

# Topic Maps: Una realidad en constante expansión

S. Sánchez Cuadrado; J. Morato; J. Llorens  
Departamento Informática Universidad Carlos III  
Av. Universidad 30 – 28911 Leganés (Madrid) España

José. A. Moreiro, C. Arellano, P. Beltrán  
Departamento Documentación Universidad Carlos III  
C/Madrid, 126 – 28903 (Getafe) España

## RESUMEN

Los *Topic Maps* son un estándar que se está implantando en el web y que posibilita la navegación conceptual. Se presenta una panorámica de los *Topic Maps* que refleje los distintos aspectos de su interés. Estos son, por un lado los que representan alguna novedad en su implantación en el web, como la navegación conceptual o la inferencia, por otro lado, los que concentran una gran actividad investigadora en este estándar, como las reglas de consistencia, validación o las propiedades que permitan la inferencia.

## PALABRAS CLAVES

*Topic Maps*, Web, Navegación conceptual, inferencia.

## INTRODUCCIÓN

Nos proponemos analizar en este informe las ventajas y líneas de investigación que ha abierto el nuevo estándar *Topic Maps* (TM). Un TM tiene como finalidad normalizar los elementos y la notación utilizada para estructurar la información mediante la construcción de una red de enlaces semánticos que relacionen diferentes recursos informativos.

TM tiene su origen en el grupo de Davenport, un foro destinado a productores de libros electrónicos que surgió a principios de la década de los 90. En 1993 se propuso la creación de una norma cuyo principal objetivo fuera posibilitar la fusión de índices impresos. Posteriormente evolucionó hacia otras estructuras (como tesauros), hasta llegar a ser una herramienta considerada en la web para la Organización, Representación y Gestión del Conocimiento. La primera versión oficial del estándar ISO/IEC data del año 2000[1].

En la práctica, la implementación de los TM se diseñó para arquitectura SGML con notación de HyTime, la DTD del estándar ISO/IEC 13250:2000 se ha mantenido en HyTime. No obstante, esta notación ha caído en desuso tras la aparición de XML. En este caso, las causas del desplazamiento de la notación de HyTime se deben a la creación de una DTD para crear *Topic Maps* en XML, denominada XTM, y al auge que ha obtenido XML. En cualquier caso, actualmente se pueden encontrar ejemplos de ambos lenguajes en la Web.

Dada la novedad del estándar TM y su, todavía, escasa implantación en el web, se ha considerado oportuno revisar, en

primer lugar, sus elementos principales, para luego centrarnos en su análisis y perspectivas futuras. En el presente documento se ha optado por mantener los términos originales en inglés para eliminar ambigüedades en los ejemplos de XML.

## Elementos Principales

Los elementos principales de los *Topic Maps* son los *topics*, las *occurrences* y las *associations*. Aunque hablando con rigor, *occurrences* y *sociations* representan sólo una parte del *topic*

Una persona para identificarse sin ambigüedad no solo tiene que decir su nombre, si no que también se le suelen pedir otros datos como el DNI o los nombres de los padres. Análogamente, si tengo un *topic* que sea “Blanca” para definirlo sin confusión puedo asociarlo a otros datos o documentos como que se trata de una “persona” (*topic type*), con “DNI 0000000”(occurrence), hijo de “x” y de “y”(associations). En las próximas líneas, vamos a definir los elementos fundamentales del *Topic Maps*.

## Topics

Un *topics* es el elemento principal de un TM, es el término que expresa determinado concepto o idea (o *subject*, como se denomina en el estándar). Ejemplos de *topics* pueden ser “Europa”, “persona”, “idea” o “continente”.

Los *topics* se pueden asociar con otros *topics* denominados *topic types*, p.e “Europa” puede tener un *topic type* que sea “continente”. Los *topic types* definen relaciones clase-instancia(de ahí su nomenclatura en XTM: *instanceOf*) para las relaciones superclase-subclase se debe de crear una *association* específica.

Los *topics* tienen tres características principales: su denominación (*names*), sus apariciones (*occurrences*) y su rol en las asociaciones (*role associations*). Dos *topics* con características equivalentes se consideran semánticamente idénticos. A continuación se describen estas características:

### 1) NAMES

Un *topic* puede tener varias denominaciones, pero debe estar representado por una forma base (*base name*). El *base name* es un elemento obligatorio y representa la forma usual de hacer mención al *topic*. Es complementario del identificador interno (id). Los *base name* deben de ser únicos en un dominio determinado bajo un *scope* dado.

Un aspecto interesante cuando se fusionan dos TM es el procedimiento con los distintos *base names* de los *topics* que hacen mención a un mismo *subject* (concepto). En este caso habría que normalizar los *base names* de alguna manera, para esto se recurre a los *public subject*, que es un recurso que identifica al *subject* de forma no ambigua.

Además el *topic* puede tener otras denominaciones (*alternative names*), como el *display name*, que es la forma en la que se mostrará al usuario, y el *sort name*, que es como se ordenará alfabéticamente cuando se presente un listado.

## 2) OCCURRENCES

*Occurrences* son enlaces a recursos informativos, p.e. una página web. Son elementos opcionales relevantes a un *topic* dado, es decir un *topic* puede tener cero o muchas *occurrences*. Los recursos informativos pueden ser de muchos tipos, una cita de un texto, una definición, una ley, un artículo, etc. Cada uno de los diferentes tipos documentales puede ser agrupado mediante *Occurrence roles* (p.e. diccionario, página web, imagen, etc). Hay que hacer notar que la mayoría de estas *occurrences* suelen ser externas al TM, siendo una situación análoga a la que siempre ha existido entre un tesoro y la indización documental realizada con ese tesoro.

Existen dos tipos de *occurrences*: *resourceRef* que es un enlace a un recurso externo de información; y *resourceData* que es algún dato no externo que se facilita.

Cuando se diseña un TM se suele aconsejar que el número de *occurrences* se limite a unos pocos recursos muy relevantes.

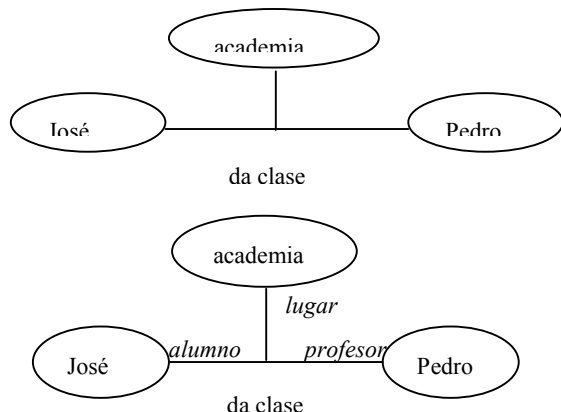
## 3) ASSOCIATIONS

Una *association* representa un enlace que establece una relación entre dos o más *topics*. Una forma de comprobarlo es tomar una frase y considerar los sustantivos de esas frases como *topics* empleando los términos de unión entre los *topics* para denominar la asociación. Por ejemplo:

- “la trementina *se obtiene de* la resina”
- “el pino silvestre *se denomina* pino albar”
- “la resina *está en* los pinos”
- “las flores masculinas *se localizan en* las ramas bajas”
- “En Cádiz *encontramos* pinsapos”
- “las piñas *están formadas por* piñones”

Las *associations* se pueden agrupar por *Association type*. De este modo se puede agrupar “se encuentra en” y “se localiza en” como una *association type* de “ubicación”.

Existen otros elementos que disminuyen la ambigüedad, así por *Association roles* hacemos referencia al rol que



desempeña determinado *topic* en una *association*. En la frase “En Cádiz *encontramos* pinsapos” tenemos que “Cádiz” desempeña un rol de lugar y el término “pinsapo” designa un tipo de árbol.

Ejemplo de *Association roles*:

En una *association*, los *topics* se denominan miembros y los miembros representan roles, por ejemplo de la frase “Pedro da clase a José en la academia” se puede obtener la siguiente representación ternaria:

Sin añadir más información no podré saber quién es el profesor y quién el alumno. Esta información se puede expresar mediante roles:

Como anteriormente, los roles “alumno”, “lugar” y “profesor” deberán ser definidos como *topics*, al igual que “José”, “Pedro” y “da clase”

La forma de expresarlo en XTM es la siguiente:

```
<topicMap>
<topic id="daclasea">
  <baseName>
    <baseNameString>Da clase
  </baseNameString>
  </baseName>
</topic>
<topic id="alumno">
  <baseName>
    <scope><topicRef
  xlink:href="#enseñanza"
    <baseNameString>Alumno</baseNameString>
  </baseName>
  </topic>
<topic> ...</topic>
<topic id="universidad">
  <instanceOf>
    <topicRef
  xlink:href="#organización"/>
  </instanceOf>
  <baseName>
    <baseNameString>Universidad</baseNameString>
  >
  </baseName>
  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#homepage"/>
    </instanceOf>
    <resourceRef
  xlink:href="http://www.uc3m.es"/>
  </occurrence>
</topic> ... </topic>
</association>
```

```

<instanceOf><topicRef
xlink:href="#daclase
a"/></instanceOf>
<member>
  <roleSpec><topicRef
xlink:href="#alumno"/></roleSpec>
  <topicRef
xlink:href="#Jose"/>
</member>
<member>
  <roleSpec><topicRef
xlink:href="#lugar"/></roleSpec>
  <topicRef
xlink:href="#universidad"/>
</member>
<member>
  <roleSpec>
    <topicRef
xlink:href="#profesor"/>
  </roleSpec>
  <topicRef
xlink:href="#Pedro"/>
</member>
</association>
...
</topicMap>

```

## Recuperación de Información con TM en la Web

### 1. Navegabilidad e Inferencia

La navegación tradicional en la web se realiza mediante hiperenlaces que asocian unos recursos informativos a otros, es decir se realiza entre *occurrences*. Uno de los principales beneficios de los TM reside en la navegabilidad, con TM se puede navegar mediante hiperenlaces entre *topics* que no tengan un recurso asociado.

No obstante, la navegabilidad entre *topics* (o *occurrences*) relacionados no tiene un sentido lógico, dicho de otro modo, por si misma la asociación que se establece entre Cádiz y Pinsapos en la frase “en Cádiz se encuentran Pinsapos” sólo afirma que Cádiz está relacionado con los Pinsapos y los Pinsapos con Cádiz, y que desde cualquiera de los dos elementos se puede acceder (“navegar”) hacia el otro.

Otro aspecto interesante de la utilización de TM en la Web es la inferencia. Es decir los mecanismos por los que podemos obtener una información no explícita en el TM.

Las inferencias que se pueden realizar a partir de un conjunto de asociaciones están marcadas por las propiedades que tienen las *association type* implicadas. Aunque en la literatura se mencionan distintas propiedades como la reflexiva, conectiva, etc. [2] en esta sección nos centraremos en la transitiva y simétrica por ser las más empleadas para inferencia y creación de TM. A continuación se analizan estas propiedades someramente:

### a) Simetría

La frase “el pino silvestre se denomina también pino albar” se podría haber escrito “el pino albar se denomina también pino silvestre” sin modificar su semántica. Frecuentemente las relaciones simétricas establecen una relación entre dos *rol types* idénticos o similares (p.e. en el ejemplo, el *rol type* de “pino albar” y de “pino silvestre” podría ser en ambos casos árbol). Las sinonimias son un ejemplo de *association type* que tiene esta propiedad.

### b) Transitividad

Es la propiedad que permite declarar implícitamente un hecho a través de varias *associations*, se suele encontrar en las *association types* de tipo superclase-subclase (pero no en la tipo-instancia) y parte-todo. Un ejemplo, a partir de “la trementina *es parte de* la resina” y “la resina *es parte de* los pinos” se puede deducir que la trementina está en los pinos.

De cualquier modo, ni con las asociaciones más conocidas la inferencia está siempre clara. Uno de estos casos, con una *association type* parte-todo, podría ser decir que “Los pies de Juan son una parte de Juan” y que “Juan es parte de la plantilla de una editorial”, resultaría extraño deducir que “el pie de Juan es parte de la plantilla de una editorial”, en este caso el problema es que la relación entre Juan y sus extremidades es mucho más fuerte que entre Juan y su puesto de trabajo.

### Inferencias basadas en conjuntos de *association type* no homogéneas

Por supuesto existen otros tipos de inferencias no transitivas, este es el caso cuando se quieren enlazar un conjunto heterogéneo de *association types*. Un ejemplo, si A es hijo de B y primo de C; y A y C son nietos de D entonces puedo decir que B es sobrino de A. La forma en que un sistema puede llegar a esta conclusión queda, hoy por hoy, fuera del alcance de TM. Téngase en cuenta que habría que estudiar que *association types* o que subconjuntos de estas y en que condiciones se pueden llegar a enlazar para llegar a una determinada deducción.

Lamentablemente, aunque los autores de TM [2] hablan a menudo de estas propiedades para hacer inferencias o ampliar un TM, ni el estándar ni las DTD publicadas incluyen esta información dentro del TM. En la práctica es una parte del sistema el que hace estas inferencias como es el caso de K-42 (<http://k42.empolis.co.uk/inferencing.htm>). De cualquier modo sobre la forma de incluirlo en el *Topic Map* existen propuestas como la de Rath[3] que contiene esta información dentro de las facetas o la de Freese [4] que sugiere contenerlo en la sintaxis XML al definir el TM incluyendo una etiqueta *rule* de las *association type*, p.e.:

```

<rule reflexive="0" transitive="0" symmetrical="0"
type="member-collection">

```

### 2. Visualización

Los TM representan una de las principales propuestas para la visualización de la web semántica [5]. Existe una problemática obvia en este tema cuando se tiene en cuenta que un TM puede tener cientos de miles de asociaciones de diferente tipología (*association types*, *roles*, *occurrences*, etc). Actualmente, entre las representaciones propuestas destacan los árboles, los

*browsers* y los gráficos. A la hora de analizar estas representaciones se debe tener en cuenta que muestren tanto información local de los *topics* que interesan al usuario, como información sobre la localización de esos *topics* en el conjunto del TM.

### 3. Recuperación

Los mecanismos de inferencia y la declaración de los distintos *type* del TM pueden mejorar en gran medida los sistemas de recuperación. Así si se pregunta por “productos extraídos de los árboles”, mediante los *topic type* podré saber que “pino” es un <árbol>, “resina” un <producto> que tiene una *association type* <obtenidos de> con “pino”, y que por la propiedad transitiva me permite llegar al <producto> “trementina”, así podré contestar a la pregunta con los términos “resina” y “trementina”.

Existen distintas propuestas sobre lenguajes para recuperar TM de la manera descrita, destacan TMQL, un lenguaje que será próximamente un estándar ISO (ver anexo), y TOLOG, un lenguaje desarrollado por Ontopia.

#### Relación con otros recursos para representar el conocimiento

Un aspecto interesante es la relación con la gestión del conocimiento (KM o Knowledge Management). Hoy en día, la Organización del Conocimiento significa que se debe facilitar a los empleados de una empresa un fácil acceso al conocimiento representado como documentos y datos de una organización. Si analizamos la diferencia entre Información y KM es la misma que entre tener un libro y saber de la materia que trata. La causa reside en que la información encontrada en el web no está estructurada, por cuanto buena parte de ella se encuentra en lenguaje natural. Existen manifestaciones en contra a este respecto [6], indicando que el concepto de la “información no estructurada” no existe. Cualquier tipo de información tiene una secuencia en la cual hay un principio, un nudo y un final, en la que existe alguna unidad de concepto, y normalmente varios niveles de jerarquía o subunidades. El lenguaje Natural (LN) tiene estructuras, aunque por sus características un ordenador no sea capaz de entenderlo con facilidad, pero eso no justifica que no tenga una estructura. No obstante es cierto que los lenguajes formales son más sencillos de automatizar que el LN. En este punto es donde la tecnología deriva a la Inteligencia Artificial, el Procesamiento del Lenguaje Natural, el análisis lingüístico y el reconocimiento Semántico.

Por lo tanto se parte de la idea de que la información representa el conocimiento o cambios en el mismo, es decir:

Conocimiento = Datos + Interpretación

Datos = Conocimiento – Interpretación

La interpretación es muy importante entendida como una representación entre conjuntos de estructuras de datos y de modelos de conjuntos de objetos en el universo del discurso, con respecto al significado de esos objetos y las relaciones entre ellos.

Las investigaciones van orientadas a descubrir estadísticamente correlaciones significativas de datos, palabras y unidades relativamente no estructurada y mediante esas correlaciones inferir significado semántico. De este modo se asume que la posición y concurrencia entre algunos contextos definidos implica una relación semántica. Las relaciones semánticas deben ser concebidas e interpretadas por un humano. Por lo tanto, los fundamentos y niveles necesarios que debe tener un sistema que pretenda parecerse a un humano son:

Nivel de interpretación

Nivel lógico

Nivel epistemológico

Nivel ontológico

Nivel Conceptual

Nivel lingüístico.

Los TM se proponen para modelar redes semánticas. Por ejemplo, si nos centramos en la relación entre un TM y un tesoro vemos que los tesauros no son más que un caso particularmente simple de TM, donde tan solo existen tres *association types* (jerarquía, sinonimia y relacionado). Así, la relación de jerarquía es una *association type* que podría denominarse “es un” o “es una clase de”. Ciertamente al diferir el modelado inicial de TM y tesauros el engarce no es siempre obvio. Por ejemplo, no parece tan evidente si la diferencia entre descriptores y no-descriptores está más ligada a la pareja [subject-topic, base name] o a [topic base name-alternative names]. Otras diferencias con los tesauros están más unidas al momento histórico en que cada propuesta tuvo lugar, así los tesauros son normalmente productos muy incorporados a una única organización (centralizados) y cuya presentación en formato electrónico es solo una característica extra, pero no un requisito. Los TM son productos descentralizados que mejoran con la cooperación de distintas organizaciones y que tienen esencialmente un formato electrónico.

## ANEXOS

### ORGANIZACIONES RELACIONADAS

- ❑ TopicMaps.Org Authoring Group (AG), es un consorcio independiente que trata de aplicar los TM a la Web
- ❑ ISO/IEC JTC 1/SC 34 Information Technology- Document Description and Processing Languages, es el comité de la ISO que desarrolla SGML y otros estándares como HyTime, Topic Maps, DSSSL, etc.
- ❑ Empresas relacionadas con TM <http://index.bonn.iz-soz.de/~sigel/veroeff/ISI-2000/resources/companies.html>

### ENLACES A DOCUMENTACIÓN SOBRE TM

- ❑ <http://www.topicmap.com/> Página muy completa sobre cualquier tema relacionado con TM
- ❑ <http://k42.empolis.co.uk/> Desarrolla software relacionado con TM. Tiene artículos interesantes
- ❑ Topic Maps links. Enlaces a TM <http://www.garshol.priv.no/download/tmlinks.html>
- ❑ [www.topicmaps.org](http://www.topicmaps.org) página oficial sobre TM
- ❑ [www.topicmaps.net](http://www.topicmaps.net) links interesantes relacionados con TM
- ❑ [http://www.ontopia.net/topicmaps/learn\\_more.html](http://www.ontopia.net/topicmaps/learn_more.html)

- ❑ Links de Ontopia la página también es interesante por el software para TM que desarrolla
- ❑ [http://www.sci.sdsu.edu/CRMSE/kfisher\\_knowrep.html](http://www.sci.sdsu.edu/CRMSE/kfisher_knowrep.html)
- ❑ Documentación sobre representación de conocimiento

#### ESTANDARES Y RECOMENDACIONES EN TM

- ❑ Topic Maps ISO 13250 -2002  
[http://www.y12.doe.gov/sgml/sc34/document/0322\\_files/iso13250-2nd-ed-v2.pdf](http://www.y12.doe.gov/sgml/sc34/document/0322_files/iso13250-2nd-ed-v2.pdf)
- ❑ XML Topic Maps (XTM) 1.0 Specification. Steve Pepper, Graham Moore, Steven R. Newcomb, Michel Biezunski, <http://www.topicmaps.org/xtm/>
- ❑ TMQL Draft (Topic Map Query Language). Ann Wrightson, Ontopia, BSI, 7 Nov 2000 (corrected 28 Nov 2000), <http://www.y12.doe.gov/sgml/sc34/document/0186.doc>

#### REFERENCIAS

- [1] Topic Maps ISO 13250 -2002  
[http://www.y12.doe.gov/sgml/sc34/document/0322\\_files/iso13250-2nd-ed-v2.pdf](http://www.y12.doe.gov/sgml/sc34/document/0322_files/iso13250-2nd-ed-v2.pdf)
- [2] Rath, H.H.; Pepper, S.(1999) Topic Maps: Introduction and Allegro, in: Proceedings of Markup Technologies 99 Conference, GCA, Alexandria, VA, 1999.
- [3] Rath, H.H.(2000) Making Topic Maps more colourful.  
<http://www.infoloom.com/gcaconfs/WEB/paris2000/S29-01.HTM>
- [4] Eric Freese. (2000) Using Topic Maps for the representation, management & discovery of knowledge. XML Europe 2000
- [5] Le Grand, Bénédicte and Soto, Michel (2002) Visualisation of the semantic web: Topic Maps Visualisation. Information Visualisation 6th International Conference. 10-12 July 2002
- [6] Steve R. Newcomb. A Perspective on the Quest for Global Knowledge Interchange. Capítulo 3 en Park, J. and Hunting. S. XML Topic Maps. Creating and Using Topic Maps for the web. 2003.