

*Semántica de Datos de Investigación en Universidades*

**Multilingüismo (y *multiescriptalismo*) en las ontologías ASIO**

www.ticarum.es · info@ticarum.es

**Descripción**

|  |  |
| --- | --- |
| Proyecto | HERCULES |
| Documento | Multilingüismo y “multiescriptalismo” en las ontologías ASIO del proyecto Hercules |
| Autor | José Emilio Labra Gayo, Emilio Rubiera Azcona |
| Destinatario |  |
| Nombre del fichero | Hercules ASIO v1.0\_multilinguismo.docx |
| Versión actual | 0.2 |
| Fecha modificación | 18/12/2019 |

**Resumen**

Descripción básica del contenido del documento.

**Control de versiones**

|  |  |  |  |
| --- | --- | --- | --- |
| **Ver.** | **Fecha** | **Autor** | **Cambios** |
| 0.1 | 22/12/2019 | WESO-IZERTIS | Versión inicial partiendo del documento de análisis de requisitos |
| 0.2 | 28/12/2019 | WESO-IZERTIS | Añadida sección “multiescriptalismo” |
| 0.3 | 13/01/2020 | WESO-IZERTIS | Contenidos para versión preliminar |

Contenido

[1 Introducción 4](#_Toc29999509)

[2 URIs 4](#_Toc29999510)

[3 Etiquetas léxicas 4](#_Toc29999511)

[4 Propiedades de documentación 5](#_Toc29999512)

[5 Otros metadatos lingüísticos 5](#_Toc29999513)

[6 “Multiescriptalismo” 6](#_Toc29999514)

[6.1 Personal 6](#_Toc29999515)

[6.2 Publicaciones 9](#_Toc29999516)

[7 Multilingüismo, “multiescriptalismo” y Wikidata/Wikibase 9](#_Toc29999517)

[8 Referencias 9](#_Toc29999518)

# Introducción

Todas las ontologías que se desarrollen estarán planteadas con un modelo de multilingüismo. Algunas de las reglas de este modelo son:

# URIs

- Preferencia por la utilización de URIs opacas[[1]](#footnote-1), especialmente para identificadores. Las URIs opacas permiten separar el identificador de la representación en lenguaje natural del recurso. Esta separación se considera una mejor práctica en entornos multilingües dado que las representaciones textuales deberían obtenerse a partir de las etiquetas multilingües (rdfs:label) en lugar del identificador del recurso.

# Etiquetas léxicas

Si en un principio se pensó utilizar únicamente etiquetas rdfs:label para todas las URIs (patrón label everything[[2]](#footnote-2)) y con un valor literal en al menos 2 idiomas: español e inglés (patrón multilingual labels[[3]](#footnote-3)) y utilizar skos:altLabel para identificar alias o nombres alternativos.

Duplicidad de las etiquetas léxicas: rdfs:label no es suficiente y skos:prefLabel, duplicándola, ofrece otras ventajas, como la relación con skos:altLabel más lógica y *modelada*.

# Propiedades de documentación

Utilización de otras representaciones en lenguaje natural del recurso mediante las propiedades rdfs:comment, dc:description, etc.

También skos ofrece otras interesantes para nuestros intereses.

# Otros metadatos lingüísticos

- Utilización de metadatos lingüísticos cuando sea necesario (patrón linguistic metadata[[4]](#footnote-4)). Aunque para muchos recursos, puede ser suficiente disponer de una representación en lenguaje natural ligera con etiquetas como rdfs:label, rdfs:comment, etc. En algunos vocabularios puede ser interesante utilizar representaciones lingüísticas más ricas como puede ser el modelo Lemon[[5]](#footnote-5), que permite representar aspectos léxicos de forma semántica. Sin embargo, Lemon puede no ser suficiente para representar todo el conocimiento léxico de forma semántica [2] y este tipo de representaciones es una línea de investigación activa. A modo de ejemplo, en el proyecto europeo Prét-á-LLOD[[6]](#footnote-6) se está actualmente trabajando en la creación y representación de datos enlazados multilingües.

- A la hora de seleccionar vocabularios para reutilizar, elegir vocabularios multilingües en la medida que sea posible[[7]](#footnote-7).

# “Multiescriptalismo”

Denominamos “multiescriptalismo”, adaptando el neologismo en inglés “multiscriptalism” de Coulmas (1996), a la faceta múltiple de las lenguas que tiene que ver no con la lengua propiamente dicha, sino con el sistema o sistemas de escritura que una lengua puede utilizar, es decir, para referirnos a la diversidad de sistemas de escritura (alfabetos, silabarios, sistemas logográficos, etc.) que se pueden encontrar en las lenguas naturales y en otros tipos de lenguajes, como por ejemplo las notaciones fonéticas.

Aunque las ontologías de ASIO en principio solamente contemplan el español y el inglés como lenguas *de trabajo*, una estructura ontológica rigurosa y bien diseñada siempre debería considerar e incluir un modelo de *internacionalización* (i18n) preparatorio para todas las *localizaciones* (l10n) que fueran necesarias.

Tener en mente estas características en las fases fundacionales del proyecto y crear una infraestructura ad hoc de antemano suele ahorrar mucho trabajo futuro, aparte de garantizar una reutilización mucho más global, que es uno de los leitmoif de la Web Semántica.

En una infraestructura ontológica como ASIO, preparada para el multilingüismo, este aspecto no debería ser desdeñado, por lo que se intentará poner un especial énfasis desde el principio en facilitar la inclusión de todas las variedades de sistemas de escritura utilizados en lenguas naturales y artificiales que sean relevantes.

## Personal

Imaginemos que la Universidad de Murcia contrata a 4 investigadores europeos como profesores:

* Un griego: Ilias Koubarakis - Κουμπαράκης, Ηλίας
* Un serbio: Aleksandar Karađorđević – Александар Карађорђевић
* Un búlgaro: Grigor Batinkov – Григор Батинков
* Un ruso: Ivan Uvarov – Иван Уваров

Todos ellos tienen una carrera académica previa en sus países de origen, donde han participado en proyectos nacionales y han publicado en sus lenguas de origen. Pero obviamente muchas veces, esa trayectoria previa aparece codificada no solamente en su lengua de origen, también en el alfabeto que usa esa lengua.

De acuerdo con la filosofía multilingüe que opera en la Unión Europea y resultaría integrador e interesante disponer de esas referencias en la lengua y alfabeto originales y, también, cuando éste .

Veamos algún ejemplo de como las URIs deberían

Para el caso griego,

rdfs:label "Ηλίας Κουμπαράκης"@el

rdfs:label "Ilias Koubarakis"@el-Latn

skos:prefLabel "Ηλίας Κουμπαράκης"@el

skos:prefLabel "Ilias Koubarakis"@el-Latn si priorizamos por ejemplo la romanización ELOT[[8]](#footnote-8).

skos:altLabel "Ilias Koubarakis"@el-Latn o como label alternativo en la romanización ISO[[9]](#footnote-9).

Para el caso búlgaro,

rdfs:label "Григор Батинков"@bg

rdfs:label "Grigor Batinkov"@bg-Latn

skos:prefLabel "Григор Батинков"@bg

skos:altLabel "Гришо Батинков"@bg

skos:prefLabel "Grigor Batinkov"@bg-Latn

Para el caso serbio,

rdfs:label "Александар Карађорђевић"@sr

rdfs:label "Aleksandar Karađorđević"@sr-Latn

skos:prefLabel "Александар Карађорђевић"@sr

skos: altLabel "Саша Карађорђевић"@sr

skos:prefLabel "Aleksandar Karađorđević"@sr-Latn

skos:altLabel "Aleksandar Karadjordjević"@sr-Latn

Para el caso ruso,

rdfs:label "Иван Уваров"@ru

rdfs:label "Ivan Uvarov"@ru-Latn

rdfs:label "Иван Уваров"@ru-Latn

skos:prefLabel "Иван Уваров"@ru

skos:prefLabel "Ivan Uvarov"@ru-Latn

skos:altLabel "Aleksandar Karadjordjević"@ru-Latn

Obviamente estamos en Murcia, y nos interesa

Garantizar el multilingüismo, por una parte, pero también el multiescriptalismo, resultaría no solamente un detalle de cortesía para con nuestros nuevos integrantes de la platilla, sino también un perfeccionamiento del modelo de datos subyacente, que internacionalizaríamos como paso previo a cualquier localización ad hoc posterior.

## Publicaciones

Como comentábamos en la sección anterior, nuestras nuevas incorporaciones extranjeras cuentan con publicaciones en sus lenguas de origen.

Ejemplos artículos o libros en alfabetos diferentes y romanizaciones.

# Multilingüismo, “multiescriptalismo” y Wikidata/Wikibase

El modelo de datos que subyace a los enlaces de Wikibase se basa en “entidades” que incluyen elementos individuales, etiquetas o identificadores para describirlos y *declaraciones* *semánticas* que atribuyen “propiedades” a cada elemento.

En principio, las etiquetas o identificadores soportan potencialmente el tratamiento multilingüe, aunque queda por analizar si este soporte alcanza plenamente los diferentes sistemas de escritura, cómo lo lleva a cabo y si su procedimiento es realmente conforme a los requisitos de la Web Semántica y los principios de Datos Enlazados.

# Referencias

Coulmas, F. (1996). *The Blackwell Encyclopedia of Writing Systems*. Oxford, U.K.: Blackwell Publishers.

1. http://www.weso.es/MLODPatterns/Opaque\_URIs.html [↑](#footnote-ref-1)
2. http://www.weso.es/MLODPatterns/Label\_everything.html [↑](#footnote-ref-2)
3. http://www.weso.es/MLODPatterns/Multilingual\_labels.html [↑](#footnote-ref-3)
4. http://www.weso.es/MLODPatterns/Linguistic\_metadata.html [↑](#footnote-ref-4)
5. http://lemon-model.net/lemon# [↑](#footnote-ref-5)
6. https://www.pret-a-llod.eu/ [↑](#footnote-ref-6)
7. http://www.weso.es/MLODPatterns/Multilingual\_vocabularies.html [↑](#footnote-ref-7)
8. Ελληνικός Οργανισμός Τυποποίησης [Ellīnikós Organismós Typopoíīsīs, "Hellenic Organization for Standardization"]. ΕΛΟΤ 743, 2η Έκδοση [ELOT 743, 2ī Ekdosī, "ELOT 743, 2nd ed."]. ELOT (Athens), 2001. [↑](#footnote-ref-8)
9. United Nations Group of Experts on Geographical Names, Working Group on Romanization Systems. Report on the Current Status of United Nations Romanization Systems for Geographical Names: "Greek". United Nations (New York), 2003. [↑](#footnote-ref-9)