

Tag Ontology for Taguette

Alejandro Adorjan
Universidad ORT Uruguay
Montevideo, Uruguay
adorjan@ort.edu.uy

1 INTRODUCCIÓN

1.1 Presentación Global en Video

<https://youtu.be/hV7zkdVvXjk>

1.2 Código fuente de la ontología

https://github.com/aadorian/onto_2023

1.3 Código fuente de la presente documentación

<https://www.overleaf.com/8518561658kxjypxrmysky#91428a>

1.4 Código fuente del plugin de Protege

<https://github.com/aadorian/plugin/>

1.5 Motivación

Investigadores de áreas sociales frecuentemente en su proceso de análisis utilizan el etiquetado de conceptos. Distintas herramientas actualmente son utilizadas por investigadores que aplican el etiquetado como método de análisis de datos, como por ejemplo Taguette [10]. Si bien la conceptualización de etiquetado está presente en la literatura de ontologías, actualmente no se aplica en dicha aplicación. A su vez, desconocemos que se apliquen en otros CAQDAS (Computer-Assisted Qualitative Data Analysis Software) como ATLAS.ti [1], Dedoose [2] y MAXQDA [11]. El etiquetado es un proceso de categorización de información con el objetivo de describir, categorizar o comentar contenido digital [9].

El etiquetado cumple la función de referencia semántica y, por lo tanto, es una de las actividades más importantes en la Web social [5]. El etiquetado es un medio principal para agregar metadatos a los recursos en el entorno Web 2.0. El principal foco de investigación del artículo [5] es la posibilidad de utilizar Upper Tag Ontology (UTO) con el objetivo de perfilar las características de tres redes sociales de etiquetado diferentes y establecer comportamientos de etiquetado de sus usuarios.

UTO es una propuesta de etiquetado considerando tres sistemas de etiquetado social, Delicious¹, Flickr² y Youtube³, para analizar su comportamiento de etiquetado social [7]. Con base en este análisis, Ding et al. [7] propone la ontología de etiquetas superior (UTO), que se origina a partir de la ontología de etiquetas propuesta por Tom Gruber [8]. Las etiquetas, según Ding et

al.[5] son una forma emergente de metadatos sociales inductivos que se generan por la sinergia de usuarios.

En síntesis, UTO es una ontología de nivel superior para el etiquetado social, que está diseñada para evitar la complejidad y la posible redundancia inherente a los vocabularios de etiquetado generados por los usuarios [5].

En la Tabla 1 se presentan los ocho conceptos que proponen Ding et. al [4] asociados de ‘Tagging’, ‘Tag’, ‘Tagger’, ‘Object’, ‘Source’, ‘Comment’, ‘Date’, ‘Vote’ correspondientes a la ontología Upper Tag Ontology (UTO) For Integrating Social Tagging Data. Formalmente UTO se como :

La definición de ontología O propuesta por Ding et. al. [4] refiere a $C = \{c_i, i \in \mathbb{N}\}$ es un conjunto finito de conceptos, y $R = \{(c_i, c_k), i, k \in \mathbb{N}\}$ un conjunto finito de relaciones establecidas entre los conceptos C . C y R respectivamente se conceptualizan como:

$C = \{ \text{Tag, Tagging, Object, Tagger}$

$\text{Source, Date, Comment, Vote} \}$

$R = \{ \text{hasRelatedTag, hasTag, hasObject, hasSource,}$

$\text{hasDate, hasCreator, hasComment, hasVote} \}$

2 DESCRIPCIÓN DEL DOMINIO

El dominio refiere a los siguientes conceptos: el equipo de investigación, las preguntas de investigación y los artefactos generados durante el proceso (entrevistas autoadministradas). El conjunto de relaciones entre dichos conceptos refieren a las etiquetas de esos artefactos en relación a las preguntas de investigación formuladas por el equipo.

El equipo de investigación es parte del proyecto MENTOR (seMantic Exploration aNd curaTion of Open Hybrid Research) y la temática asociada a los artefactos refiere a las prácticas de lectura de estudiantes de comunicación de la FIC.

La propuesta de desarrollo de la ontología inicialmente gue indagar si era posible la aplicación de ontologías existentes de etiquetado en la aplicación Taguette [10].

3 ROL DE LA ONTOLOGÍA

El sistema en el cual interviene la ontología es un framework que apoya el trabajo de los investigadores que aplican metodologías de investigación cualitativas, el cual facilita el proceso de etiquetado de datos. La implementación en si como demostración esta disponible en el siguiente video <https://www.youtube.com/watch?v=hV7zkdVvXjk>

El principal rol de la ontología refiere a integrar los conceptos de etiquetado con el proceso de investigación. El principal consistió en analizar el modelo de dominio existente, asociarlo con las preguntas e hipótesis de investigación que se realizan y buscar la posibilidad de inferir conceptos asociados.

¹Delicious proveía un servicio de gestión de marcadores sociales en web, los cuales permitían agregar los marcadores en los navegadores y categorizarlos con un sistema de etiquetado denominado folksonomías <https://es.wikipedia.org/wiki/Delicious>.

²Flickr es un sitio web que permite almacenar, ordenar, buscar, vender y compartir fotografías o videos en línea, a través de Internet. <https://es.wikipedia.org/wiki/Flickr>

³Youtube es un sitio dedicado a compartir videos <https://es.wikipedia.org/wiki/YouTube>

Table 1: Tabla de Conceptos de Upper Tag Ontology (UTO) For Integrating Social Tagging Data [4].

Concepto	Sinónimos	Descripción	Tipo de Valor	Instancias
Tagging		El etiquetado es el concepto creado para enlazar otros conceptos. En sí mismo, no tiene ningún significado real.	string	
Tag	Keyword	Una etiqueta es una palabra clave que un Etiquetador asigna a un Objeto.	string	diseño, Web2.0, diseño instruccional, tutoriales, etc.
Tagger	User ID	Un Etiquetador es un ID de usuario para la persona que asigna una etiqueta a un Objeto.	string	sborelli
Source	Social network	Una Fuente es un sitio en línea donde se aloja la relación entre Etiqueta y Objeto.	string	Delicious, Flickr, YouTube, etc.
Date	Time	Una Fecha es la marca de tiempo de la actividad de etiquetado. El formato es "Mmm AA."	date	Jun 07
Object	Online object, Resource	Un Objeto es una entidad a la que un Etiquetador asigna una Etiqueta. Puede ser un marcador (URL), foto, video, etc.	string	www.commoncraft.com
Comment	Note	Un Comentario es una declaración añadida a un Objeto o Etiqueta por un Etiquetador durante la actividad de etiquetado.	string	The CommonCraft Show, Common Craft – Diseño Social para la Web
Vote	Favorite	El etiquetado puede verse como un acto de votación. Un Voto puede ser el número de Taggers diferentes que asignan una Etiqueta a un marcador en Delicious, a una foto indicada como favorita en Flickr o al número de estrellas otorgadas a un vídeo en YouTube.	integer	103 (i.e., 103 taggearon y pusieron el bookmark en Delicious como favoritos)

4 ESPECIFICACIÓN DE REQUERIMIENTOS.

Como investigador cualitativo **quiero** etiquetar los datos de una entrevista *para* poder analizar y reflexionar respecto a las preguntas de investigación que tengo planteadas.

Como investigador cualitativo **quiero** obtener información respecto de como se vinculan las etiquetas en una entrevista *para* relacionarlas con las preguntas de investigación.

Como investigador cualitativo **quiero** ser informado de inconsistencias respecto a preguntas de investigación y etiquetas realizadas *para* poder establecer una consistencia en el proceso de análisis de la investigación.

Caso de Uso Principal

En la Figura 2 presentamos el Modelo Conceptual UTO modificado en el cual estamos realizando el refactorio de UTO. En la extensión proponemos incorporar los conceptos asociados a la investigación como 'ResearchMethod', 'ResearchContext', 'Participant', 'TagCriteria' preservando los otros conceptos propuestos en UTO.

En la Tabla 2 presentamos la integración y comparación de la propuesta de este trabajo de etiquetado en relación a la ontología UTO.

En particular la documentación asociada a la especificación fue creada por un plugin de Protegé llamado Widoco <https://github.com/dgarijo/Widoco>.

5 ESCENARIO PLANTEADO

El escenario planteado es el realización de etiquetado por parte de los investigadores cualitativos, con la idea de poder establecer una observación de los aspectos de análisis mas profundo de su entendimiento. En particular se seleccionó el escenario de un subconjunto de la abstracción de 'etiquetado'.

Donde este etiquetado corresponde con el contexto metodológico. Este conjunto metodológico establece un conjunto de procedimientos, métodos e instrumentos que son conceptualizados en las clases Metodologías, Marco Teórico y Reporte de investigación (en la ontología estan en ingles). Estas conceptualizaciones son disjuntas entre si y por tanto llevadas a la abstracción correspondiente.

Es relevante señalar que el escenario planteado es una simplificación de las características que los investigadores cualitativos realizan al modelar una abstracción y generar en una categorización de elementos una teoría basada en la construcción analítica de los textos.

Table 2: Tabla de Conceptos de Refactoring del Upper Tag Ontology (UTO) For Integrating Social Tagging Data para ser utilizado en la ontología a desarrollar.

Concepto	Sinónimos	Descripción	Tipo de Valor	Instancias
Tagging	Idem UTO	El etiquetado es el concepto creado para enlazar otros conceptos. En sí mismo, no tiene ningún significado real.	string	
Tag	IDEM	Una etiqueta es una palabra clave que un Etiquetador asigna a un Objeto.	string	Prácticas y recursos para estudiar
Tagger	Renombrado con Researcher	El investigador es el responsable de realizar el etiquetado .	string	winocur
Source	Social network	Una Fuente es un sitio en línea donde se aloja el proyecto de etiquetado.	string	http://mentor-me.info/project
Date	Time	Una Fecha es la marca de tiempo de la actividad de etiquetado. TODO: reformatear a un timestamp mas adecuado	date	07 Jun 2023
Object	Online object, Resource	Un Objeto es una entidad a la que un Etiquetador asigna una Etiqueta.	string	http://mentor-me.info/raw/document.doc
Comment		Un Comentario es una declaración añadida a un Etiqueta por un Investigador durante la actividad de etiquetado.	string	Ej: La respuestas a esta pregunta se clasificaron en la Sección Prácticas de lectura, aquí solo se dejó la parte de la respuesta que se relaciona con..
Vote		El etiquetado en un mismo documento puede hacer referencia a varias instancias del tag.Cantidad de Tags en el documento	integer	3 (3 veces se menciona el documento y se referenció al tag correspondiente)
TagCriteria		Criterio establecido en etapas previas de la investigación.	string	Criterio de Codificación Axial, Criterio de Codificación Abierta u otro tipo de Codificación
Participant		Persona participante en el estudio o individuos cuyas características y respuestas fisiológicas y/o conductuales son objeto de estudio en un proyecto de investigación.	string	
ResearchMethod		Método de investigación utilizado	string	entrevista, entrevista autoadministrada
ResearchContext		Contexto general de investigación, donde se establecen las preguntas de investigación, hipótesis y hallazgos	string	RQ1 ¿Cuáles son los recursos y estrategias que utilizan los estudiantes avanzados de comunicación para organizar el estudio de forma individual y/o grupal? Objetivos Explorar los diversos modos de leer y las estrategias de estudio de los estudiantes de comunicación del último ciclo de la carrera Hallazgos Los recursos, habilidades y competencias digitales no suponen el reemplazo de los recursos tradicionales para leer y estudiar.

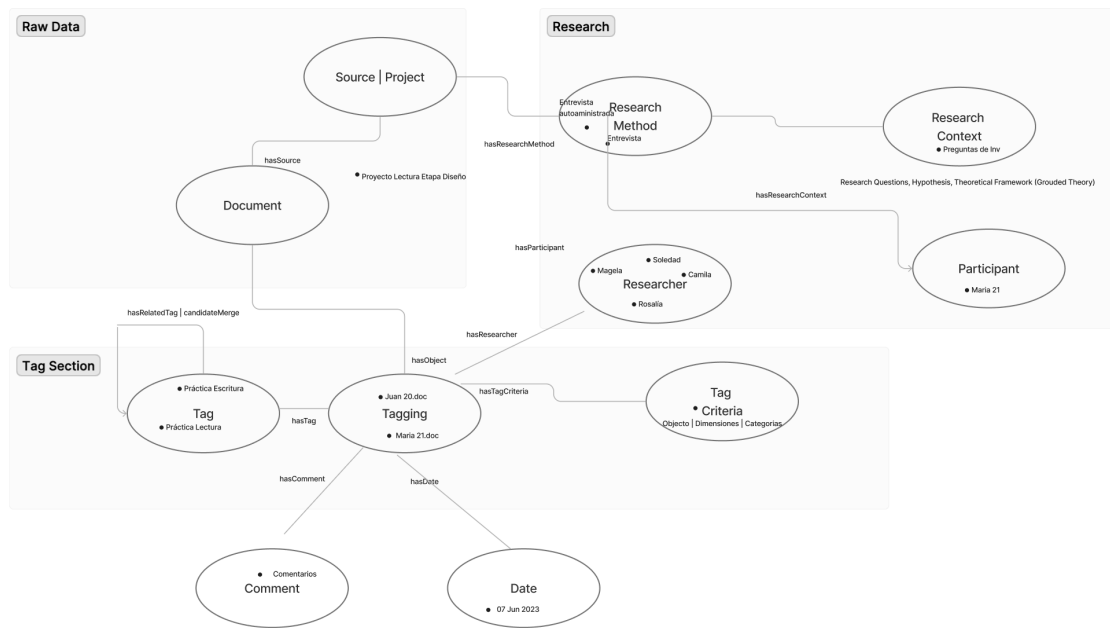


Figure 1: Refactoreo de este trabajo de refactorio/extensión de Upper Tag Ontology (UTO) For Integrating Social Tagging Data

Los principales supuestos de este trabajo es que es posible ‘objetivizar’ un concepto complejo de la realidad y por medio de una representación de lógica establecer un orden estructural.

La propuesta de este trabajo consistió en refactoriar la ontología Upper Tag Ontology for Integrating Social Tagging Data [4] con el objetivo de ser aplicada en Tagguette [10] y extender el concepto de etiquetado a un contexto de investigación cualitativa.

Lección Aprendida

La principal lección aprendida de este trabajo fue el hecho que la generación de una estructura lógica de TBOX permitió ir desarrollando un diseño en tiempo real con el razonador en ejecución. Este punto a nivel de diseño de software conceptualmente es relevante por la abstracción en sí. La creación de un plugin considero que me permitió entender en detalle la API que permite identificar las principales operaciones de abstracción de un modelo de dominio y establecer las reglas de conceptualización del modelo, tanto a nivel conceptual como la semántica en sí. A su vez, entrar en mayor detalle de profundidad en el trabajo de campo del Doctorado.

6 DESCRIPCIÓN DE BUENAS PRÁCTICAS Y VERIFICACIONES DE CALIDAD

La verificación de buenas prácticas en este trabajo si bien siguieron un "estandar" fueron escasas dado el conocimiento inicial de la temática y la complejidad de la misma. Considero que la ‘calidad’ en sí fue la adecuación al uso en el cual se construyo desde un modelo existente "UTO" que y la instanciación de la práctica.

En las instancias del modelo se utilizaron ejemplos aplicados en un trabajo experimental de campo que se corresponde con el proyecto MENTOR.

La documentación fue generada por Wicoco <https://github.com/dgarijo/Widoco/releases> Esta disponible en <https://github.com/aadorian/Ontospy>

UTO	FOAF
Tagger	= Person ⊆ Agent
Object	= Document ⊆ Image
hasTag	= depiction = topic
hasCreator	= maker

Table 3: UTO alineamiento con FOAF

6.1 Código Fuente Proyecto y Ontología

https://github.com/aadorian/onto_2023.git

Nota: Se utilizan las definiciones provistas de conceptos de métodos cualitativos del siguiente Glosario de Términos ⁴

Se utilizó Protégé [12] como herramienta de diseño y se ejecutaron los razonadores integrados para verificar la consistencia del modelo. Como prueba de concepto prototipamos un plugin en Protégé [12] que si bien no integró el concepto de “tagging” (por detalles de tiempos) permite visualizar en forma global la API de OWL2.

7 TRABAJO A FUTURO

Como trabajo a futuro quedó pendiente la utilización de la API de OpenAI | ChatGPT para la generación de IA de la notación en Manchester por ejemplo y documentación en conjunto con la descripción de la ontología generada como documentación.

8 CONCLUSIÓN

El presente trabajo permitió integrar aspectos relevantes en el trabajo global de investigación dado que la estructura y conceptualización a nivel técnico permite realizar un modelo de diseño

⁴<https://files.eric.ed.gov/fulltext/ED611000.pdf>

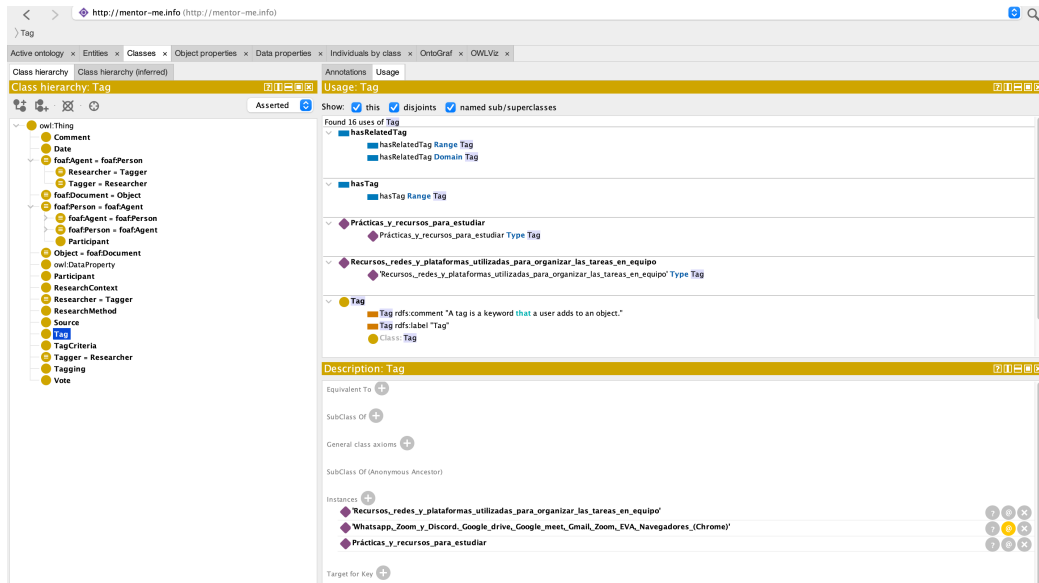
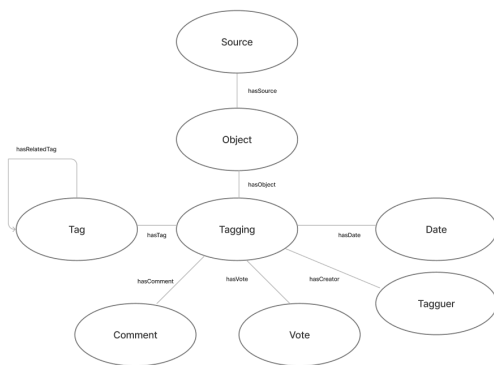


Figure 2: Ejemplo aplicado de avance Protege.



UTO- Upper Tag Ontology (UTO)

Figure 3: Diseño del UTO propuesto por Ding et al .Upper Tag Ontology (UTO) For Integrating Social Tagging Data [4].

en conjunto con un “razonador” que sigue reglas establecidas deductivamente, en relación a la abstracción que se desea modelar. A nivel de ejecución, consideramos que es relevante tener este aspecto de “copiloto” de diseño conceptual que permita corregir e inferir las reglas de los modelos. A su vez encontramos que a nivel de metodologías cualitativas es un punto interesante la “categorización” de la abstracción perse.

9 ANEXO

9.1 Representación presentada en la propuesta de UTO

En la siguiente tabla se describen las relaciones entre los conceptos de la Tabla 1.

9.2 Caso de uso Mentor

En la presente figura se conceptualiza los documentos utilizados en una análisis de categorías y el mapeo correspondiente con los conceptos de la ontología refactorizada

El código fuente de avance de la ontología esta disponible en https://github.com/aadorian/onto_2023.git

Comando de Generación de Documentación `java -jar /Users/user/Desktop/widoco-1.4.20-jar-with-dependencies_JDK-11-ontFile.jar /Users/user/Desktop/onto_2023.owl -getOntologyMetadata -includeImportedOntologies -displayDirectImportsOnly -uniteSections -outFolder /Users/user/Desktop`

REFERENCES

- [1] ATLAS.ti. last accessed July 2023. ATLAS.ti. <https://atlasti.com>.
- [2] Dedoose. last accessed July 2023. Dedoose. <https://www.dedoose.com/>.
- [3] Ying Ding, E Jacob, James Caverlee, Michael Fried, and Zhixiong Zhang. 2009. Profiling social networks. *D-Lib Magazine* 15, 3/4 (2009), 1082–9873.
- [4] Ying Ding, Elin K Jacob, Michael Fried, Ioan Toma, Erjia Yan, Schubert Foo, and Staša Milojević. 2010. Upper tag ontology for integrating social tagging data. *Journal of the American Society for Information Science and Technology* 61, 3 (2010), 505–521.
- [5] Ying Ding, Elin K Jacob, Zhixiong Zhang, Schubert Foo, Erjia Yan, Nicolas L George, and Lijiang Guo. 2009. Perspectives on social tagging. *Journal of the American Society for Information Science and Technology* 60, 12 (2009), 2388–2401.
- [6] Ying Ding, Ioan Toma, Michael Fried, Sin-Jae Kang, and Zhixian Yan. 2008. Integrating social tagging data: upper tag ontology (UTO). In *2008 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, 460–466.
- [7] Ying Ding, Ioan Toma, S Kang, Michael Fried, and Zhixian Yan. 2008. Data mediation and interoperation in social web: Modeling, crawling and integrating social tagging data. In *WWW2008 Workshop on Social Web Search and Mining (SWSM 2008)*, April. 21–25.
- [8] Thomas Gruber. 2007. Ontology of folksonomy: A mash-up of apples and oranges. *International Journal on Semantic Web and Information Systems (IJSWIS)* 3, 1 (2007), 1–11.
- [9] Pauline Rafferty. 2018. Tagging: Reviews of concepts in Knowledge Organization. (2018).
- [10] Rémi Rampin and Vicky Rampin. 2021. Taguette: open-source qualitative data analysis. *Journal of Open Source Software* 6, 68 (2021), 3522.
- [11] VERBI Software. last accessed July 2023. MAXQDA. <http://maxqda.com>.
- [12] Stanford University. Accessed on July 2023. Protege. <https://protege.stanford.edu>.

Table 4: Resumen de características de etiquetado social en Delicious, Flickr y YouTube [3].

Características	Delicious	Flickr	YouTube
Comunidad	Personas interesadas en compartir marcadores sobre la web, programación, etc.	Fotógrafos profesionales y no profesionales interesados en compartir fotografías	Personas interesadas en compartir videos sobre cualquier tema.
Temas Principales	Recursos relacionados con TI, video, música y noticias	Características de las fotografías, incluyendo colores, ubicaciones, años, estaciones	Géneros (música, humor) y respuestas afectivas (sexy, caliente, nuevo)
Comportamiento de Etiquetado	Actividad clave con muchos usuarios participando en actividades de etiquetado	Pocos usuarios participan en el etiquetado	Pocos usuarios participan en el etiquetado
Dinámica	Sí	No	No
Temas Emergentes (Tendencias)	Diseño web, tutoriales, web 2.0	No aparentes	No aparentes
Temas en Declive	Internet, investigación, XML, seguridad, Java	No aparentes	No aparentes
Estabilidad	Conjunto de etiquetas más populares cambia un 10-20% cada año en promedio	El conjunto de etiquetas cambia, pero algunas categorías de etiquetas (color, ubicación, año) son relativamente estables	El conjunto de etiquetas cambia, pero las categorías de etiquetas (género, reacción afectiva) son relativamente estables
Vocabulario de Etiquetado	Relativamente estable	Categorías estables de etiquetas	Categorías estables de etiquetas
Enfoque de Etiquetado	Contenido de los recursos marcados	Características de las fotografías	Géneros, respuestas afectivas a los videos
Propósito Principal del Etiquetado	Acceder y compartir	Almacenar y recuperar	Describir y calificar



Figure 4: Ejemplo de visualización de instancia por Ding et al.[4]

Relación	Dominio	Rango	Cardinalidad	Tipo OWL	Propiedades	Relación Inversa
<i>hasTag</i>	Object	Tag	1	Object Property	Transitive, Symmetric	<i>isTagOf</i>
<i>hasRelatedTag</i>	Tag	Tag	-	Object Property	Transitive	-
<i>hasCreator</i>	Tagging	Tagger	1	Object Property	-	<i>isCreatorOf</i>
<i>hasObject</i>	Tagging	Object	1	Object Property	Transitive	<i>isObjectOf</i>
<i>hasDate</i>	Tagging	Date	1	Object Property	-	-
<i>hasSource</i>	Tagging	Source	1	Object Property	-	<i>isSourceOf</i>
<i>hasComment</i>	Tagging	Comment	1	Object Property	-	<i>isCommentOf</i>
<i>hasVote</i>	Comment	Vote	1	Object Property	-	<i>isVoteOf</i>

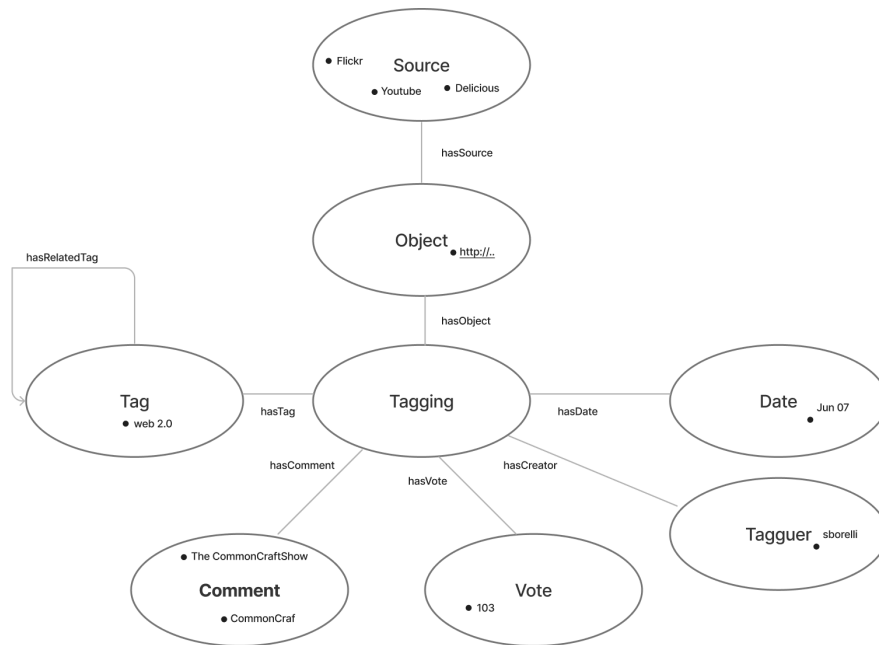


Figure 5: Ejemplo de instancia presentado por Ding et al.[4].

Table 5: Ejemplo Tagging en RDF presentado en el artículo Integrating social tagging data: upper tag ontology (UTO) [6]

Subject	Predicate	Object
http://uto.deri.at/b47145e3-5417-4d6e-885e-43469bc18ca4	http://uto.deri.at/has_tag	http://del.icio.us/tag/web2.0
http://uto.deri.at/b47145e3-5417-4d6e-885e-43469bc18ca4	http://uto.deri.at/has_tag	http://del.icio.us/tag/semanticweb
http://uto.deri.at/b47145e3-5417-4d6e-885e-43469bc18ca4	http://uto.deri.at/has_tag	http://del.icio.us/tag/deri
http://uto.deri.at/b47145e3-5417-4d6e-885e-43469bc18ca4	http://uto.deri.at/has_tag	http://del.icio.us/tag/innsbruck
http://uto.deri.at/b47145e3-5417-4d6e-885e-43469bc18ca4	http://uto.deri.at/has_comment	Deri home
http://uto.deri.at/b47145e3-5417-4d6e-885e-43469bc18ca4	http://uto.deri.at/has_vote	2468
http://uto.deri.at/b47145e3-5417-4d6e-885e-43469bc18ca4	http://uto.deri.at/has_tagger	MichaelFried
http://uto.deri.at/b47145e3-5417-4d6e-885e-43469bc18ca4	http://uto.deri.at/has_date	Jul 07
http://uto.deri.at/b47145e3-5417-4d6e-885e-43469bc18ca4	http://uto.deri.at/has_object	http://www.deri.org
http://del.icio.us/tag/web2.0	http://uto.deri.at/has_related_tag	http://del.icio.us/tag/tools
http://del.icio.us/tag/web2.0	http://uto.deri.at/has_related_tag	http://del.icio.us/tag/blog
http://del.icio.us/tag/web2.0	http://uto.deri.at/has_related_tag	http://del.icio.us/tag/design
http://del.icio.us/tag/web2.0	http://uto.deri.at/has_related_tag	http://del.icio.us/tag/community
http://www.deri.org	http://uto.deri.at/has_source	http://del.icio.us

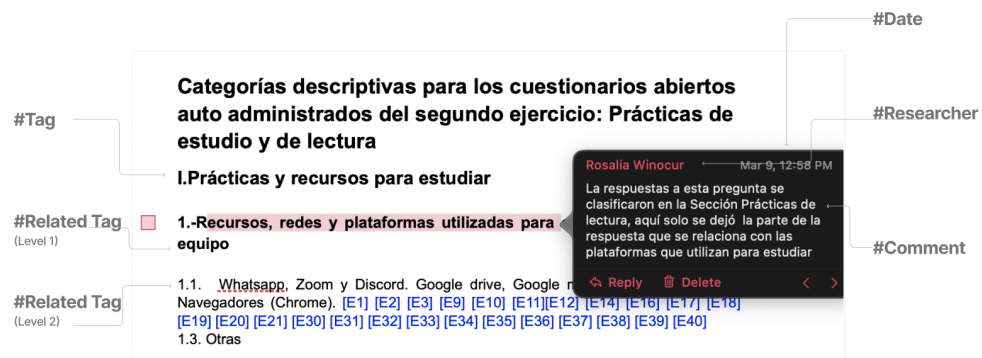


Figure 6: Caso de estudio MENTOR: Ejemplo de instancia.