

TermiGraph

Terminología en RDF

DOCUMENTACIÓN TERMIGRAPH: CONVERSOR DE EXCEL-UNTERM

23 de abril de 2026

TerésIA



Índice

1. Introducción	2
2. Estructura del recurso	2
3. TermiGraph paso a paso	4
3.1. Paso 1: Cargar el recurso	4
3.2. Paso 2: Indicar los metadatos del recurso	5
3.3. Paso 3: Descarga del archivo	6
4. Terminología en RDF con Ontolex-lemon	6
4.1. Web Semántica y RDF	7
4.2. Datos terminológicos y el modelo Ontolex-lemon	7
4.3. Modelo de transformación de TeresIA	9
A. Ejemplo de datos transformados	I

1. INTRODUCCIÓN

El presente documento describe el uso y funcionamiento de TermiGraph¹, un conversor de terminologías a RDF desarrollado en el marco del proyecto nacional TeresIA², cuyo objetivo es proporcionar un punto centralizado de acceso a terminologías.

El contenido se centra específicamente en la conversión de terminologías en hojas de cálculo de Excel que siguen el esquema de UNTERM. En la sección 2 se detalla cómo deben estructurarse los datos de entrada que el usuario desea transformar. En la sección 3 se ofrece una guía paso a paso para utilizar la interfaz de conversión. Finalmente, la sección 4 introduce brevemente el concepto de datos enlazados y explica cómo se representan en RDF los datos originales tras la transformación.

2. ESTRUCTURA DEL RECURSO

Este conversor surge a raíz de la colaboración entre el proyecto TeresIA y la base de datos terminológica de Naciones Unidas, UNTERM. Este conversor requiere como entrada un archivo Excel (.xlsx). En cuanto a la estructura de los datos, debe tenerse en cuenta lo siguiente: Cada concepto, junto con sus términos asociados, debe figurar en la misma fila. En la primera columna debe incluirse el identificador del concepto, por ejemplo, concepto_1. A continuación deben dejarse dos columnas vacías, que podrán emplearse para registrar notas internas que no serán transformadas. En la quinta columna comienza la zona destinada a la transformación de los datos terminológicos. Esta sección está organizada por idioma, siguiendo bloques estructurales de 15 columnas, cuyo orden debe respetarse estrictamente:

1. **Términos preferentes:** Esta columna debe llevar como encabezado el nombre del idioma al que corresponden los términos preferentes. El idioma puede indicarse mediante su denominación en español (por ejemplo, español), en inglés (Spanish) o a través de sus códigos normalizados ISO 639-1 (es), ISO 639-2 (spa) o ISO 639-3 (spa). También pueden utilizarse denominaciones en otras lenguas —como euskera, catalán o gallego— siempre que estén registradas en Lexvo³. Esta columna define el idioma de referencia del bloque de 15 columnas. En cuanto al contenido, pueden registrarse varios términos preferentes, separados mediante punto y coma (;).
2. **Fuente del término preferente:** En esta columna se indican las fuentes de procedencia de los términos preferentes, ya se trate de referencias bibliográficas o de enlaces. Si se incluyen varios términos, las fuentes deben aparecer en el mismo orden y separadas con punto y coma (;), de modo que cada fuente se corresponda con el término situado en la misma posición.

¹<https://termigraph.teresia.linkeddata.es>

²<https://proyectoteresia.org/>

³<http://www.lexvo.org/>

3. **Notas del término preferente:** Aquí pueden añadirse observaciones, aclaraciones o comentarios relativos a los términos preferentes. Si existen varios términos, las notas deben separarse mediante punto y coma (;), manteniendo la correspondencia en el mismo orden.
4. **Contextos del término preferente:** En esta columna pueden incorporarse ejemplos de uso que ilustren el empleo de los términos preferentes en diferentes contextos lingüísticos o discursivos. Si se registran varios términos, los ejemplos deben separarse mediante punto y coma (;), respetando el orden secuencial para asegurar la correspondencia término-contexto.
5. **Términos alternativos:** En esta columna pueden consignarse variantes terminológicas o términos alternativos del mismo idioma. Pueden registrarse varios, separándolos mediante punto y coma (;).
6. **Fuente de los términos alternativos:** Se indican aquí las fuentes de los términos alternativos, ya sean referencias bibliográficas o enlaces. Si se listan varios términos alternativos, las fuentes deben aparecer en el mismo orden y separadas con punto y coma (;), garantizando la correspondencia con cada término.
7. **Notas de los términos alternativos:** En esta columna pueden añadirse observaciones, aclaraciones o comentarios respecto a los términos alternativos. Si se incluyen varios, las notas deben separarse mediante punto y coma (;), manteniendo la correspondencia posicional.
8. **Contextos de los términos alternativos:** Aquí se registran ejemplos de uso de los términos alternativos en distintos contextos lingüísticos o discursivos. Cuando existan varios términos, los ejemplos deben separarse con punto y coma (;), siguiendo el orden establecido.
9. **Acrónimos y siglas:** En esta columna pueden registrarse las siglas o acrónimos empleados en el idioma correspondiente. Si existen varios elementos, deben separarse mediante punto y coma (;).
10. **Fuente de los acrónimos:** Se indican las fuentes de procedencia de cada acrónimo o sigla, ya sean referencias bibliográficas o enlaces. Si se incluyen varios, las fuentes deben separarse con punto y coma (;), respetando la correspondencia posicional con los acrónimos listados.
11. **Notas de los acrónimos y siglas:** En esta columna pueden añadirse observaciones o comentarios sobre los acrónimos y siglas. Si se registran varios, las notas deben separarse con punto y coma (;), manteniendo el mismo orden de referencia.
12. **Contextos de los acrónimos y siglas:** Aquí pueden recogerse ejemplos de uso de los acrónimos y las siglas en diferentes contextos discursivos. Si se listan varios, los ejemplos deben separarse mediante punto y coma (;), conservando la correspondencia con los elementos previos.
13. **Validez por idioma:** Indica si la entrada terminológica correspondiente a ese idioma ha sido validada o no.

- 14. **Definición:** Definición del concepto en el idioma correspondiente. Esta definición es aplicable a todos los términos asociados en dicho bloque: preferentes, alternativos y acrónimos/siglas.
- 15. **Nota conceptual:** Nota relativa al concepto en el idioma correspondiente. Abarca todos los términos (preferentes, alternativos y acrónimos/siglas) incluidos en el bloque.

Si no se dispone de datos para un campo determinado (por ejemplo, la fuente de un término), la columna debe dejarse vacía, pero nunca eliminarse. De lo contrario, la configuración del bloque quedará desalineada y los datos posteriores se transformarán incorrectamente.

Consultar el anexo para ver un ejemplo completo o descargar la plantilla correspondiente.

3. TERMIGRAPH PASO A PASO

A continuación, se explica paso a paso cómo convertir una terminología en formato Excel-UNTERM con TermiGraph.

3.1. Paso 1: Cargar el recurso

Paso 1: CARGA DEL RECURSO

Tipo de archivo: *

-- Selecciona un tipo --

Archivo: *

Examinar...

No se ha seleccionado ningún archivo.

Limpiar

Continuar

* Obligatorio

Figura 1: TermiGraph paso 1 (carga del recurso)

Tal y como se muestra en la Figura 1, en la primera pantalla del conversor se ha de cargar el recurso que se quiera convertir, indicando su tipo.

Para ello, primero se debe indicar el *Tipo de archivo*. Esta selección condicionará el conversor que se quiere utilizar. En el caso de los glosarios monolingües, se debe seleccionar ‘**Glosario monolingüe (lista)**’.

A continuación, en *Archivo*, se debe cargar **cargar el fichero** a convertir. Para ello, pulse ‘Examinar’. Es importante tener en cuenta que dicho fichero debe estar en formato de texto plano,

es decir, debe tener la **extensión .txt**. Asimismo, tal y como se explica en la Sección 2, **cada línea** debe contener **un sólo término, sin signos de puntuación** (salvo que sean parte del término).

Una vez se haya cumplimentado la primera página, pulse ‘Continuar’. En caso de querer borrar los datos, puede pulsar ‘Limpiar’.

3.2. Paso 2: Indicar los metadatos del recurso

Paso 2: METADATOS DEL RECURSO

Título del recurso *

Dominio (EuroVoc) *

Buscar...

DERECHO

derechos y libertades

derechos políticos

derechos sociales

derechos fundamentales

lucha contra la discriminación

Autor del recurso *

Nota: Para añadir múltiples autores, separar mediante punto y coma (;). Por ejemplo: Marta López; Juanito García

Enlace al recurso original

Limpiar

Convertir

* Obligatorio

Figura 2: TermiGraph paso 2 (metadatos del recurso)

A través de la segunda página del conversor se registran los metadatos del glosario (ver Figura 2).

En primer lugar, es obligatorio introducir el nombre del recurso en el campo *Título del recurso*.

A continuación, en el campo *Dominio (EuroVoc)*, se deben seleccionar uno o varios dominios temáticos relacionados con el glosario. Dado que TermiGraph ha sido desarrollada en coherencia con los demás servicios de TeresIA, las opciones disponibles se han limitado para facilitar la posterior desambiguación de términos. En particular, se emplea el esquema establecido por

EuroVoc⁴, el tesoro multidisciplinario de la Unión Europea.

En el campo *Autor del recurso*, de carácter obligatorio, se debe indicar el autor o los autores del glosario. Tal como se señala en la nota, **si existen varios autores, deben separarse mediante punto y coma (;)**. Por ejemplo: Marcela Sánchez Giménez; Pedro López Rosas. Cabe destacar que, al convertirse los datos en un grafo interconectado, no se conserva ningún orden jerárquico entre los autores.

De forma opcional, puede añadirse un enlace al recurso original en el campo *Enlace al recurso original*. Aunque no es obligatorio, se recomienda completarlo para aumentar la visibilidad del glosario, especialmente si se prevé su incorporación al portal de TeresIA.

Una vez completados todos los metadatos, se puede proceder a la conversión del recurso pulsando el botón *Convertir*. Si se desea eliminar la información introducida, puede utilizarse la opción *Limpiar*.

3.3. Paso 3: Descarga del archivo

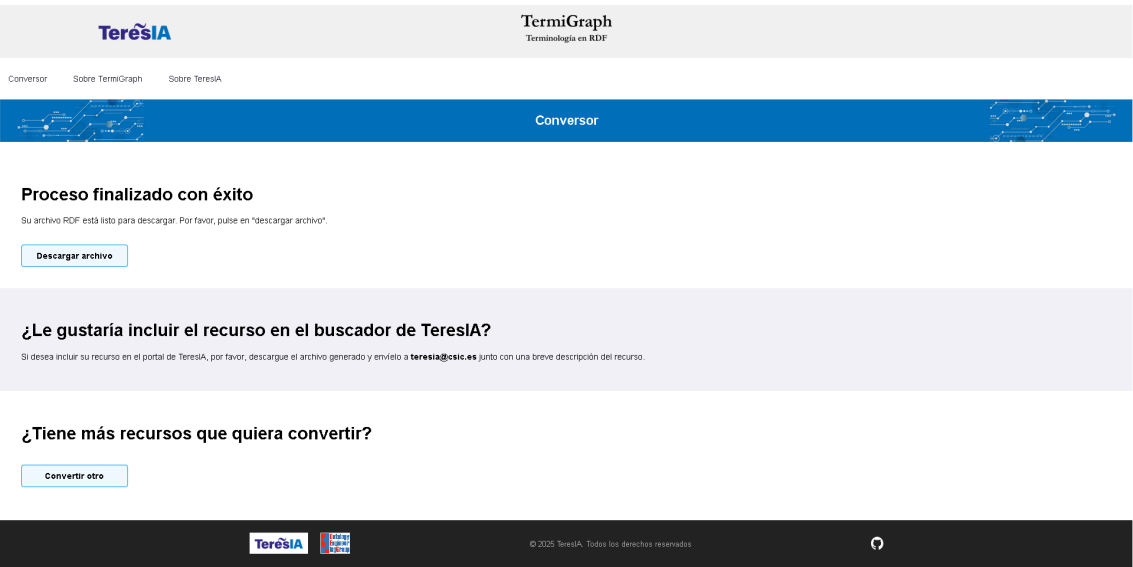


Figura 3: TermiGraph paso 3 (descarga)

Mientras el glosario se convierte, un mensaje de espera aparecerá en pantalla. El conversor puede tardar varios minutos en convertir el recurso, especialmente si el recurso es grande.

Finalmente, cuando el proceso de conversión se completa correctamente, el archivo resultante puede descargarse pulsando en *Descargar archivo*.

En caso de querer añadir su recurso al portal de recursos de TeresIA, puede enviar el recurso convertido al correo electrónico que se ofrece como contacto. Asimismo, se recomienda incluir en el mensaje una breve descripción del recurso, para los usuarios del portal. Se recuerda que, debido a la naturaleza del proyecto, **solo se aceptarán recursos que incluyan alguno de los idiomas oficiales de España**.

En caso de querer convertir un nuevo recurso, es posible volver a la primera página del conversor pulsando en *Convertir otro*.

⁴<https://eur-lex.europa.eu/browse/eurovoc.html?locale=es>

4. TERMINOLOGÍA EN RDF CON ONTOLEX-LEMON

Esta sección tiene como objetivo explicar la estructura del archivo RDF generado. Para ello, se inicia con una breve introducción a la Web Semántica, al estándar RDF y a varios conceptos fundamentales que facilitan su comprensión. A continuación, se presenta el modelo Ontolex-lemon, el modelo de referencia para la representación de datos lingüísticos y la base empleada en TeresIA para la transformación de los términos. Finalmente, se describe la estructura del RDF resultante, detallando cómo se representan los términos originales en el modelo transformado.

4.1. Web Semántica y RDF

El RDF (Resource Description Framework)⁵ es un formato estándar propuesto por World Wide Web Consortium (W3C) y sirve como base de los datos enlazados. Este formato organiza la información en tripletas compuestas por un sujeto, un predicado y un objeto; y el conjunto de tripletas genera un grafo. El sujeto, que es el primer elemento de la triplete, representa el recurso sobre el cual se proporciona información y se identifica mediante un Identificador Uniforme de Recursos (URI, por sus siglas en inglés). El predicado, que ocupa la segunda posición, describe la relación o propiedad que conecta al sujeto con el objeto, siendo también representado por un URI. Por último, el objeto, tercer elemento de la triplete, define el valor asociado a la propiedad del sujeto, pudiendo ser un literal (como texto) o un URI.(Chaves-Fraga et al., 2022)

Para estandarizar la representación de datos, se utilizan las ontologías, que son modelos o esquemas semánticos diseñados para describir formalmente un dominio. Las ontologías incluyen diversos componentes, como clases, propiedades e instancias. Las clases definen categorías o tipos de elementos dentro de un ámbito de conocimiento. Las propiedades representan las relaciones o atributos que vinculan clases o instancias entre sí. Por último, las instancias son ejemplos concretos o específicos que forman parte de una clase.

1. Sujeto: es aquello de lo que se habla, el recurso principal de la afirmación.
2. Predicado: es la propiedad o relación que conecta al sujeto con algo más.
3. Objeto: es el valor o recurso que completa el significado de la relación establecida por el predicado.

4.2. Datos terminológicos y el modelo Ontolex-lemon

Aunque no hay una sola forma de representar los datos lingüísticos en RDF, Ontolex-lemon (en adelante, Ontolex) (McCrae et al., 2017) se ha convertido en la ontología o modelo más utilizado para representar recursos lexicográficos y terminológicos en RDF (Cimiano et al. 2011, Di Buono et al. 2020). Esta ontología cuenta con cuatro clases principales, tal y como se puede observar en la Figura 4: Concepto Léxico (*Lexical Concept*), Sentido Léxico (*Lexical Sense*), Entrada Léxica (*Lexical Entry*) y Forma (*Form*). La clase Entrada Léxica se utiliza para representar una palabra, frase o unidad léxica en una lengua determinada. Las diferentes realizaciones morfológicas de la entrada se expresan a través de la clase Forma. En cuanto al Sentido Léxico, se utiliza para

⁵<https://www.w3.org/RDF/>

proporcionar una conexión semántica entre la Entrada Léxica y un concepto de la ontología. Por último, un Concepto Léxico representa un concepto o idea abstracta que unifica significados entre sentidos léxicos y lenguas.

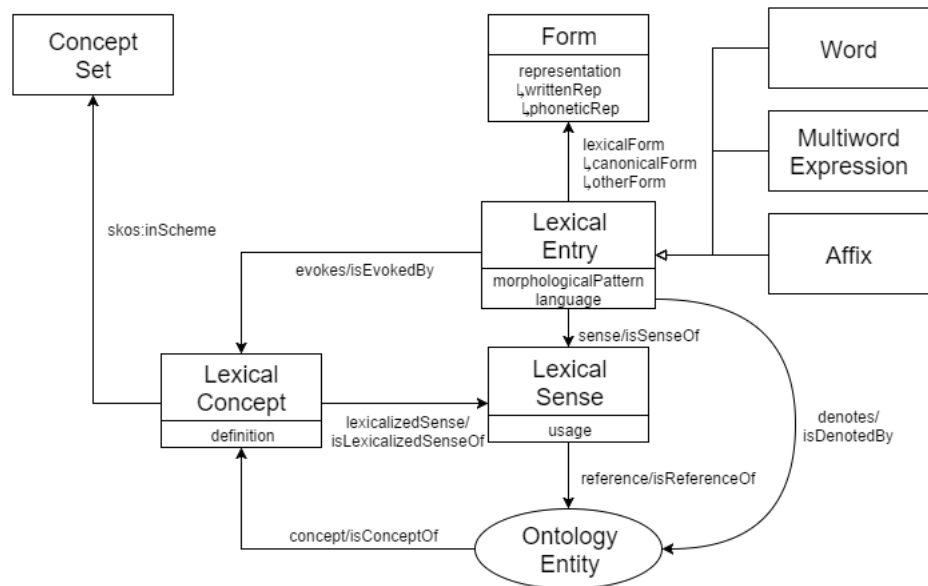


Figura 4: Diagrama del núcleo de Ontolex

Aunque Ontolex cubre la mayoría de las necesidades de representación que necesitan los glosarios monolingües, se utilizan otras ontologías para la representación de algunos aspectos como los idiomas o los datos de autoría. En concreto, se usan:

1. **Simple Knowledge Organization System (SKOS)**⁶: un estándar del W3C, utilizado habitualmente para representar datos jerárquicos en tesauros, sistemas de clasificación u otros tipos de sistemas de organización.
2. **Lexvo**⁷: Lexvo.org (de Melo, 2015) es una plataforma de datos enlazados que proporciona identificadores únicos y descripciones semánticas de lenguas y variantes lingüísticas. Su objetivo es vincular información lingüística de diferentes fuentes y ofrecerla en múltiples idiomas para apoyar la interoperabilidad y la reutilización de datos en la Web Semántica.
3. **Meta-Share**⁸: ontología que describe recursos lingüísticos y tecnologías de procesamiento del lenguaje, incluyendo sus propiedades, tipos, relaciones y condiciones de uso.
4. **FOAF (Friend of a Friend)**⁹: ontología usada para describir a las personas, sus actividades y sus relaciones.
5. **DCMI Metadata Terms**¹⁰: vocabulario muy extendido para describir metadatos sobre recursos, como documentos, imágenes, conjuntos de datos o cualquier otro tipo de entidad digital o física. Por ejemplo, permite representar datos como el título o el idioma de un elemento. Esta ontología también se conoce como Dublin Core Terms (DCTerms).
6. **RDF Schema (RDFS)**¹¹: una extensión de vocabulario de RDF que proporciona mecanismos para describir la estructura, la semántica y las relaciones de los datos. Este vocabulario, en

⁶<https://www.w3.org/2009/08/skos-reference/skos.html>

⁷<http://www.lexvo.org/>

⁸<http://w3id.org/meta-share/meta-share/2.0.0>

⁹<https://xmlns.com/foaf/spec/>

¹⁰<http://purl.org/dc/terms/>

¹¹<http://www.w3.org/2000/01/rdf-schema#>

Léxico con el Sentido Léxico del término relacionado. Por otro lado, el Sentido Léxico permite también representar la normatividad de un término. Este se expresa a través de la propiedad `lexinfo:normativeAuthorization`, que apunta a una de las instancias predefinidas de la clase `lexinfo:NormativeAuthorization`. En concreto, los términos consignados en la columna de términos preferentes (véase la Sección 2) se indican mediante la instancia `lexinfo:preferredTerm`. Finalmente, el Sentido Léxico mantiene el vínculo con la fuente original del término mediante la propiedad `dct:source`.

Asimismo, se crea un **Concepto Léxico** (`ontolex:LexicalConcept`) por cada término. Habitualmente, esta clase alberga definiciones, notas u otra información conceptual, pero en los glosarios monolingües no se dispone de estos datos, por lo que la clase queda sin contenido terminológico adicional. Al igual que el resto de clases de Ontolex, el Concepto Léxico enlaza también con el recurso original a través de `dct:source`.

Todos los Conceptos Léxicos se agrupan en un **Conjunto de Conceptos** (`ontolex:ConceptSet`) mediante la propiedad `skos:inScheme`. A través de este conjunto se accede al Recurso Léxico-Conceptual (`ms:LexicalConceptualResource`), donde se almacenan los metadatos del recurso.

Los metadatos se describen conforme a la ontología Meta-Share. Para ello, se genera un **Recurso Lexico-Conceptual** (`ms:LexicalConceptualResource`). El nombre del recurso se indica mediante `ms:resourceName`, y el idioma mediante `ms:language`, siguiendo el mismo criterio de estandarización con Lexvo mencionado anteriormente. Como ya se ha mencionado previamente, toda la información representada mediante Ontolex (Forma, Entrada Léxica, Sentido Léxico y Concepto Léxico) enlaza con este recurso para garantizar la trazabilidad de los datos originales en el grafo.

El **dominio temático** del recurso se especifica mediante `ms:domain`, que enlaza con una instancia de EuroVoc (`skos:Concept`). Además, la propiedad `ms:resourceCreator` vincula el recurso con su autor, representado como una instancia de `ms:Actor`, cuyo nombre se indica mediante `rdfs:label`. Finalmente, el enlace al recurso original proporcionado por el usuario se representa mediante `dct:source`.

A. EJEMPLO DE DATOS TRANSFORMADOS

A continuación, se ilustra el proceso de transformación de un recurso. Por un lado, las Figuras 6 y 7 muestran, respectivamente, la carga en la interfaz de TermiGraph de la terminología que se desea convertir y la posterior cumplimentación de los metadatos del recurso.

Paso 1: CARGA DEL RECURSO

Tipo de archivo: *

EXCEL (UNTERM)

Archivo: *

Examinar...

ejemplo_excel_unterm.xlsx

Limpiar

Continuar

* Obligatorio

Paso 2: METADATOS DEL RECURSO

Título del recurso *

Ejemplo animales

Dominio (EuroVoc) *

ciencias

derechos políticos

derechos sociales

derechos fundamentales

lucha contra la discriminación

deberes del ciudadano

derechos económicos

derechos humanos

[100224] ciencias naturales y aplicadas ✖

Autor del recurso *

Nota: Para añadir múltiples autores, separar mediante punto y coma (;). Por ejemplo: Marta López; Juanito García

Autor1 Apellido1; Autor2 Apellido2

Enlace al recurso original

https://oeg.fi.upm.es/

Limpiar

Convertir

* Obligatorio

Figura 6: Carga de recurso en TermiGraph

Figura 7: Metadatos en TermiGraph

Por otro lado, las Figuras 8, 9 y 10 muestran el contenido del archivo cargado en la Figura 6.

	A	B	C	D
1	ID Concepto	Notas a ignorar	Estado de la ficha	Dominio
2	concept_1	Notas internas	completa	Animales
3	concept_2	Notas internas	ok	Animales; naturaleza

Figura 8: Ejemplo de Excel formato UNTerm, columnas A-D

	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	Español	Fuente término español	Nota término español	Contexto término español	Alternativo Español	Fuente término español alt	Nota término español alt	Contexto término español	Acronímico español	Fuente acronímico	Notas acronímico	Contexto acronímico	Validas español	Definición español	Notas español (concepto)
2	Gato	Diccionario 1		Me gusta tu gato	Felis catus		Término científico		GT;G	; Un libro					Concepto histórico
3	Perrito		Término coloquial		Canis familiaris; Canis lupus familiaris	; Revista A	Término científico; Otro término científico						validado	Mamífero que se suele considerar animal de compañía	

Figura 9: Ejemplo de Excel formato UNTerm, columnas E-S

1

T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
Inglés	Fuente término	Nota término	Contexto término	Alternativo inglés	Fuentes término	Nota término	Contexto término	Acrónimos inglés	Fuente acrónimo	Nota acrónimo	Contexto acrónimo	Valides inglés	Definición inglés	Notas inglés (concepto)
Cat				House Cat										
Dog														

Figura 10: Ejemplo de Excel formato UNTerm, columnas T-AH

Referencias

Chaves-Fraga, D., Óscar Corcho, and Ruckhaus, E. (2022). Guía práctica para la publicación de datos enlazados. *datos.gob.es*.

Cimiano, P., Buitelaar, P., McCrae, J., and Sintek, M. (2011). Lexinfo: A declarative model for the lexicon-ontology interface. *Web Semantics: Science, Services and Agents on the World Wide Web*.

de Melo, G. (2015). Lexvo.org: Language-related information for the Linguistic Linked Data cloud. *Semantic Web*, 6(4):393–400.

Di Buono, M. P., Cimiano, P., Elahi, M. F., and Grimm, F. (2020). Terme-a-llod: Simplifying the conversion and hosting of terminological resources as linked data. In *Proceedings of the 7th Workshop on Linked Data in Linguistics (LDL-2020)*, pages 28–35.

McCrae, J. P., Bosque-Gil, J., Gracia, J., Buitelaar, P., and Cimiano, P. (2017). The Ontolex-Lemon model: development and applications. In *Proceedings of eLex 2017 conference*, pages 19–21.