# Documentación

17.8.2014

Empiezo a etiquetar Mateo. Utilizo

Utilizo la siguiente expresión regular:

(^|[^>A-z])(Jesús)([^<A-z]|$)

\1<rs key="per1">\2</rs>\3

Cambiado en general la etiqueta <name> por <rs>

## Atributos

### type

* organization
* place
* person

### key

* ids en el excel!

# Trabajo por hacer

No he puesto en relación el type y el prefijo de si es persona o lugar.

Colocar etiquetas en tiempo. Las he ido documentando en una pestaña del Excel.

Colocar etiquetas de speech. Casi todas, claro, son de Jesús. Algunos ejemplos para sacar expresiones regulares:

* diciendo: "He
* le dijeron: --¿Por qué
* dijo: --Porque
* les dijo: 'Un hombre
* se Cumplió lo dicho por medio del profeta diciendo: Abriré mi

No estaría mal también las narraciones, las parábolas y cosas así. No sé con qué etiqueta :S

Una manerade etiquetar automáticamente el speach sería colocar el último atributo de persona, es decir, buscar por la última etiqueta <name y rescatar el atributo key.

Si se hace esto, habría que tener en cuenta solo aquellas etiquetas que con el atributo person u organization; y no place!

Si se hace eso, hay que tenerse en cuenta que muchas veces el fin del parlmento no se termina de ninguna manera, sencillamente se termina con un punto. Esto significa que si se lleva la etiqueta <sp hasta el final del <seg, habrá que revisarlo porque habrá errores.

# Dudas

Decido no etiquetar los pronombres.

Decido no etiquetar los nombres comunes (leproso, suegra, etcétera). Sí queda etiquetado el centurión.

Mat 9.1: llegó a su propia ciudad? Nazaret o Belén?

Día del juicio y Tribulación de aquellos Días es lo mismo??

En general no se marcan los sustantivos comunes en plural, como "reyes" o "discípulos".

Sí marco "ángeles" y cualquier ángel sin nombre propio con el número 13

Mat 12 1: marco sábado? cómo?

Hay varias menciones a profeta que habría que desentrañar de manera más compleja que una expresión regular.

Espíritu como demonio no lo he marcado. Si lo he hecho con ángel... ¿?

Hay varias veces que aparece espíritu en mayúscula y sin una clara referencia a si es al Espíritu Santo, a Dios o a Jesús.

Mateo 13.1, hay una referencia al mar. Estaría bien señalar cuál.

Mat 13.55, aparece por segunda vez mencionados los hermanos de Jesús: Jacobo, José, Simón Judas.

No estoy seguro de qué hacer con los parlamentos (u otras unidades) que duran varios versículos. Ponerlos por encima de los versículos podría sacar problemas con los parlamentos que comienzan en medio de un versículo (cosa muy frecuente). Ponerlos bajo el versículo haría repetir muchíiiiisimas etiquetas...

tierra como lugar???

Desarrollo de trabajo

2014.08.17: Hecho hasta 16

Mat.17.1: aparece un monte alto, cuál?

2014 08 19: hecho hasta 19

Mat.20.18 debería codificar sacerdotes?

Mat.20.19 debería codificar escribas?

Mat.20.19 : tercer Día? codificar?

Mat.21.2 aparece una aldea cerca del monte de los olivos ¿?

Mat.21.42 aparecen las escrituras. Habría que codificarlo!!! En Mat.22.36 se menciona "la ley"

Mat.22.17 codificado César, aunque no es una persona, sino un puesto!

Mat.23.16 aparece "Guías" (anteriormente tb en minúscula), parece que es un tipo de cargo, similar a escriba o fariseo.

Mat.26.18 se menciona una ciudad ¿?

280823: Terminado etiquetación de personas, lugares y organizaciones. Hay que revisar ahora los momentos y sobre todo los speaches!!!!

## 3.10.2014

Después de darle bastantes vueltas, llego a la conclusión de que la mejor manera de encontrar fragmentos de diálogo con la persona hablante en el mismo versículo es con la siguiente expresión regular:

("(per[0-9]+).\*?[^=])"([^>].\*?[^=])"([^>])

\1<!--revisar--><sp who="\2">\3</sp>\4

Mejorada

("(per[0-9]+)((?!"per).)\*>[^<]\*?)"([^>]\*?[^\r])"([^>])

\1<!--revisar--><sp who="\2">\4</sp>\5

Para encontrar los diálogos con hablantes antes, utilizo la siguiente expresión regular:

((per[0-9]\*)((?!"per).)\*>[^<]\*?)"([^>]\*?[^\r]\*)"

\1<!--revisar--><sp who="\2">\4</sp>

Para encontrar diálogos con hablantes donde no se cierra la cita directamente, utilizo estas expresión regular:

((per[0-9]\*)((?!"per).)\*>[^<]\*?)"([^>]\*?[^\r"]\*)\r

\1<!--revisar--><sp who="\2">\4</sp>

Después de darle una cuantas vueltas, encuentro una mejor manera:

(<rs key="(per[0-9]+)">((?!"per).)\*?>[^<]\*?)"([^>\r]\*?<\*.\*?)"?$

\1<!--revisar--><sp who="\2">\4</sp>

Explicación

(buscamos la etiqueta <rs key="(per+número)">después una cadena de caracteres que no contenga "per, después el caracter > después cualquier caracter que no contenga <) después comilla, (después cualquier caracter que no sea ni > ni salto de línea, después < después cualquier caracter y después opcionalmente ")

Aunque parezca difícil de creer, esto: >[^<]\*?)"([^>\r]+?< sirve para buscar comillas que no estén dentro de una tag.

Para buscar oraciones con una persona y una --

("(per[0-9]+).\*?[^!])--([^>].\*?)$

\1<!--revisar--><sp who="\2">\3</sp>

Para buscar oraciones sin per

("(per[0-9]+).\*?[^!])--([^>].\*?)$

\1<!--revisar--><sp who="\2">\3</sp>

10.11.2014

Revisado y codificados los parlamentos hasta el capítulo 7

24.11.2014

Mat 8:28: región de los gadarenos? qué lugar es?

Mateo 9:1: "Llegó a su propia ciudad"? Nazaret?

Automatizar "el día del juicio". Tb "fin del mundo"

Mateo 14.3: este Felipe es correcto???

## 26.11.2014

Hasta el capítulo 19

## 7.1.2015:

hasta 22.24

Hasta 26.51

## 8.1.2015:

Se repite a veces "ancianos del pueblo". Dudo sobre si codificarlo.

Muchas veces aparece mencionado los principales sacerdotes y los ancianos de manera seguida. Podríamos etiquetarlos como Sanedrín?

Modificados automáticamente los <sp key por <sp who

Codificado hasta el final  , incluido el primer capítulo con la larguísima geneología.

## Por hacer

Crear esquema de documento.

Revisar si he utilizado de la mejor manera las etiquetas de <sp>. ¿Coloco más?

Revisar etiqueta <seg>

## 30.1.2016

Se cambia said por q

Se coloca un tipo con los siguientes valores:

* spoken: por defecto
* written: cuando se señala que debía ser escrito o no puede haber otra opción por el contexto
* unknown: cuando
* dream:
* prayer:
* soCalled: cuando lo señalado es una suposición de la persona
* idea: cuando no se ha dicho, pero se señala como algo que se podría decir o que se dirá

No puedo poner diferentes valores, pero debaría poder.

## 2016.2.13

Para empezar a preparar Juan utilizo esta búsqueda:

for $s in distinct-values(//@key) return ($s, distinct-values(//\*[@key=$s]))

## 2016.02.20

Decido reutilizar partes de Mateo para preparar Juan. Saco una lista de las cadenas de texto que están en rs junto con su ids. Para sacar esto utilizo un archivo XSLT llamado id-text.xsl

Lo limpio mediante expresiones regulares extrayendo aquellas cadenas que tenían varios atributos (sé que se puede hacer programando, pero no me apetecía...).

Manualmente saco aquellas cadenas que son ambivalentes (María, Juan, Señor, etcétera). Los datos están en /TEIBibel/documentacion/datos/rs con indices de mateo.ods. Aquellas cadenas que no son ambivalentes, las coloco automáticamente mediante un archivo python (TEIBibel/programacion/python/putidsinstrings.py) en Juan. Ese archivo también es bastante mejorable, qué le vamos a hacer...

Hay que revisarlo y además colocar con cuidado las que sí son ambivalentes. Cabría pensar en alguna manera de semiautomatizar también el trabajo con las ambivalentes, por ejemplo sacando aquellas cadenas que en una gran mayor parte de los casos no resultan ambivalentes.