerca de la pertenencia a una entidad o a una clase natural.

Por tanto, OntoClean puede ser entendida como una heurística para el razonamiento y un sistema de inferencia que establece que la compatibilidad entre las metapropiedades que definen la esencia de los conceptos, determina si un concepto puede subsumir a otro y viceversa.

Sin embargo, una teoría global de referencia y categorización, independiente de cualquier dominio, problema y tarea como la que OntoClean propone no es posible. Trabajos recientes en psicología cultural muestran diferencias cognitivas sistemáticas entre asiáticos y occidentales, y algunos trabajos apuntan que esto se extiende también a las intuiciones filosóficas [25]

Parece que los creadores de OntoClean consideran sus propias intuiciones con respecto a los referentes de los objetos, y la de sus colegas filosóficos como universales, y la evidencia sugiere que está mal asumir “a priori”, la universalidad de sus propias intuiciones semánticas.

Si los últimos 30 años de investigación en ciencia cognitiva nos han enseñado algo, es que muchas de nuestras intuiciones,

tendencias y procesos cognitivos son inescrutables a través de la mera la reflexión (filosófica).

De hecho, es por esto que la psicología cognitiva es tan útil. Si simplemente pudiésemos descubrir lo que hacen nuestras mentes, a través de una reflexión disciplinada, y esto es lo que hacen los filósofos que estudian la referencia, no tendríamos necesidad de hacer experimentos con terceras personas.

Además, es interesante notar que sólo las relaciones “is-a” y “part-of” están fuertemente definidas en OntoClean. Consecuentemente, cualquier otra relación que se salga del ámbito de la metodología no puede ser descrita o controlada.

Por tanto, el limitado número de abstracciones con las que OntoClean pretende dominar la complejidad de cualquier dominio, y estructurar una ontología no es suficiente.

### B. Archonte

Archonte se basa en el trabajo de [26], quién establece que inclusive para dominios bien definidos, las normas que fijan el significado de una palabra y de su referente (v.g. su concepto) no pueden ser previstas, y que el significado de las palabras es inmanente a una situación dada y contexto de utilización. Archonte alega que provee a los conceptos con un significado dependiente del dominio y la tarea, por medio de las similitudes y diferencias que una unidad lingüística tiene con otras unidades vecinas en el mismo contexto de uso, y que estas diferencias y similitudes pueden encontrarse analizando un corpus.

Esta metodología requiere el uso de herramientas de extracción terminológica para descubrir palabras candidatas a conceptos del corpus. Una vez que se ha hecho esto, utiliza un conjunto de principios o propiedades diferenciales [13], para crear y estructurar una taxonomía básica o “backbone taxonomy” (posiblemente basada en la relación “is-a”), donde las diferencias y similitudes entre conceptos son expresados en lenguaje natural. En esencia, esto es muy parecido a las taxonomías o redes semánticas que pueden ser creadas usando las definiciones de los sustantivos en los diccionarios electrónicos. [27].

Sin embargo, mientras que las taxonomías de estos diccionarios contienen circularidades y rupturas en la representación de conocimiento, Archonte provee un método manual con el cual construir taxonomías lingüísticas apropiadamente estructuradas a partir de las palabras y con la ayuda de un experto.

Ya que estos principios van adheridos a los conceptos, aquí yace la similitud con OntoClean. Para estructurar una ontología, son los conceptos (no las relaciones) los que deben tener un conjunto de (meta)propiedades que determinen si una relación semántica puede existir entre ellos.

No obstante, aunque Archonte establece que es dependiente del dominio y la tarea, está claro que es solamente dependiente del dominio. Los conceptos y relaciones son obtenidos al procesar un corpus dependiente del dominio, pero el corpus en si es independiente de cualquier tarea, así como también el conjunto de propiedades usadas para organizar los conceptos.

## IV. CARACTERIZANDO LAS RELACIONES CON PROPIEDADES

Dada la evidencia, está claro que el control de las relaciones semánticas no puede hacerse utilizando solamente las propiedades de los conceptos, y con un conjunto limitado de relaciones.

Además, dicho control no es ni algo que deba hacerse al final (v.g., como lo hace OntoClean) ni manualmente (v.g., como lo hace Archonte), sino permanentemente. Por tanto, el entorno de desarrollo debe ejercer este tipo de control a lo largo del proceso de autoría.

Consecuentemente, nuestra posición es que si a las relaciones semánticas, no a los conceptos, se les da un conjunto de propiedades que definan su intensión; entonces, utilizando estas propiedades como una ayuda, la(s) herramienta(s) de autoría puede hacerle preguntas significativas al usuario para evitarle tomar decisiones de modelado incorrectas.

Se puede discutir que el uso de propiedades para representar la semántica de las relaciones no es algo nuevo. Sin embargo, al contrario de lo que se ha hecho en el pasado [28], creemos que estas propiedades deberían estar enfocadas en el dominio a ser modelado, el problema a ser resuelto y la tarea a llevar a cabo, en vez de darle a estas propiedades un carácter universal.

En consecuencia, dividimos estas propiedades en algebraicas denotando aquellas propiedades necesarias para hacer silogismos válidos (v.g., transitividad, asimetría, reflexividad), y propiedades intrínsecas representando hechos son difíciles de formalizar, es decir, propiedades dependientes del problema y la tarea. La Fig. 4 muestra este enfoque.

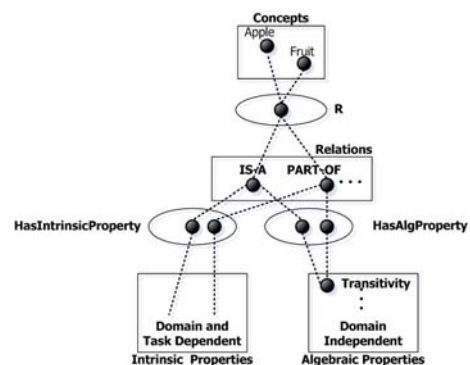


Fig. 4. Representando relaciones con propiedades algebraicas e intrínsecas

Sin embargo, la construcción de la ontología y la aplicación de las propiedades no se hace directamente. Necesitamos un estado inicial antes de empezar a estructurar la ontología y usar las propiedades antes mencionadas.

Por tanto, en vez de construir la ontología de la raíz hacia abajo, seguimos un enfoque “bottom-up”, donde la idea es movernos de un conjunto de términos hacia un grupo de conceptos que referencian estos términos, para así formar una estructura lexico-conceptual primitiva (ELCP) desprovista de relaciones taxonómicas, cuyos conceptos serán organizados

más tarde, como una ontología, mediante un conjunto de relaciones semánticas.

En esta ELCP, cada concepto tiene una definición intensional dada en lenguaje natural. Además, cada definición intensional contiene una colección de palabras claves, seleccionadas por el usuario, las cuales están ya contenidas en la definición, que nos ayudarán, por medio de la herramienta de autoría, a relacionar un concepto con otros conceptos y términos. La Fig. 5 muestra este enfoque.

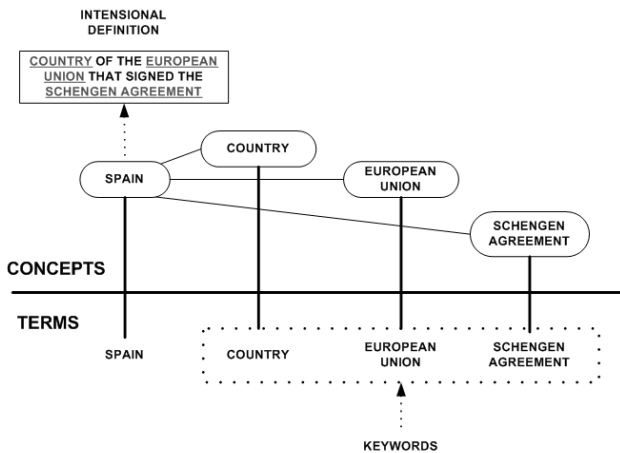


Fig. 5. Ejemplo de una Estructura Lexico-Conceptual Primitiva

En la Fig. 5, el concepto “Spain” tiene una definición con tres palabras clave (“Country”, “European Union” y “Schengen agreement”). Estas tres palabras son usadas para asociar los conceptos “Spain”, “Country”, “European Union” y “Schengen agreement”.

Aquí es importante subrayar que las herramientas de autoría deben asegurar la completitud de la ELCP, es decir, si en una definición intensional, el autor selecciona una palabra clave que no existe, entonces, esta palabra debe ser creada e incluida como parte del conjunto de términos y conceptos de la ELCP.

Además, si se da la situación donde un término dado sea polisémico, entonces, las herramientas deben señalar este hecho, y forzar al desarrollador a escoger entre los diferentes conceptos que denotan al término.

Una vez que la ELCP se ha completado, entonces podemos empezar a organizar los conceptos usando un conjunto de relaciones significativas. Sin embargo, varias preguntas interesantes surgen aquí. Primero, ¿hay un conjunto de relaciones que sea común a todos los dominios, problemas y tareas? Segundo, ¿cual de todas las posibles relaciones es la primera relación que se debe utilizar para estructurar un dominio?

Para la primera pregunta, afirmamos que de la misma manera que el conjunto vacío ( $\emptyset$ ) es parte de todos los conjuntos posibles, hay un conjunto de relaciones que siempre será parte de cualquier conjunto de relaciones, el conjunto formado por las relaciones “is-a” y “part-of”: {“is-a”, “member-of”}.

En cuanto a la segunda pregunta, la primera relación que debe ser usada es “member-of”, ya que siempre empezamos la construcción de un modelo del dominio identificando o seleccionando un conjunto de objetos de interés. Entonces, la relación “is-a” aparece al descubrir las propiedades comunes de todos los miembros pertenecientes a un conjunto extensión. La Fig. 6, ejemplifica esta idea.

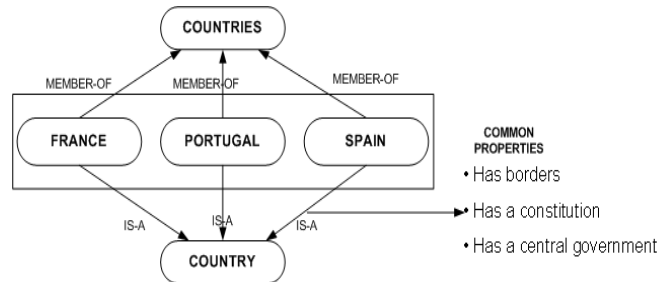


Fig. 6. Usando las relaciones “member-of” e “is-a” para Construir un Modelo del Dominio

Otra relación que puede aparecer frecuentemente es la “component-of”, que denota un objeto o concepto compuesto por otros, que en la ausencia de alguno de sus componentes deja de existir.

La razón de ser de estas ideas es que con estas tres relaciones, como veremos más tarde, podemos echar abajo errores comunes y modelar correctamente ontologías que han sido previamente mal modeladas. Pero primero, daremos un ejemplo de la utilización de las propiedades algebraicas e intrínsecas.

Por ejemplo, supongamos que estamos tratando de desarrollar una ontología para dar soporte a un sistema de información legal, para tratar temas relacionados con los movimientos entre fronteras en países signatarios del tratado de Schengen, y que ya hemos construido nuestra ELCP.

En este punto, si quisiéramos enlazar el concepto “Spain” con el concepto “Schengen agreement” usando la relación “is-a”, el sistema sólo permitiría esta operación si “Spain” cumpliera con todas las propiedades algebraicas e intrínsecas desde su nivel de profundidad hasta la raíz. La Fig. 7 muestra esta idea.

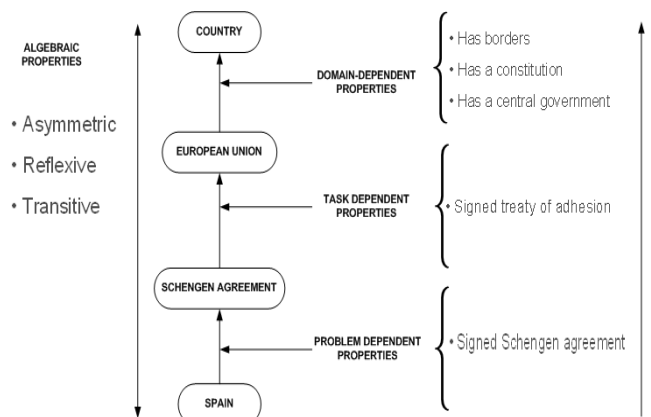


Fig. 7. Propiedades algebraicas e intrínsecas a lo largo de una sola relación

En consecuencia, tenemos que para cualquier dominio, problema y tarea, una ontología estará determinada, primero por una ELCP, y luego por un conjunto de relaciones semánticas, donde cada relación tiene su propio grupo de propiedades algebraicas e intrínsecas.

Ahora, para probar la utilidad de el conjunto común de relaciones mencionadas anteriormente, remodelemos (sin propiedades algebraicas e intrínsecas ya que es un ejemplo trivial) el error “both testes is-a testis” de [22] representado en la Fig. 3.

Primero tenemos que los conceptos “undescended testis”, “left testis” y “right testis” representan un concepto extensión del cual son miembros, denotado por el plural “testes”. La Fig. 8 ilustra este hecho.

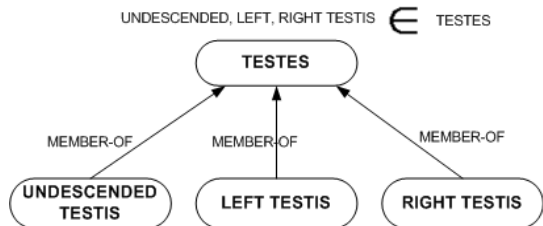


Fig. 8. La Estructuración comienza con la Relación “member-of”

El concepto “testis” es creado como resultado de la abstracción de las propiedades comunes de todos los “testes” y la relación “is-a” aparece naturalmente. La fig. 9 muestra este proceso.

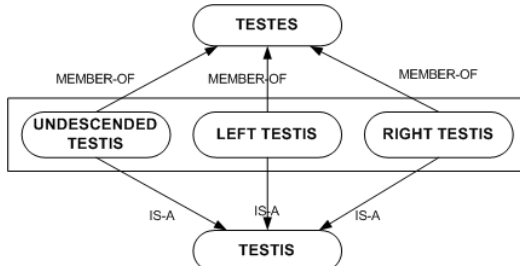


Fig. 9. Testis es la abstracción de tres “testes” diferentes

Después de esto, queda claro que “both testes” es un subconjunto especial del conjunto extensión “testes”, y que la siguiente relación que aparece naturalmente es “component-of”, tal y como se muestra en la Fig. 10.

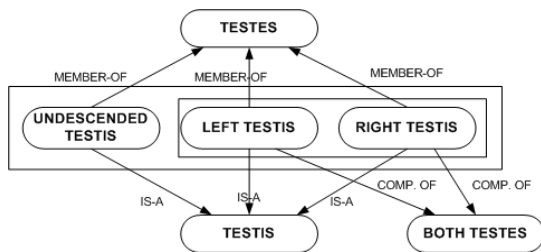


Fig. 10. “Both testes” es un subconjunto especial de “Testes”

La justificación detrás de las figures 8, 9 y 10 es que debemos agotar las relaciones basadas en conjuntos (set theoretic relations) antes de movernos hacia otras relaciones

que no lo son como la “adjacent to” que hemos descrito en [28]. También implica que: a) una sola relación (“is-a” o “subsumption”) siempre nos inducirá a errores; b) “member-of” e “is-a” siempre deben contemplarse; y c) la complejidad de los dominios debe ser atajada relación por relación.

V. DISEÑANDO UN DICCIONARIO ELECTRÓNICO BASADO EN ONTOLOGÍA

En nuestro trabajo previo [1], [2] desarrollamos un par de herramientas para la construcción y consulta de diccionarios basados en ontologías con propósitos educativos. Sin embargo, ambas tanto los modelos conceptuales como las herramientas tenían varios defectos con respecto a la representación de relaciones.

El primer modelo, como puede verse en la Fig. 11, sólo permitía crear una ontología con estructura de árbol, y con una sola relación implícita de significado indefinido. El segundo modelo, que se muestra en la Fig. 12, aunque permitía la construcción de una ontología en forma de grafo acíclico, retenía el mismo defecto en términos de representación de relaciones.

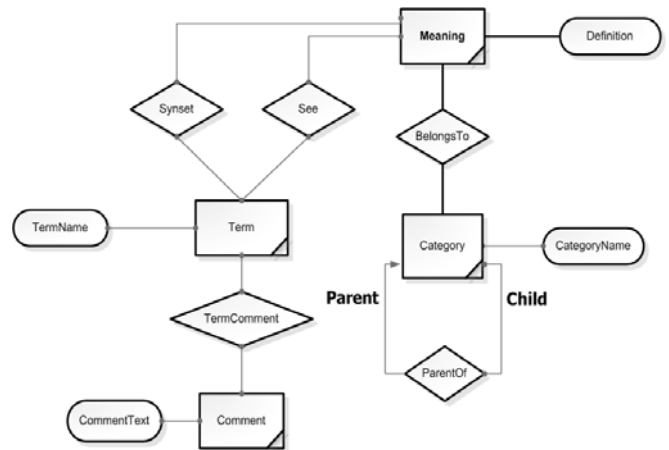


Fig. 11. Modelo E-R para un diccionario monolingüe con una ontología en forma de árbol

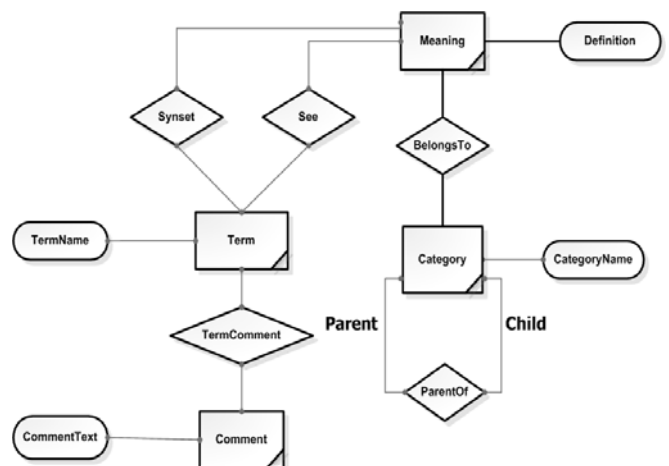


Fig. 12. Modelo E-R de un diccionario monolingüe con una ontología en forma de grafo acíclico

Pare poder superar estas limitaciones e incluir las ideas que presentamos en la sección 4, introducimos un nuevo modelo E-R de una diccionario electrónico basado en ontología (ver Fig. 13), que representa el resultado de la primera fase de diseño (modelado conceptual).

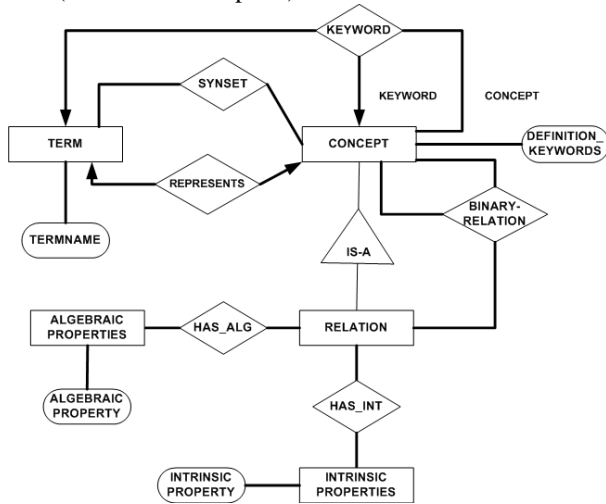


Fig. 13. Modelo E-R para un diccionario con múltiples relaciones

#### A. Conjuntos de Entidades

El conjunto de entidades “Term” representa todos los terminos que componen el diccionario, y cada término está simbolizado por un nombre representado por el atributo “TermName”.

El conjunto de entidades “Concept” denota el significado de las palabras y tiene un atributo: “Definition\_Keywords”. Este atributo representa la definición intensional de cada concepto, así como el conjunto de palabras clave que pueden ser parte de cada definición.

El conjunto de entidades “Relation” representa todas las relaciones en la ontología. El conjunto de entidades “Algebraic properties” e “Intrinsic properties” corresponde al grupo de propiedades algebraicas e intrínsecas que una relación dada puede tener respectivamente.

#### B. Conjuntos de Relaciones

El conjunto de relaciones “Synset” representa a un conjunto de términos correlacionados con un concepto dado. “Synset” tiene una cardinalidad de muchos a muchos para denotar la sinonimia y la polisemia, ya que un concepto puede ser expresado por varios términos y un término puede representar varios conceptos.

El conjunto de relaciones “Represents” se usa para establecer que para un conjunto de sinónimos, hay un término que es representativo del concepto denotado por ese conjunto. “Represents” tiene una cardinalidad de uno a uno porque, se asume que solamente un término del conjunto de sinónimos puede ser representativo, y que es poco probable que el mismo término pueda ser representativo de conjunto de sinónimos diferente.

El conjunto de relaciones “Keyword” establece que un concepto puede estar relacionado con otro concepto por medio

de un término (una de las palabras clave en la definición) y que este término denota a un concepto. “Keyword” tiene una cardinalidad de muchos a muchos de su lado recursivo porque un concepto puede estar relacionado (a través de sus palabras clave) con uno o más conceptos, y es de uno a muchos entre “Term” y “Concept” porque cada concepto está denotado por un solo término.

El conjunto de relaciones “Binary relation” se usa para establecer que un concepto está relacionado con otro concepto a través de una relación semántica binaria. Su cardinalidad es de muchos a muchos porque un concepto puede estar enlazado a muchos conceptos, y un concepto puede estar enlazado con otros conceptos a través de muchas relaciones.

El conjunto de relaciones “Has\_Alg” y “Has\_Int” representa la idea de que una relación puede tener un conjunto de propiedades algebraicas y propiedades intrínsecas respectivamente. Ambos conjuntos tienen una cardinalidad de muchos a muchos porque una relación puede tener varias propiedades algebraicas e intrínsecas y la misma propiedad algebraica o intrínseca puede estar presente en diferentes relaciones.

## VI. CONCLUSIONES Y TRABAJO FUTURO

Estructurar una ontología y enlazar conceptos a través de relaciones semánticas son procesos de toma de decisiones que normalmente permanecen oscuros y subjetivos. Los entornos de desarrollo de ontologías tradicionales no consideran como se toman este tipo de decisiones, y como una computadora puede ayudar a mejorar la claridad de tales decisiones.

Creemos que este proceso debe estar basado en la cooperación provista por un dialogo continuo entre el usuario y el sistema. El corolario de este dialogo será el uso coherente de las relaciones semánticas, destinando poder de computo, a un punto específico de la construcción de ontologías, que sufre de errores o deficiencias por parte del usuario.

Para llevar esto a cabo, hemos propuesto un enfoque de estructuración de ontologías dividido en dos niveles. En el primero, nos movemos de los términos a los conceptos para crear una ELCP. En el segundo, en vez de seguir el principio estándar de abstracción de clases, utilizamos uno de construcción de conjuntos, que junto con las propiedades algebraicas e intrínsecas puede ser usado para estructurar coherentemente una ontología.

Un hecho que cabe señalar, es que estas ideas no solamente ayudarán en la estructuración de la ontología con respecto a una tarea dada dentro de un dominio, pero también servirá para enseñar como utilizar apropiadamente estas relaciones en la construcción de diccionarios para el aprendizaje de idiomas.

Las relaciones escogidas pueden variar de dominio en dominio y de tarea en tarea. Sin embargo, dos relaciones que siempre debemos tener en cuenta en el proceso de construcción del diccionario son la “member-of” y la “is-a”, ya que aparecen de manera natural en la construcción del modelo de un dominio.

También introducimos y presentamos un sólido modelo conceptual para diccionarios monolingües basados en ontologías, que toma en cuenta las cuestiones de representación de conocimiento con respecto a la caracterización de relaciones semánticas con propiedades y su uso para el control y verificación durante el proceso de construcción de la ontología.

Basándonos en este modelo conceptual, actualmente estamos en el proceso de construir las herramientas de autoría de diccionarios. El trabajo futuro incluye instalar estas herramientas en el aula, así como probarlas mediante la creación y consulta de diccionarios específicos para ver si los objetivos pedagógicos previstos son cumplidos.

#### REFERENCIAS

- [1] Vaquero, A., Sáenz, F. and López, C., "Herramientas para la Creación de Diccionarios Monolingües con Objetivos Pedagógicos". Challenges 2003 – 5<sup>th</sup> International Symposium on Computers in Education (SHE, 2003), 2003.
- [2] A. Vaquero, F.J. Álvarez, and F. Sáenz, "Learning Linguistic Concepts through the Construction of Dictionaries with a DAG shaped Taxonomy". In *New Trends and Technologies in Computer-Aided Learning for Computer-Aided Design*, Springer, 2005, pp. 91-106.
- [3] Bilac, S. and Zock, M., "Towards a User-Friendly Dictionary Interface". In *Proceedings of the Papillon 2003 Workshop*, 2003.
- [4] Zeltzen, D. and Addison, R. K., "Responsive Virtual Environments". *Communications of the ACM*, 40(8), 1995.
- [5] Zock, M., "Electronic Dictionaries for Men, Machines or for both?" Third Papillon seminar, 2002.
- [6] Cabrera, A., "Informática educativa: La revolución constructivista". In *Informática y Automática*, 28(1), 1995
- [7] Vygotsky, L. *Thought and Language*, The MIT press, 1986.
- [8] Zeiliger, R., Belisle, C., Cerratto, T., "Implementing a Constructivist Approach to Web Navigation Support". In *proceedings of the ED-MEDIA'99 Conference*, 1999.
- [9] Tversky, B., "Some ways that maps and diagrams communicate". In *Freksa, C., Brauer, W., Habel, C., and Wender, K. F.(eds). Spatial cognition II: Integrating abstract theories, empirical studies, formal models, and practical applications*. Springer-Verlag, Berlin, 2000.
- [10] Dicheva, D., Dichev, C., Wang, D., "Visualizing Topic Maps For E-Learning". In *proc. of the Workshop on Applications of Semantic Web in E-Learning*, 2005.
- [11] Guarino, N., "Some Ontological Principles for Designing Upper Level Lexical Resources". In *proc. of the First International Conference on Language Resources and Evaluation (LREC)*, 1998, pp. 527-534.
- [12] Guarino, N. and Welty, C., "An Overview of OntoClean". In *Steffen Staab and Rudi Studer (Eds.): The Handbook on Ontologies*, 2005, pp. 151-172.
- [13] Bachimont, B., Isaac, A. and Troncy, R., "Semantic Commitment for Designing Ontologies: A Proposal". In *Proc. Of the 13th International Conference on Knowledge Engineering and Knowledge Management*, LNAI 2473, 2002, pp. 114-121.
- [14] Rath, H.H., "Making topic maps more colourful". In *proceedings of XML Europe 2000 Conference*, 2000. Available at: <http://www.gca.org/papers/xml europe 2000/papers/s29-01.html>
- [15] Sáenz, F. and Vaquero, A., "Applying Relational Database Development Methodologies to the Design of Lexical Databases". *Database Systems 2005, IADIS Virtual Multi Conference on Computer Science and Information Systems (MCCSIS 2005)*, 2005.
- [16] Hasan, Maruf, Takeuchi, K., Sornlertlamvanich, V. and Isahara. H. "Putting NLP Tools into Action: Incorporating NLP Tools in Web-based Language Learning Environment". In *Proceedings of the Asian Symposium on Natural Language Processing to Overcome Language Barriers, IJCNLP-04*, 2004, pp. 85-90.
- [17] Ide, N., and Veronis, J., "Extracting Knowledge Bases from Machine-Readable Dictionaries: Have we wasted our time?" In *Proceedings of the First International Conference on Building and Sharing of Very Large-Scale Knowledge Bases*, 1993, pp. 257-266.
- [18] Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K. & Teng, R. *Five Papers on WordNet*. CSL Report 43, 1990.
- [19] Morante, R. and Martí, M.A. "EuroWordNet as a resource for learning Spanish verbs". In *Proceedings of the First International Conference on WordNet*, 2002, pp.231-238.
- [20] Burgun, A. and Bodenreider, O., "Aspects of the Taxonomic Relation in the Biomedical Domain". In *proc. of the 2<sup>nd</sup> International Conference on Formal Ontologies in Information Systems*, 2001, pp. 222-233
- [21] Spackman, K.A., Campbell, K.E, and Côté, "SNOMED RT: a reference terminology for health care". In *proc. of the Am. Med. Inform. Assoc. Symp.*, 1997, pp. 640-644.
- [22] Ceusters, W., Smith, B., Flanagan, J., "Ontology and Medical Terminology: Why Description Logics Are Not Enough". In *Proc. of the Conference: Towards an Electronic Patient Record (TEPR 2003)*, 2003, pp. 183-195.
- [23] Barrett, H.C., "On the Functional Origins of Essentialism". In *Mind and Society* 3(2), 2001, pp. 1-30.
- [24] Medin, D. & Ortony, A., "Psychological Essentialism". In *S. Vosniadou & A. Ortony (eds.) Similarity and Analogical Reasoning*. Cambridge University Press, 1989, pp. 179-195.
- [25] Machery, E. et al, "Semantics Cross-Cultural Style". In *Cognition* 92, 2004, pp.B1-B12.
- [26] Rastier, F., Cavazza, M., Abeillé, A., "Sémantique pour l'analyse. De la linguistique à l'informatique", Masson, 1994.
- [27] Wilks, Y., Sator, B., Guthrie, L. *Electric Words: Dictionaries, Computers and Meanings*. MIT Press, 1996.
- [28] Alvarez, F. Vaquero, A., Sáenz, F., Buenaga, M. *Neglecting Semantic Relations: Consequences and Proposals*. In *Proceedings of the IADIS International Conference on Intelligent Systems and Agents (ISA 2007)*, 2007, pp. 99-108.



**Antonio Vaquero Sánchez** recibió su grado de licenciado en ciencias físicas en 1961 y su grado de Doctor en Física en 1967. Actualmente es catedrático de universidad y parte del departamento de Ingeniería de Software e Inteligencia Artificial de la Facultad de Informática de la Universidad Complutense de Madrid (UCM). Sus intereses de investigación incluyen: diseño lógico, sistemas de enseñanza y aprendizaje, sistemas de autor, procesamiento de lenguaje natural, recursos lingüísticos, representación de conocimiento y ontologías.



**Francisco Alvarez Montero** tiene una licenciatura en Informática por parte del Instituto Tecnológico de Culiacán (ITC). Recibió su diploma de estudios avanzados (DEA) en 2004 de la Universidad Complutense de Madrid. Es parte del departamento de Ingeniería de Software e Inteligencia Artificial de la Facultad de Informática de la Universidad Complutense de Madrid (UCM) donde actualmente trabaja en su tesis doctoral con una beca de la



Universidad Autónoma de Sinaloa (UAS). Sus intereses de investigación incluyen: representación de conocimiento, ontologías y bases de datos relacionales.

**Fernando Sáenz Pérez** recibió su grado de licenciado en ciencias físicas en 1988 y su doctorado en física en 1995, ambos por parte de la Universidad Complutense de Madrid. Actualmente es profesor titular de universidad y parte del departamento de Ingeniería de Software e Inteligencia Artificial de la Facultad de Informática de la Universidad Complutense de Madrid (UCM). Sus intereses de investigación incluyen: ontologías, bases de datos y programación declarativa. Ha formado parte de varios proyectos de investigación de I+D en colaboración con corporaciones como Repsol YPF, realizando transferencias de investigación. Además, participa activamente en varios proyectos de código libre.