

November 06, 2020

從宗教研究到數位宗教研究： 以TEI格式來標記宗教研究 期刊論文的經驗與反思

天主教輔仁大學宗教學系 張名揚 陳奕甫



數位人文與宗教研究

- 1949年義大利耶穌會神父Roberto Busa在IBM的幫助下，將聖多瑪斯阿奎那(St. Thomas Aquinas)的所有作品用電腦儲存，並建立一套可搜索的編制索引
- 臺灣數位典藏建置：法鼓山和CBETA online設置
 - 超過20年
 - 從數位典藏到數位研究工具
- 嘗試建立宗教文本標籤規範，作為數位研究之基礎
 - 短期目標：宗教研究論文標記
 - 長期目標：供研究者使用的數位文本版本及工具



文字標記

- 標記 markup 是什麼？
 - 在文字上加上標記，可以增添純文字不能表達的資料 (information)
- 為什麼要markup?
 - 把更多資料放進數位文本中
 - 目的：把我們（意會而）知道的，（以明示的方式）讓機器知道。



文字標記

- 兩主要類別：「排版或顯示格式」的標記 & 「資料結構或內容」的標記
 - 「排版或顯示格式」標記：e.g. `\textcolor{red}{-1000元}`
 - HTML (Hypertext Markup Language), TeX, etc.
 - 「資料結構或內容」標記：e.g. `<measure unit="TWD" value="-1000">負債臺幣壹仟圓</measure>`
 - XML



文字標記

- 兩主要類別：「排版或顯示格式」的標記 & 「資料結構或內容」的標記
 - 「排版或顯示格式」標記：e.g. `\textcolor{red}{-1000元}`
 - HTML (Hypertext Markup Language), TeX, etc.
 - 「資料結構或內容」標記：e.g. `<measure unit="TWD" value="-1000">負債臺幣壹仟圓</measure>`
 - XML



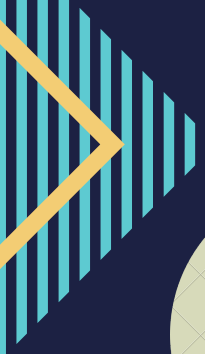
XML 文字標記例子

```
<lg>  
  <l>一切有為法</l><l>如夢幻泡影</l>  
  <l>如露亦如電</l><l>應作如是觀</l>  
</lg>
```

<lg> = 偈頌

<l> = ?

- 來源：周邦信，〈標記語言的應用〉，《佛教圖書館館訊》，第24期。
<http://www.gaya.org.tw/journal/m24/24-main5.htm>



→ **OxGarage XSLT**

平台
⋮



TEI

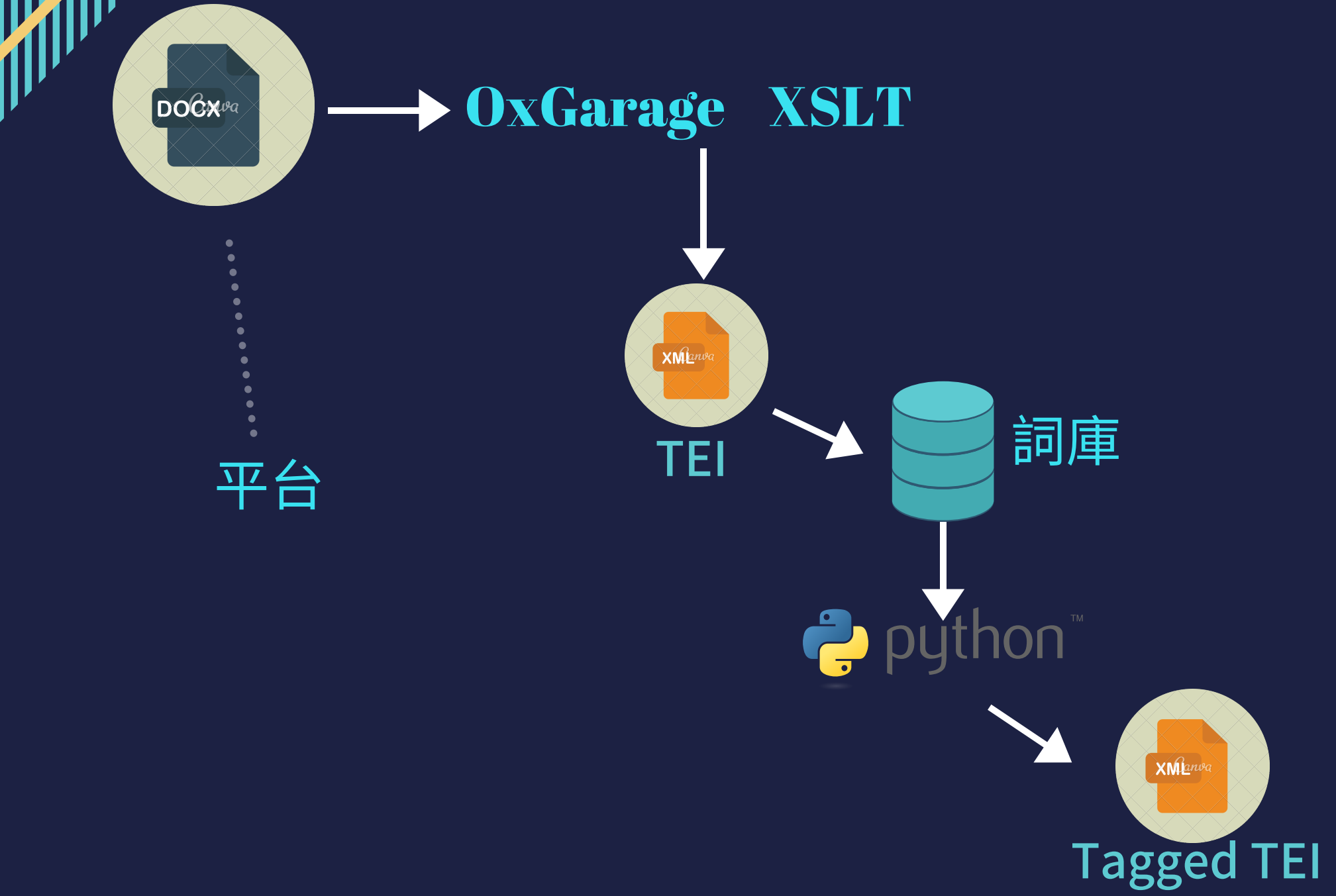


詞庫

python™



Tagged TEI





數位人文平台測試



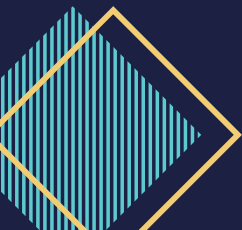
標記操作流程



詞庫建立

結語

數位人文平台 測試





匯入格式

MARKUS

中研院
數位人文研究
平台

UTF-8純文字檔

UTF-8文字檔

貼上文字

符合平台規範之
XML檔

MARKUS檔案
_MARKUS.html

MARKUS檔案
_MARKUS.html



標記方式

MARKUS

中研院
數位人文研究
平台

手動標記

手動標記

串聯資料庫，
直接標記

自動標記

單一文本
權威檔配對

上傳或匯入
既有權威檔，
與文本點選配對，
即完成標記

將若干關鍵詞(欲
標記詞)，標記於
一個標記名稱

關鍵字標記

多文本
配對多權威檔

同上，選擇多個權
威檔與文本

同時以若干標記名
稱，標記若干文本
中，若干詞彙

批次標記



匯出格式

MARKUS

中研院
數位人文研究
平台

MARKUS存檔
(_MARKUS.html)

XML

供瀏覽器檢視

網頁(排版)
_layout.html

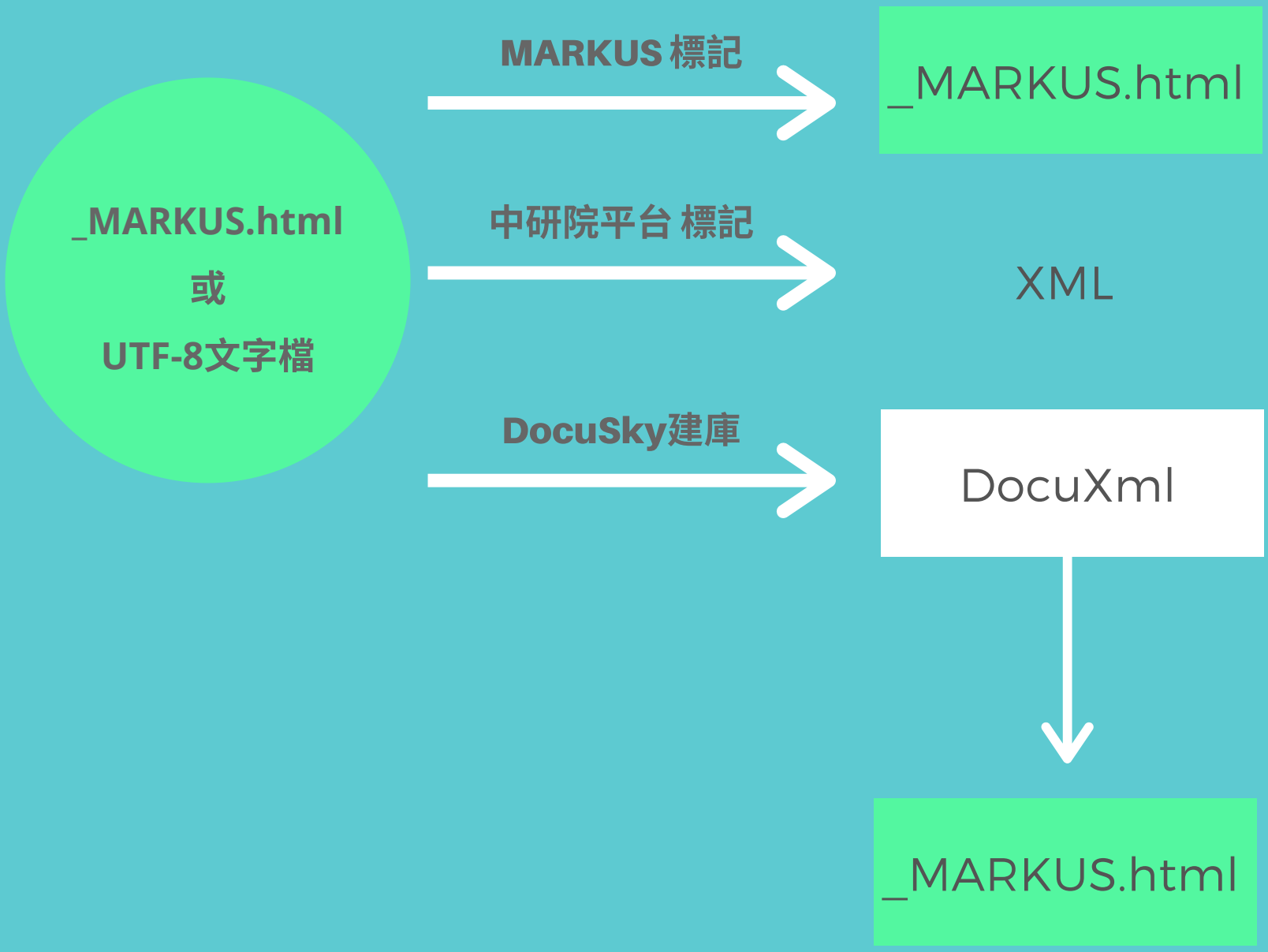
含標記之XML

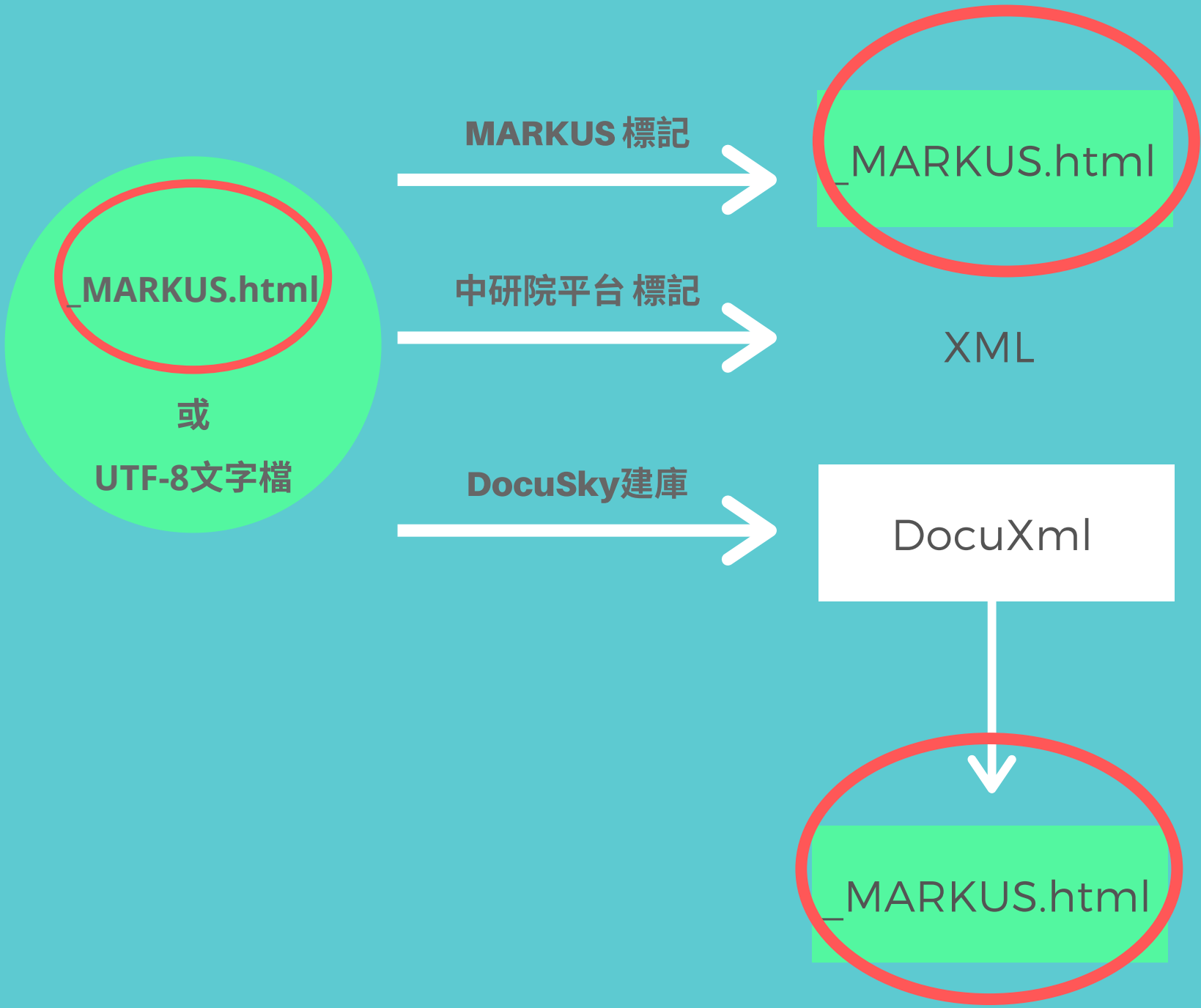
符合TEI標準之
XML檔

TEI(測試)
_tei.xml

含關係詞之XML

標記與關係詞皆
含括之XML

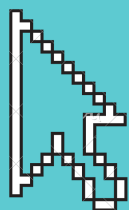
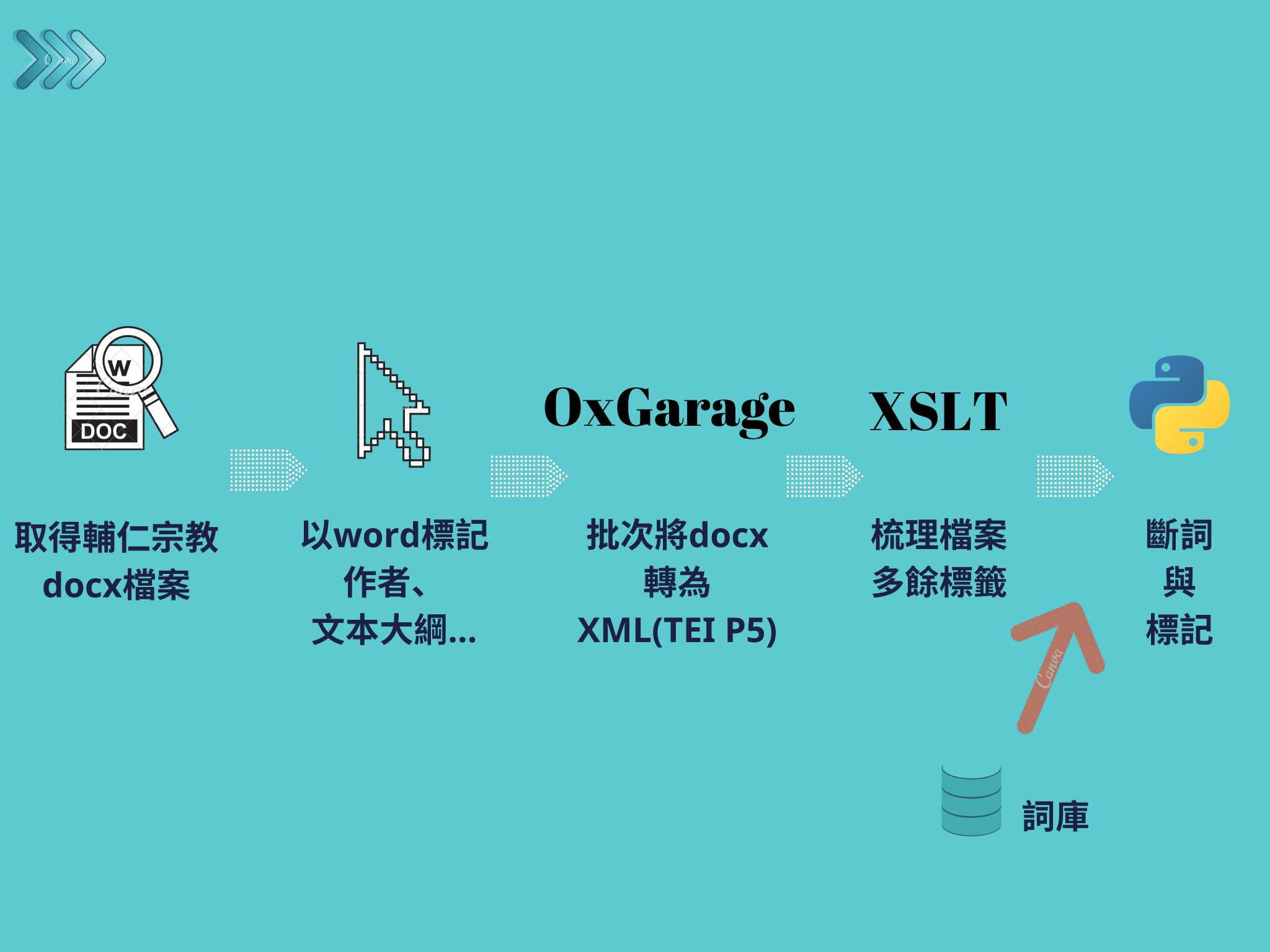




標記操作流程



標記宗教文本初步嘗試

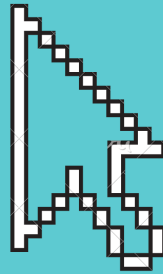


OxGarage

XSLT



詞庫



**以word標記
作者、文本大綱...**



生命教育與生命的誕生-艾克哈大師靈心神學的對話

陳德光

輔仁大學宗教學系教授

提要

筆者曾把生命教育了解作「感通」與「創生」的教育。「感通」與「創生」預設一種人學或靈學，至今，人學或靈學的論述仍然不足，人學或靈學與「感通」與「創生」之間的關係尚未討論。

本文寫作的目的與內容就是介紹基督宗教「靈心神學」(theologia cordis)，尤其是艾克哈大師(Meister Eckhart, 約1260-1327/8)論「聖子的誕生」(the birth of the Son)的道理，補充生命作為「感通」與「創生」的主題，建立生命教育的人學與宗教學的基礎。

此外，本文把艾克哈大師與本地文化作對比或對話，其中涉及東西方宗教「靈心之鏡」的比喻，突顯基督宗教傳統對開發本地生命教育的意義。

關鍵字：

艾克哈大師，聖子的誕生，靈心神學，靈心之鏡，生命教育



16-2.doc - 內容

- 一般
- 安全性
- 自訂
- 詳細資料
- 以前的版本

屬性	值
描述	
標題	生命教育與生命的誕生-艾克哈大師靈心神...
主旨	
標籤	
類別	
註解	
來源	
作者	陳德光



提要

關鍵字：

引言

一、生命教育回顧與前瞻

▲ 二、艾克哈論聖子的誕生

（一）相同子性的誕生

（二）自生與回生

（三）新生不息

（四）肖像與去障

（五）誕生與三元人學

三、東西「靈心鏡子」的對話

四、結論

▲ 參考書目：



```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>生命教育與生命的誕生-艾克哈大師
7         <author>陳德光</author>
8       </titleStmt>
```



生命教育與生命的誕生-艾克哈大師靈心神學的對話

陳德光

輔仁大學宗教學系教授

提要

筆者曾把生命教育了解作「感通」與「創生」的教育。「感通」與「創生」預設一種人學或靈學，至今，人學或靈學的論述仍然不足，人學或靈學與「感通」與「創生」之間的關係尚未討論。

本文寫作的目的與內容就是介紹基督宗教「靈心神學」(theologia cordis)，尤其是艾克哈大師(Meister Eckhart, 約1260-1327/8)論「聖子的誕生」(the birth of the Son)的道理，補充生命作為「感通」與「創生」的主題，建立生命教育的人學與宗教學的基礎。

此外，本文把艾克哈大師與本地文化作對比或對話，其中涉及東西方宗教「靈心之鏡」的比喻，突顯基督宗教傳統對開發本地生命教育的意義。

關鍵字：

艾克哈大師，聖子的誕生，靈心神學，靈心之鏡，生命教育



<p rend="center bold">生命教育與生命的誕生-艾克哈

<p rend="center">陳德光</p>

<p rend="center">輔仁大學宗教學系教授</p>

<p rend="bold">提要</p>

<p>筆者曾把生命教育了解作「感通」與「創生」的教育。

<p>本文寫作的目的與內容就是介紹基督宗教「靈心神學」
Eckhart, 約1260-1327/8) 論「聖子的誕生」 (the



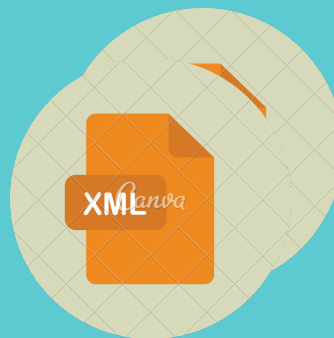
OxGarage

OxGarage

批次將docx

轉為

XML(TEI P5)





OxGarage Conversion

Choose the file, upload images and press convert

Select file to convert: [?](#)

選擇檔案 未選擇任何檔案

- Hide advanced options [?](#)

- Convert text only, don't include any images from the document
- Download images located on the internet
- Copy and rename all provided images

Upload images: [?](#)

You can upload image files and .zip files containing images

You do not need to upload any images, as images are included in the document file.

Conversion: Microsoft Word (.docx) -> TEI P5 XML Document

Language

Select conversion profile:

URL for conversion with current properties:

```
https://oxgarage-paderborn.tei-c.org/ege-webservice/Conversions/docx%3Aapplication%3Avnd.openxmlformats-officedocument.wordprocessingml.document/TEI%3Atext%3Axml/conversion?properties=<conversions><conversion index="0"><property id="oxgarage.getImages">true</property><property id="oxgarage.getOnlineImages">true</property><property id="oxgarage.lang">en</property><property id="oxgarage.textOnly">>false</property><property id="pl.psnc.dl.ege.tei.profileNames">default</property></conversion></conversions>
```

URL for conversion with default properties:

```
https://oxgarage-paderborn.tei-c.org/ege-webservice/Conversions/docx%3Aapplication%3Avnd.openxmlformats-officedocument.wordprocessingml.document/TEI%3Atext%3Axml/
```

Convert

Reset



Select file to convert: ?

Upload images: ?

You can upload image files and .zip files containing images

選擇檔案

未選擇任何檔案

You do not need to upload any images, as images are included in the document file.

- Hide advanced options ?

- Convert text only, don't include any images from the document
- Download images located on the internet
- Copy and rename all provided images

Conversion: Microsoft Word (.docx) -> TEI P5 XML Document

Language en ▼

Select conversion profile: default ▼

URL for conversion with current properties:

```
https://oxgarage-paderborn.tei-c.org/ege-webservice/Conversions/docx%3Aapplication%3Avnd.openxmlformats-officedocument.wordprocessingml.document/TEI%3Atext%3Axml/conversion?properties=<conversions><conversion index="0"><property id="oxgarage.getImages">true</property><property id="oxgarage.getOnlineImages">true</property><property id="oxgarage.lang">en</property><property id="oxgarage.textOnly">>false</property><property id="pl.psnc.dl.ege.tei.profileNames">default</property></conversion></conversions>
```

URL for conversion with default properties:

```
https://oxgarage-paderborn.tei-c.org/ege-webservice/Conversions/docx%3Aapplication%3Avnd.openxmlformats-officedocument.wordprocessingml.document/TEI%3Atext%3Axml/
```



OxGarage Conversion

TEI Consortium 網站提供 OxGarage 檔案轉換工具
(open source project)

經 Web 介面 oxgarage2.tei-c.org
可選擇來源檔格式 (txt, doc, docx, odt, ...)
及輸出格式 (txt, doc, odt, xml [TEI P5])

本計畫：把 doc 和 docx 檔
轉換成 TEI P5 XML documents



OxGarage Conversion

使用 https service

<https://oxgarage-paderborn.tei-c.org/ege-webservice/Conversions/docx%3Aapplication%3Avnd.openxmlformats-officedocument.wordprocessingml.document/TEI%3Atext%3Axml/>

(因來源格式和輸出格式不同，https 地址也不同)



OxGarage Conversion

OxGarage 檔案轉換工具的 https service

更可加入如何處理檔案內圖像的 properties

```
...TEI%3Atext%3Axml/conversion?properties=<conversions>
```

```
<conversion index="0"><property id="oxgarage.getImages">>true
```

```
... </conversions>
```

以 script 來送出 https service request 來批次處理期刊 Word 檔 (共122個)

XSLT

梳理檔案

多餘標籤

處理的標籤： hi, p, cell

保留之 attribute： center, bold, underline, doubleunderline, italic, start, end, color([non-black]), background-color([non-white]) ...

刪除之 attribute： justify, color(#000000), background-color(#ffffff)

```
<hi rend="color(#000000)bold">内容</hi>
```



```
<hi rend="bold">内容</hi>
```


XSL使用(1)

```
<xsl:template  
match="hi[@rend='color(#000000)bold']">
```

```
  <xsl:copy>
```

```
    <xsl:attribute  
name="rend">bold</xsl:attribute>
```

```
  <xsl:apply-templates/>
```

```
</xsl:copy>
```

```
</xsl:template>
```

XSL使用(2)

```
<xsl:template match="cell[contains(@rend,'color(#000000)')]
|cell[contains(@rend,'justify')]">
  <xsl:copy>
    <xsl:for-each select="@*">
      <xsl:choose>
        <xsl:when test="name(.)='rend'">
          <xsl:if test=". != 'color(#000000)' and != 'justify' and
          != 'justify color(#000000)' and
          != 'justify color(#000000) background-color(#ffffff)'">
            <xsl:attribute name='rend'>
              <xsl:value-of select="normalize-space(replace(replace(
replace(., 'color\(#000000\)', ''), 'justify',
' '), 'background-color\(#ffffff\)', ''))"/>
            </xsl:attribute>
          </xsl:if>
        </xsl:when>
      </xsl:choose>
    </xsl:for-each>
  </xsl:copy>
</xsl:template>
```

[...]

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" data-bbox="10 300 500 330">
  <author data-bbox="10 340 500 370">教育與生命的誕生-艾克哈大師靈心神學的對話</author>
  <titleStmt data-bbox="10 380 500 410"></titleStmt>
  <editionStmt data-bbox="10 420 500 450"></editionStmt>
  <publicationStmt data-bbox="10 460 500 490"><p>no available</p></publicationStmt>
  <sourceDesc data-bbox="10 500 500 530"><p>OpenOffice</p></sourceDesc>
  <fileDesc data-bbox="10 540 500 570"></fileDesc>
  <me data-bbox="10 580 500 610"><date>2019-02-25T07:13:00</date></me>
  <text data-bbox="10 620 500 710">
    <body data-bbox="10 630 500 710">
      <p rend="center color(#000000)bold" data-bbox="10 640 500 670">生命教育與靈  

        的對話</p>
      <p rend="center" data-bbox="10 680 500 710">陳德光</p>
      <p rend="justify color(#000000)bold" data-bbox="10 720 500 750">提要</p>
    </body>
  </text>
</TEI>
```

```
<listChange data-bbox="510 290 990 320">
  <change data-bbox="510 330 990 360"><name>if</name><date></date></change>
</listChange>
</revisionDesc>
</teiHeader>
<text data-bbox="510 490 990 520">
  <body data-bbox="510 530 990 710">
    <p rend="center bold" data-bbox="510 570 990 600">生命教育與靈  

    <p rend="center" data-bbox="510 610 990 640">陳德光</p>
    <p rend="center" data-bbox="510 650 990 680">輔仁大學宗教學系  

    <p rend="bold" data-bbox="510 690 990 720">提要</p>
  </body>
</text>
```



斷詞與標記



原理

搜尋

天主教會
天主
教會



取代

```
<rs type="sect">天主教會</rs>  
  <rs type="god">天主</rs>  
<rs type="community">教會</rs>
```



原理

csv檔

	A <input type="text" value="▼"/>	B
1	天主教會	<code><rs type="sect">天主教會</rs></code>
2	一性論	<code><rs type="theology">一性論</rs></code>
3	一體一位論	<code><rs type="theology">一體一位論</rs></code>
4	二性一位的結	<code><rs type="theology">二性一位的結合</rs></code>
5	十二宗徒訓誨	<code><rs type="document">十二宗徒訓誨錄</rs></code>
6	十字架神學	<code><rs type="theology">十字架神學</rs></code>



```
import csv
```

```
import os
```

```
with open(r'tagdic.csv') as freq:
```

```
    reader = csv.reader(freq)
```

```
    freq_list = list(reader)
```

```
sourcepath = os.listdir('InputFiles/')
```

```
for file in sourcepath:
```

```
    inputfile = 'InputFiles/' + file
```

```
    print('Conversion is ongoing for:'+ inputfile)
```

```
    with open(inputfile,'r') as inputfile:
```

```
        filedata = inputfile.read()
```

```
    destinationpath = 'OutputFiles/' + file
```

```
    for i in freq_list:
```

```
        filedata = filedata.replace(i[0],i[1])
```

```
    with open(destinationpath,'w') as file:
```

```
        file.write(filedata)
```

詞庫

批次讀取資料夾所有文本

依序讀取詞庫欄位



過度標記與重複標記

程式執行次序	詞庫欄位排序	執行結果
1	天主教會	<code><rs type="sect">天主教會</rs></code>
2	天主	<code><rs type="sect"><rs type="god"> 天主</rs>教會</rs></code>
3	教會	<code><rs type="sect"><rs type="god">天主 </rs><rs type="community">教會</rs> </rs></code>



為了避免 過度標記與重複標記 我們嘗試了斷詞加入space

< p > 天主教會 於 二十 世紀 六十年 代 所 召開 的
梵蒂岡 第二 屆 大公 會議



為了避免 過度標記與重複標記 我們嘗試了斷詞加入space

< p > 天主教會 於 二十 世紀 六十年代 所召開 的
梵蒂岡 第二 屆 大公會議

< p ><rs type="sect">天主教會</rs>於 二十 世
紀 六十年代 所召開 的 梵蒂岡 第二 屆 大公會議

<p><rs type="sect">天主教會</rs>於二十世紀六
十年代所召開的梵蒂岡第二屆大公會議

詞庫建立



分類

類別

標籤

神祇

<rs type="god"><rs>

思想與教義

<rs type="<rs type="theology">eology"><rs>

節日

<rs type="season"><rs>

建築

<rs type="architecture"><rs>

藝術

<rs type="art"><rs>

經驗

<rs type="experience"><rs>

流派

<rs type="sect"><rs>

團體

<rs type="community"><rs>

分類

類別

標籤

教制

<rs type="polity"></rs>

生物靈體

<rs type="beings"></rs>

地名

<rs type="place"></rs>

事件

<rs type="event"></rs>

物品

<rs type="object"></rs>

牧靈用語

<rs type="pastoral"></rs>

職稱

<rs type="role"></rs>

人物

<rs type="person"></rs>

經典

<rs type="scripture"></rs>

文件

<rs type="document"></rs>

儀式

<rs type="rite"></rs>

行為

<rs type="behavior"></rs>



morethanbooks / XML-TEI-Bible

<> Code



Issues 0



Pull requests 0

GitHub is h

	A	B	C	D	E
1		id	normalizedName	clarification	variants
4	2	#per2	Abraham		Abram
5	3	#per3	Isaac		
6	4	#per4	Jacob		Israel
7	5	#per5	Judá		
8	6	#per6	Tamar		
9	7	#per7	Fares		
10	8	#per8	Zéraj		
11	9	#per9	Hesrón		
12	10	#per10	Herodes		
13	11	#per11	Maria	madre de Jesús	
14	12	#per12	José	padre de Jesús	

	I	J	K	L	M
	Gender	normalizedName	normalizedName	type	nature
	male	Abraham	Abraão	person	superior-human
	male	Isaak	Isaque	person	human
	male	Jakob	Jacó	person	human
	male	Juda	Judá	person	human
	female	Tamar	Tamar	person	human
	male	Perez	Perez	person	human
	male	Serach	Zerá	person	human
	male	Hezron	Esrom	person	human
	male	Herodes	Herodes	person	human
	female	Maria	Maria	person	human

<rs key="#per2">Abraham</rs>
<rs key="#per2">Abraham</rs>:
ostro sobre su rostro, y se ri
key="#per14">Dios</rs>: <q who
con <rs key="#per2">Abraham</rs>
a su <rs key="#per140">hijo</rs>
a su <rs key="#per140">hijo</rs>
y nueve años de edad cuando c
ey="#per2">Abraham</rs> y su <

分類參考：內政部宗教資訊、神學辭典

視角：多瑪斯學派、聖母

舊新約巴比倫象徵



結語

未來可能性

- wiki檢閱
- 關鍵詞分類統計

尚待解決

- 斷詞、分類細節
- 

Thanks for listening

