

# CWordTM Toolkit Usage on the Holy Bible (CUV)

This Jupyter notebook demonstrates how to use the package "CWordTM" on the Holy Bible (Chinese Union Version - Traditional Chinese):

- 1. Utility Features
- 2. Text Visualization - Word Cloud
- 3. Text Summarization
- 4. Pivot Table
- 5. OT Quotes
- 6. Topic Modeling - LDA, BERTopic and NMF

CWordTM Toolkit's Documentation: <https://cwordtm.readthedocs.io>

```
In [1]: import warnings
warnings.filterwarnings('ignore')

In [2]: # Import the Package CWordTM
import cwordtm
from cwordtm import *

In [3]: cwordtm.__version__

Out[3]: '0.6.4'
```

## 1. Utility Features

```
In [4]: # Load the whole Bible (Chinese Union Version)
bible = "cuvs.csv"
cdf = util.load_word(bible, info=True)

Loading file 'D:\Dev\Anaconda3\lib\site-packages\cwordtm\data\cuvs.csv' ...

Dataset Information:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 31102 entries, 0 to 31101
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype
---  ---
0    book        31102 non-null  object
1    book_no     31102 non-null  int64
2    chapter     31102 non-null  int64
3    verse       31102 non-null  int64
4    text        31102 non-null  object
5    testament   31102 non-null  int64
6    category    31102 non-null  object
7    cat         31102 non-null  object
8    cat_no      31102 non-null  int64
dtypes: int64(5), object(4)
memory usage: 2.1+ MB
```

### Extract Partial Scripture

```
In [5]: # Extract OT Scripture
otdf = util.extract(cdf, testament=0)
otdf.head()
```

```
Out[5]:
```

	book	book_no	chapter	verse	text	testament	category	cat	cat_no
0	Gen	1	1	1	起初 神創造天地。	0	Torah	tor	0
1	Gen	1	1	2	地是空虛混沌。淵面黑暗。 神的靈運行在水面上。	0	Torah	tor	0
2	Gen	1	1	3	神說、要有光、就有了光。	0	Torah	tor	0
3	Gen	1	1	4	神看光是好的、就把光暗分開了。	0	Torah	tor	0
4	Gen	1	1	5	神稱光為晝、稱暗為夜。 有晚上、有早晨、這是頭一日。	0	Torah	tor	0

```
In [6]: # Extract Gospels (The first four book in NT)
gos = util.extract(cdf, category='gos')
gos.head()
```

Out[6]:

	book	book_no	chapter	verse	text	testament	category	cat	cat_no
23145	Mat	40	1	1	亞伯拉罕的後裔、大衛的子孫、耶穌基督的家譜。〔後裔子孫原文都作兒子下同〕	1	Gospel	gos	5
23146	Mat	40	1	2	亞伯拉罕生以撒。以撒生雅各。雅各生猶大和他的弟兄。	1	Gospel	gos	5
23147	Mat	40	1	3	猶大從他瑪氏生法勒斯和謝拉。法勒斯生希斯崙。希斯崙生亞蘭。	1	Gospel	gos	5
23148	Mat	40	1	4	亞蘭生亞米拿達。亞米拿達生拿順。拿順生撒門。	1	Gospel	gos	5
23149	Mat	40	1	5	撒門從喇合氏生波阿斯。波阿斯從路得氏生俄備得。俄備得生耶西。	1	Gospel	gos	5

In [7]:

```
# Extract the Book of Romans
rom = util.extract(cdf, book='Rom')
rom.head()
```

Out[7]:

	book	book_no	chapter	verse	text	testament	category	cat	cat_no
27930	Rom	45	1	1	耶穌基督的僕人保羅、奉召為使徒、特派傳 神的福音。	1	Pauline Epistles	pau	7
27931	Rom	45	1	2	這福音是 神從前藉眾先知、在聖經上所應許的、	1	Pauline Epistles	pau	7
27932	Rom	45	1	3	論到他兒子我主耶穌基督。按肉體說、是從大衛後裔生的、	1	Pauline Epistles	pau	7
27933	Rom	45	1	4	按聖善的靈說、因從死裡復活、以大能顯明是 神的兒子。	1	Pauline Epistles	pau	7
27934	Rom	45	1	5	我們從他受了恩惠、並使徒的職分、在萬國之中叫人為他的名信服真道。	1	Pauline Epistles	pau	7

In [8]:

```
# Extract Verse 28 of Chapter 8 of Romans
rom8_28 = util.extract(cdf, book='Rom', chapter=8, verse=28)
rom8_28.text
```

Out[8]:

28143 我們曉得萬事都互相效力、叫愛 神的人得益處、就是按他旨意被召的人。  
Name: text, dtype: object

In [9]:

```
# Another way to extract Rom 8:28
rom8_28 = util.extract2(cdf, filter='Rom 8:28')
rom8_28.text
```

Out[9]:

28143 我們曉得萬事都互相效力、叫愛 神的人得益處、就是按他旨意被召的人。  
Name: text, dtype: object

In [10]:

```
# Extract Rom 8:1~10
rom8_1_10 = util.extract2(cdf, filter='Rom 8:1-10')
list(rom8_1_10.text)
```

Out[10]:

['如今那些在基督耶穌裡的、就不定罪了。',  
'因為賜生命聖靈的律、在基督耶穌裡釋放了我、使我脫離罪和死的律了。',  
'律法既因肉體軟弱、有所不能行的、\u3000神就差遣自己的兒子、成為罪身的形狀、作了贖罪祭、在肉體中定了罪案。',  
'使律法的義、成就在我們這不隨從肉體、只隨從聖靈的人身上。',  
'因為隨從肉體的人、體貼肉體的事。隨從聖靈的人、體貼聖靈的事。',  
'體貼肉體的就是死。體貼聖靈的乃是生命平安。',  
'原來體貼肉體的、就是與\u3000神為仇。因為不服\u3000神的律法、也是不能服。',  
'而且屬肉體的人、不能得\u3000神的喜歡。',  
'如果\u3000神的靈住在你們心裡、你們就不屬肉體、乃屬聖靈了。人若沒有基督的靈、就不是屬基督的。',  
'基督若在你們心裡、身體就因罪而死、心靈卻因義而活。']

2. Text Visualization - Word Cloud

In [11]:

```
# Extract the NT Scripture for Word Cloud
text_list = util.get_text_list(cdf[cdf.testament==1]) # Load New Testament Scripture
text_list[:10]
```

Out[11]:

['亞伯拉罕的後裔、大衛的子孫、耶穌基督的家譜。〔後裔子孫原文都作兒子下同〕',  
'亞伯拉罕生以撒。以撒生雅各。雅各生猶大和他的弟兄。',  
'猶大從他瑪氏生法勒斯和謝拉。法勒斯生希斯崙。希斯崙生亞蘭。',  
'亞蘭生亞米拿達。亞米拿達生拿順。拿順生撒門。',  
'撒門從喇合氏生波阿斯。波阿斯從路得氏生俄備得。俄備得生耶西。',  
'耶西生大衛王。大衛從烏利亞的妻子生所羅門。',  
'所羅門生羅波安。羅波安生亞比雅。亞比雅生亞撒。',  
'亞撒生約沙法。約沙法生約蘭。約蘭生烏西亞。',  
'烏西亞生約坦。約坦生亞哈斯。亞哈斯生希西家。',  
'希西家生瑪拿西。瑪拿西生亞們。亞們生約西亞。']

In [12]:

```
# White background with no image mask
viz.chi_wordcloud(text_list)
```

```
Building prefix dict from the default dictionary ...
Loading model from cache C:\Users\johnnyc\AppData\Local\Temp\jieba.cache
Loading Chinese vocabulary 'D:\Dev\Anaconda3\lib\site-packages\cwordtm\data\bible_vocab.txt' ...

Loading model cost 0.606 seconds.
Prefix dict has been built successfully.
Building prefix dict from D:\Dev\Anaconda3\lib\site-packages\cwordtm\dictionary\dict.txt.big.txt ...
Loading model from cache C:\Users\johnnyc\AppData\Local\Temp\jieba.u819f581d122236472c15f14136fc73d9.cache
Loading model cost 0.961 seconds.
Prefix dict has been built successfully.
D:\Dev\Anaconda3\lib\site-packages\wordcloud\wordcloud.py:106: MatplotlibDeprecationWarning: The get_cmap function w
as deprecated in Matplotlib 3.7 and will be removed two minor releases later. Use ``matplotlib.colormaps[name]`` or
``matplotlib.colormaps.get_cmap(obj)`` instead.
    self.colormap = plt.cm.get_cmap(colormap)
```



```
In [13]: # Use internal image mask
viz.chi_wordcloud(text_list, bg='black', image=1)
```

```
Building prefix dict from D:\Dev\Anaconda3\lib\site-packages\cwordtm\dictionary\dict.txt.big.txt ...
Loading model from cache C:\Users\johnnyc\AppData\Local\Temp\jieba.u819f581d122236472c15f14136fc73d9.cache
Loading Chinese vocabulary 'D:\Dev\Anaconda3\lib\site-packages\cwordtm\data\bible_vocab.txt' ...

Loading model cost 0.962 seconds.
Prefix dict has been built successfully.
D:\Dev\Anaconda3\lib\site-packages\wordcloud\wordcloud.py:106: MatplotlibDeprecationWarning: The get_cmap function w
as deprecated in Matplotlib 3.7 and will be removed two minor releases later. Use ``matplotlib.colormaps[name]`` or
``matplotlib.colormaps.get_cmap(obj)`` instead.
    self.colormap = plt.cm.get_cmap(colormap)
```



### 3. Text Summarization

```
In [14]: # Extract and summarize Chapter 8 of Romans
rom8 = util.extract2(cdf, 'Rom 8')
ta.summary_chi(rom8)
```

```
Building prefix dict from D:\Dev\Anaconda3\lib\site-packages\cwordtm\dictionary\dict.txt.big.txt ...
Loading model from cache C:\Users\johnnyc\AppData\Local\Temp\jieba.u819f581d122236472c15f14136fc73d9.cache
Loading Chinese vocabulary 'D:\Dev\Anaconda3\lib\site-packages\cwordtm\data\bible_vocab.txt' ...
```

```
Loading model cost 0.915 seconds.
```

```
Prefix dict has been built successfully.
```

```
Out[14]: ['因為隨從肉體的人、體貼肉體的事、隨從聖靈的人、體貼聖靈的事',
          '而且屬肉體的人、不能得神的喜歡',
          '如果神的靈住在你們心裡、你們就不屬肉體、乃屬聖靈了。人若沒有基督的靈、就不是屬基督的']
```

### 4. Pivot Table

#### Show Bible Scripture Statistics through a Pivot Table

```
In [15]: util.set_rows()
pivot.stat(cdf, chi=True)
```

```
Book category information can be shown by invoking 'util.bible_cat_info()'
```

Out[15]:

	chapter	verse	text
category	book_no	book	
Torah	1	Gen	50 1533 51460
	2	Exo	40 1213 40057
	3	Lev	27 859 29228
	4	Num	36 1288 41654
	5	Deu	34 959 35904
Sub-Total			187 5852 198303
OT History	6	Jos	24 658 25794
	7	Jug	21 618 24375
	8	Rut	4 85 3362
	9	1Sa	31 810 32218
	10	2Sa	24 695 26697
	11	1Ki	22 816 30530
	12	2Ki	25 719 29626
	13	1Ch	29 942 30231
	14	2Ch	36 822 33445
	15	Ezr	10 280 10094
	16	Neh	13 406 14739
	17	Est	10 167 6571
Sub-Total			249 7018 267682
Ketuvim	18	Job	42 1070 24294
	19	Psm	150 2461 64303
	20	Pro	31 915 19986
	21	Ecc	12 222 7228
	22	Son	8 117 3979
Sub-Total			243 4785 119790
Major Prophets	23	Isa	66 1292 50630
	24	Jer	52 1364 58005
	25	Lam	5 154 4721
	26	Eze	48 1273 49218
	27	Dan	12 357 14646
Sub-Total			183 4440 177220
Minor Prophets	28	Hos	14 197 7455
	29	Joe	3 73 2642
	30	Amo	9 146 5487
	31	Oba	1 21 838
	32	Jon	4 48 1675
	33	Mic	7 105 4307
	34	Nah	3 47 1716
	35	Hab	3 56 1974
	36	Zep	3 53 2216
	37	Hag	2 38 1484
	38	Zec	14 211 8235
	39	Mal	4 55 2658
Sub-Total			67 1050 40687
Gospel	40	Mat	28 1071 32774
	41	Mak	16 678 20482
	42	Luk	24 1151 35526

	category	book_no	book	chapter	verse	text
		43	Jhn	21	878	27378
	Sub-Total			89	3778	116160
	NT History	44	Act	28	1007	33762
	Sub-Total			28	1007	33762
	Pauline Epistles	45	Rom	16	433	14365
		46	1Co	16	437	14154
		47	2Co	13	257	9260
		48	Gal	6	149	4817
		49	Eph	6	155	4591
		50	Phl	4	104	3408
		51	Col	4	95	3312
		52	1Ts	5	89	2983
		53	2Ts	3	47	1624
		54	1Ti	6	113	3579
		55	2Ti	4	83	2639
		56	Tit	3	46	1544
		57	Phm	1	25	717
	Sub-Total			87	2033	66993
	General Epistles	58	Heb	13	303	10428
		59	Jas	5	108	3534
		60	1Pe	5	105	3905
		61	2Pe	3	61	2343
		62	1Jn	5	105	3779
		63	2Jn	1	13	463
		64	3Jn	1	15	478
		65	Jud	1	25	1030
	Sub-Total			34	735	25960
	Apocalypse	66	Rev	22	404	15606
	Sub-Total			22	404	15606
	Total			1189	31102	1062163

```
In [16]: util.bible_cat_info()
```

	category	cat	book_list	nbooks
0	Torah	tor	Gen Exo Lev Num Deu	5
1	OT History	oth	Jos Jug Rut 1Sa 2Sa 1Ki 2Ki 1Ch 2Ch Ezr Neh Est	12
2	Ketuvim	ket	Job Psm Pro Ecc Son	5
3	Major Prophets	map	Isa Jer Lam Eze Dan	5
4	Minor Prophets	mip	Hos Joe Amo Oba Jon Mic Nah Hab Zep Hag Zec Mal	12
5	Gospel	gos	Mat Mak Luk Jhn	4
6	NT History	nth	Act	1
7	Pauline Epistles	pau	Rom 1Co 2Co Gal Eph Phl Col 1Ts 2Ts 1Ti 2Ti Ti...	13
8	General Epistles	epi	Heb Jas 1Pe 2Pe 1Jn 2Jn 3Jn Jud	8
9	Apocalypse	apo	Rev	1

5. OT Quotes

Identify Cited Sources in OT Scripture for some NT verses

```
In [17]: rom10 = util.extract2(cdf, 'Rom 10')
         quot.show_quot(rom10, lang='chi')
```

```
Building prefix dict from D:\Dev\Anaconda3\lib\site-packages\cwordtm\dictionary\dict.txt.big.txt ...
Loading model from cache C:\Users\johnnyc\AppData\Local\Temp\jieba.u819f581d122236472c15f14136fc73d9.cache
Loading Chinese vocabulary 'D:\Dev\Anaconda3\lib\site-packages\cwordtm\data\bible_vocab.txt' ...
Loading file 'D:\Dev\Anaconda3\lib\site-packages\cwordtm\data\cuw.csv' ...
Loading file 'D:\Dev\Anaconda3\lib\site-packages\cwordtm\data\book_categories.csv' ...
( 1) 羅 10:5 摩西寫著說、『人若行那出於律法的義、就必因此活著。』
```

```
Loading model cost 0.917 seconds.
```

```
Prefix dict has been built successfully.
```

```
( 2) 羅 10:6 惟有出於信心的義如此說、『你不要心裡說、誰要升到天上去呢。就是要領下基督來。
( 3) 羅 10:8 他到底怎麼說呢。他說、『這道離你不遠、正在你口裡、在你心裡。』就是我們所傳信主的道。
( 4) 羅 10:11 經上說、『凡信他的人、必不至於羞愧。』
( 5) 羅 10:13 因為『凡求告主名的、就必得救。』
( 6) 羅 10:15 若沒有奉差遣、怎能傳道呢。如經上所記、『報福音傳喜信的人、他們的腳蹤何等佳美。』
( 7) 羅 10:16 只是人沒有都聽從福音。因為以賽亞說、『主阿、我們所傳的有誰信呢。』
    -> 0.6723 賽 53:1 我們所傳的、〔或作所傳與我們的〕有誰信呢。耶和華的膀臂向誰顯露呢。
( 8) 羅 10:18 但我說、人沒有聽見麼。誠然聽見了。『他們的聲音傳遍天下、他們的言語傳到地極。』
( 9) 羅 10:19 我再說、以色列人不知道麼。先有摩西說、『我要用那不成子民的、惹動你們的憤恨。我要用那無知的民、觸動你們的怒氣。』
    -> 0.5403 申 32:21 他們以那不算為神的、觸動我的憤恨、以虛無的神、惹了我的怒氣、我也要以那不成子民的、觸動他們的憤恨、
以愚昧的國民、惹了他們的怒氣。
(10) 羅 10:20 又有以賽亞放膽說、『沒有尋找我的、我叫他們遇見。沒有訪問我的、我向他們顯現。』
    -> 0.6651 賽 65:1 素來沒有訪問我的、現在求問我。沒有尋找我的、我叫他們遇見。沒有稱為我名下的、我對他們說、我在這裡、
我在這裡。
(11) 羅 10:21 至於以色列人、他說、『我整天伸手招呼那悖逆頂嘴的百姓。』
    -> 0.6086 賽 65:2 我整天伸手招呼那悖逆的百姓、他們隨自己的意念行不善之道。
```

## 6. Topic Modeling

### LDA Modeling

```
In [18]: # Build an LDA Model on the OT Scripture
         lda = tm.Lda_process(bible, cat=1, chi=True, eval=True, timing=True)
```

```

Loading Bible 'D:\Dev\Anaconda3\lib\site-packages\cwordtm\data\cuV.csv' ...
Corpus loaded!
Text preprocessed!
Text trained!
If no visualization is shown,
  you may execute the following commands to show the visualization:
  > import pyLDAvis
  > pyLDAvis.display(lda.vis_data)
Visualization prepared!

```

Topics from LDA Model:

```

[(0,
 '0.053*"耶和華" + 0.028*"說" + 0.015*"神" + 0.012*"必" + 0.008*"以色列" + 0.005*"主" + '
 '0.005*"沒有" + 0.005*"百姓" + 0.005*"埃及" + 0.005*"這是"',),
 (1,
 '0.007*"神" + 0.006*"耶和華" + 0.006*"說" + 0.006*"必" + 0.005*"祭司" + 0.005*"作" + '
 '0.004*"王" + 0.004*"約櫃" + 0.004*"利未人" + 0.004*"以色列"',),
 (2,
 '0.026*"耶和華" + 0.015*"不可" + 0.010*"說" + 0.010*"潔淨" + 0.010*"祭司" + 0.009*"兒子" '
 '+ 0.008*"以色列" + 0.008*"作" + 0.007*"喫" + 0.006*"面前"',),
 (3,
 '0.020*"公" + 0.015*"作" + 0.014*"一隻" + 0.009*"羊羔" + 0.008*"郊野" + 0.008*"燔祭" + '
 '0.008*"兒子" + 0.007*"屬城" + 0.007*"素祭" + 0.007*"耶和華"',),
 (4,
 '0.026*"耶和華" + 0.014*"必" + 0.012*"神" + 0.009*"求" + 0.007*"中" + 0.007*"說" + '
 '0.006*"惡人" + 0.005*"願" + 0.005*"我要" + 0.004*"公義"',),
 (5,
 '0.027*"耶和華" + 0.026*"說" + 0.016*"以色列" + 0.014*"王" + 0.012*"神" + 0.009*"所羅門" '
 '+ 0.008*"兒子" + 0.007*"耶路撒冷" + 0.007*"僕人" + 0.006*"面前"',),
 (6,
 '0.017*"子孫" + 0.011*"王" + 0.009*"猶大" + 0.008*"說" + 0.008*"埃及" + 0.008*"耶和華" '
 '+ 0.005*"末底改" + 0.005*"法老" + 0.005*"巴比倫" + 0.004*"祭司"',),
 (7,
 '0.044*"說" + 0.016*"耶和華" + 0.010*"大衛" + 0.009*"兒子" + 0.008*"王" + 0.008*"以色列" '
 '+ 0.007*"掃羅" + 0.005*"僕人" + 0.005*"沒有" + 0.005*"父親"',),
 (8,
 '0.033*"耶和華" + 0.027*"以色列" + 0.016*"摩西" + 0.016*"說" + 0.015*"兒子" + '
 '0.007*"吩咐" + 0.007*"王" + 0.006*"亞倫" + 0.006*"支派" + 0.006*"中"',),
 (9,
 '0.027*"兒子" + 0.010*"說" + 0.010*"生" + 0.009*"耶和華" + 0.007*"必" + 0.004*"中" + '
 '0.004*"族" + 0.004*"以色列" + 0.003*"摩押" + 0.003*"死"')]
```

Model Evaluation Scores:

```

Coherence: 0.3556461157424597
Perplexity: -8.22263582330936
Topic diversity: 0.001295926134366632
Topic size distribution: 0.0034263338228095936

```

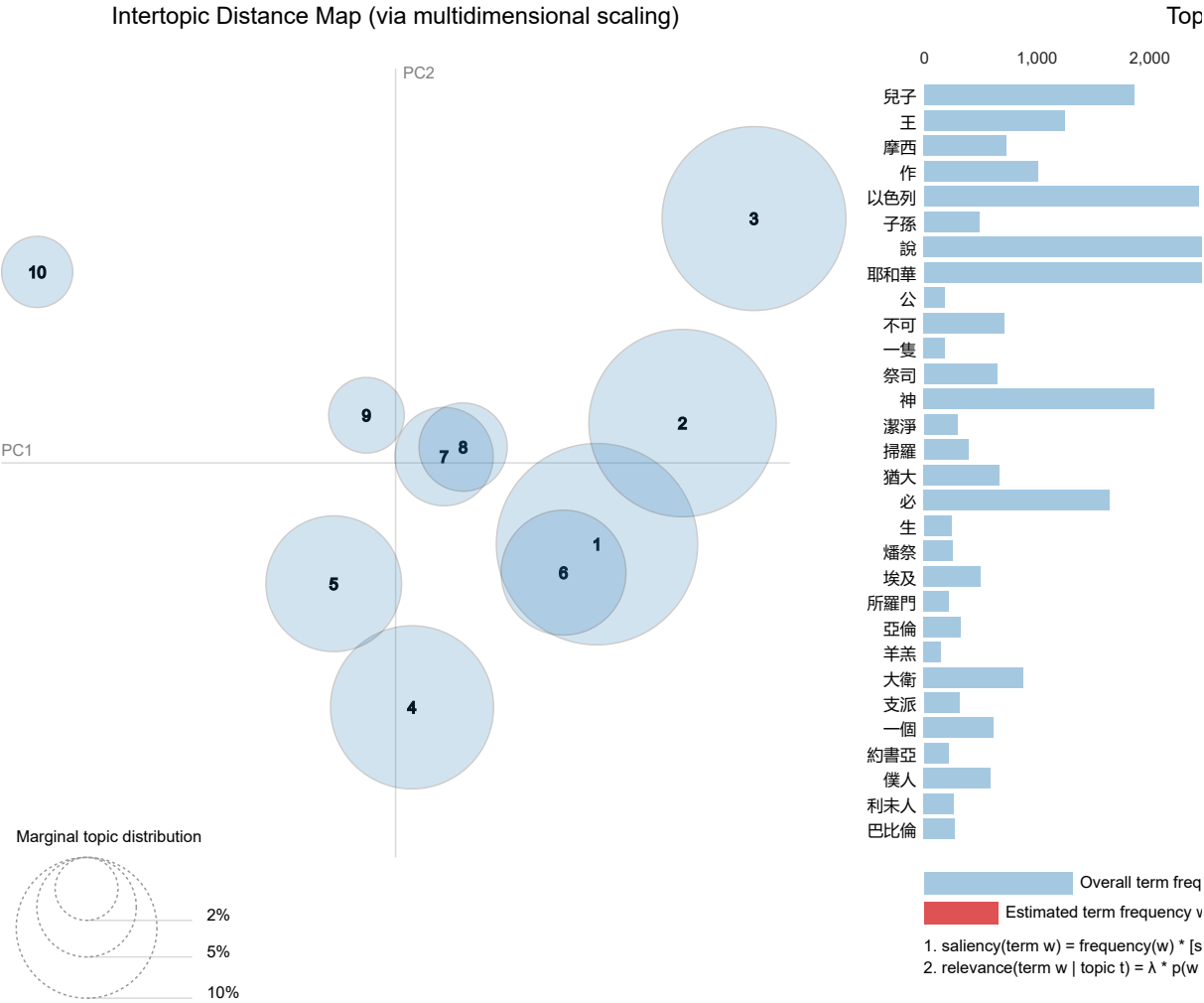
Finished 'lda\_process' in 33.9946 secs

```

In [19]: # Show LDA Model Visualization
import pyLDAvis
pyLDAvis.display(lda.vis_data)

```





## BERTopic Modeling

```
In [20]: # Build a BERTopic Model on the OT Scripture
btm = tm.btm_process(bible, cat=1, chi=True, eval=True, timing=True)

Loading Bible 'D:\Dev\Anaconda3\lib\site-packages\cwordtm\data\cuw.csv' ...
Corpus loaded!
Chinese text preprocessed!

Some weights of the model checkpoint at bert-base-chinese were not used when initializing BertModel: ['cls.predictio
ns.transform.LayerNorm.weight', 'cls.predictions.transform.dense.bias', 'cls.predictions.bias', 'cls.seq_relationshi
p.bias', 'cls.predictions.transform.dense.weight', 'cls.predictions.decoder.weight', 'cls.predictions.transform.Laye
rNorm.bias', 'cls.seq_relationship.weight']
- This IS expected if you are initializing BertModel from the checkpoint of a model trained on another task or with
another architecture (e.g. initializing a BertForSequenceClassification model from a BertForPreTraining model).
- This IS NOT expected if you are initializing BertModel from the checkpoint of a model that you expect to be exactl
y identical (initializing a BertForSequenceClassification model from a BertForSequenceClassification model).
D:\Dev\Anaconda3\lib\site-packages\hdbscan\hdbscan_.py:1170: DeprecationWarning: `alltrue` is deprecated as of NumPy
1.25.0, and will be removed in NumPy 2.0. Please use `all` instead.
  self._all_finite = is_finite(X)
```

Text trained!

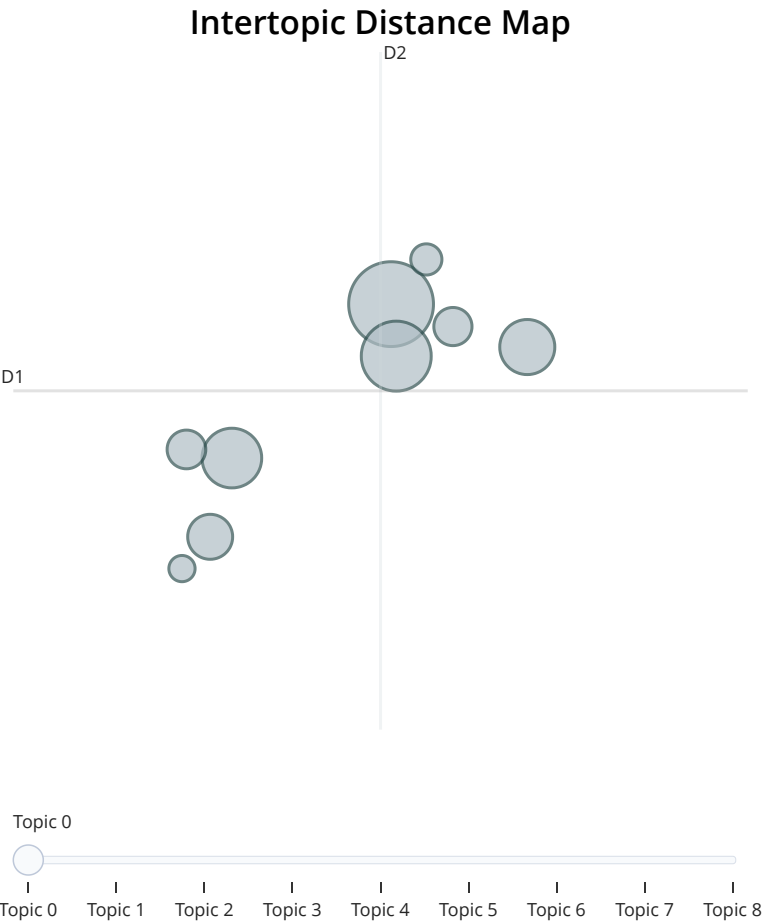
Topics from BERTopic Model:  
Topic 0: 耶和華 | 兒子 | 以色列 | 大衛 | 沒有 | 雅各 | 僕人 | 父親 | 一個 | 知道  
Topic 1: 兒子 | 耶和華 | 以色列 | 大衛 | 巴比倫 | 耶路撒冷 | 子孫 | 猶大 | 萬軍 | 非利士人  
Topic 2: 耶和華 | 不可 | 雅各 | 埃及 | 沒有 | 面前 | 地上 | 以色列 | 知道 | 亞伯拉罕  
Topic 3: 祭司 | 耶和華 | 兒子 | 以色列 | 潔淨 | 利未人 | 亞倫 | 子孫 | 面前 | 吩咐  
Topic 4: 耶和華 | 讚美 | 永遠 | 的詩 | 伶長 | 公義 | 之詩 | 交與 | 大衛 | 上行  
Topic 5: 智慧 | 惡人 | 愚昧 | 義人 | 知識 | 正直 | 言語 | 愚妄 | 聰明 | 知道  
Topic 6: 摩西 | 耶和華 | 支派 | 屬城 | 吩咐 | 郊野 | 以色列 | 一個 | 瑪拿西 | 兒子  
Topic 7: 耶和華 | 摩西 | 亞倫 | 以色列 | 不可 | 素祭 | 火祭 | 基路伯 | 獻給 | 活物  
Topic 8: 歌頌 | 耶和華 | 我要 | 歡呼 | 稱謝 | 詩歌 | 稱讚 | 萬民 | 永遠 | 讚美

Model Evaluation Scores:  
Coherence: 0.372286323281463

BERTopic Model Visualization:

```
D:\Dev\Anaconda3\lib\site-packages\plotly\io\_renderers.py:395: DeprecationWarning:
distutils Version classes are deprecated. Use packaging.version instead.

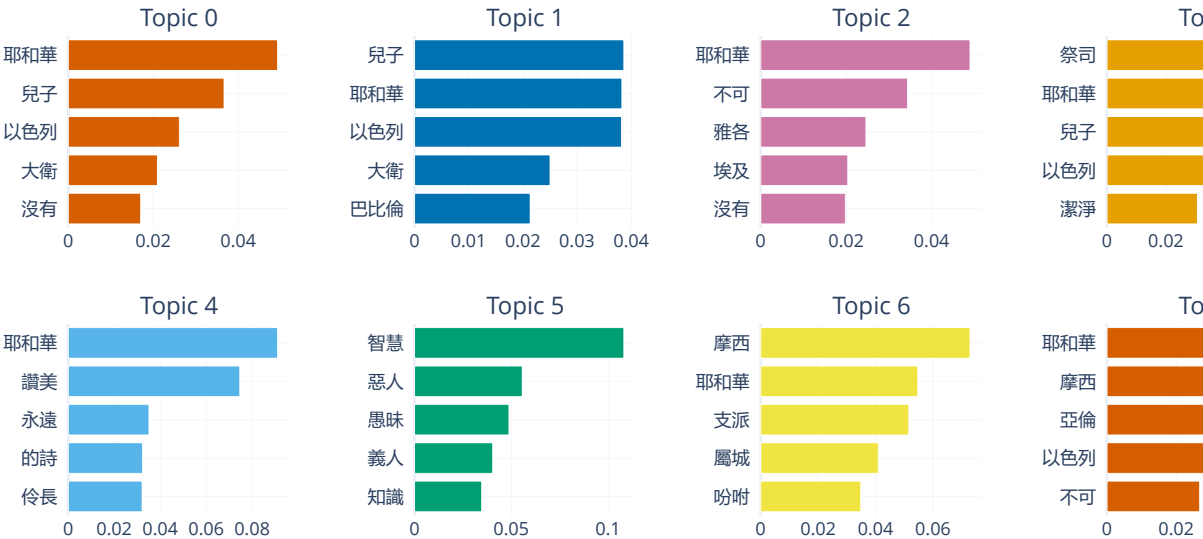
D:\Dev\Anaconda3\lib\site-packages\plotly\io\_renderers.py:395: DeprecationWarning:
distutils Version classes are deprecated. Use packaging.version instead.
```



```
D:\Dev\Anaconda3\lib\site-packages\plotly\io\_renderers.py:395: DeprecationWarning:
distutils Version classes are deprecated. Use packaging.version instead.

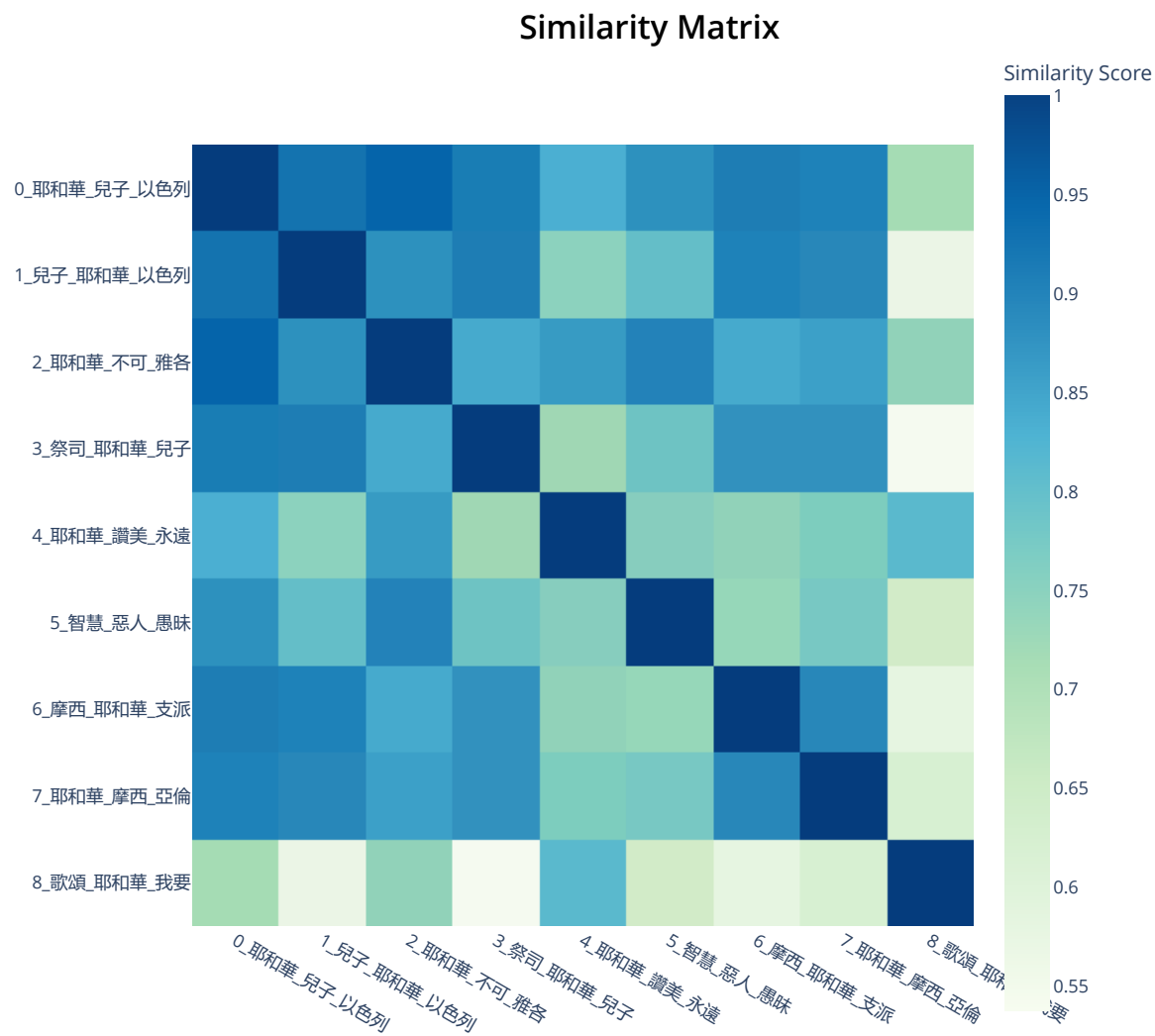
D:\Dev\Anaconda3\lib\site-packages\plotly\io\_renderers.py:395: DeprecationWarning:
distutils Version classes are deprecated. Use packaging.version instead.
```

Topic Word Scores



```
D:\Dev\Anaconda3\lib\site-packages\plotly\io\_renderers.py:395: DeprecationWarning:
distutils Version classes are deprecated. Use packaging.version instead.

D:\Dev\Anaconda3\lib\site-packages\plotly\io\_renderers.py:395: DeprecationWarning:
distutils Version classes are deprecated. Use packaging.version instead.
```



If no visualization is shown,  
you may execute the following commands one-by-one:  
btm.model.visualize\_topics()  
btm.model.visualize\_barchart()  
btm.model.visualize\_heatmap()

Finished 'btm\_process' in 84.5544 secs

### NMF Modeling

```
In [21]: nmf = tm.nmf_process(bible, num_topics=8, cat=1, chi=True, eval=True, timing=True)
```

Loading Bible 'D:\Dev\Anaconda3\lib\site-packages\cwordtm\data\cuV.csv' ...  
Corpus loaded!  
Text preprocessed!

D:\Dev\Anaconda3\lib\site-packages\gensim\models\nmf.py:578: DeprecationWarning:  
Please import `csc\_matrix` from the `scipy.sparse` namespace; the `scipy.sparse.csc` namespace is deprecated and will be removed in SciPy 2.0.0.

D:\Dev\Anaconda3\lib\site-packages\gensim\models\nmf.py:607: DeprecationWarning:  
Please import `csc\_matrix` from the `scipy.sparse` namespace; the `scipy.sparse.csc` namespace is deprecated and will be removed in SciPy 2.0.0.

D:\Dev\Anaconda3\lib\site-packages\gensim\models\nmf.py:620: DeprecationWarning:  
Please import `csc\_matrix` from the `scipy.sparse` namespace; the `scipy.sparse.csc` namespace is deprecated and will be removed in SciPy 2.0.0.

Text trained!

Topics-Words from NMF Model:

Topic 1:

說 (0.088083)  
耶和華 (0.039516)  
以色列 (0.012834)  
沒有 (0.007518)  
面前 (0.007325)  
喫 (0.006688)  
見 (0.006139)  
百姓 (0.005975)  
看見 (0.005931)  
先知 (0.005677)

Topic 2:

兒子 (0.047777)  
耶和華 (0.023131)  
作 (0.011530)  
說 (0.009809)  
耶路撒冷 (0.009204)  
僕人 (0.007863)  
生 (0.007701)  
巴比倫 (0.005934)  
父親 (0.005816)  
摩押 (0.005335)

Topic 3:

必 (0.023225)  
耶和華 (0.020787)  
摩西 (0.015048)  
神 (0.014480)  
支派 (0.007139)  
不可 (0.006407)  
吩咐 (0.006254)  
亞倫 (0.006091)  
以色列 (0.005665)  
約書亞 (0.005176)

Topic 4:

王 (0.038957)  
兒子 (0.023522)  
以色列 (0.021094)  
大衛 (0.016886)  
掃羅 (0.015428)  
中 (0.007632)  
一個 (0.006893)  
說 (0.005689)  
所羅門 (0.004547)  
非利士人 (0.004371)

Topic 5:

耶和華 (0.068373)  
公 (0.016609)  
作 (0.011502)  
一隻 (0.011399)  
祭司 (0.011106)  
子孫 (0.009801)  
以色列 (0.008523)  
神 (0.008372)  
這是 (0.007268)  
猶大 (0.007225)

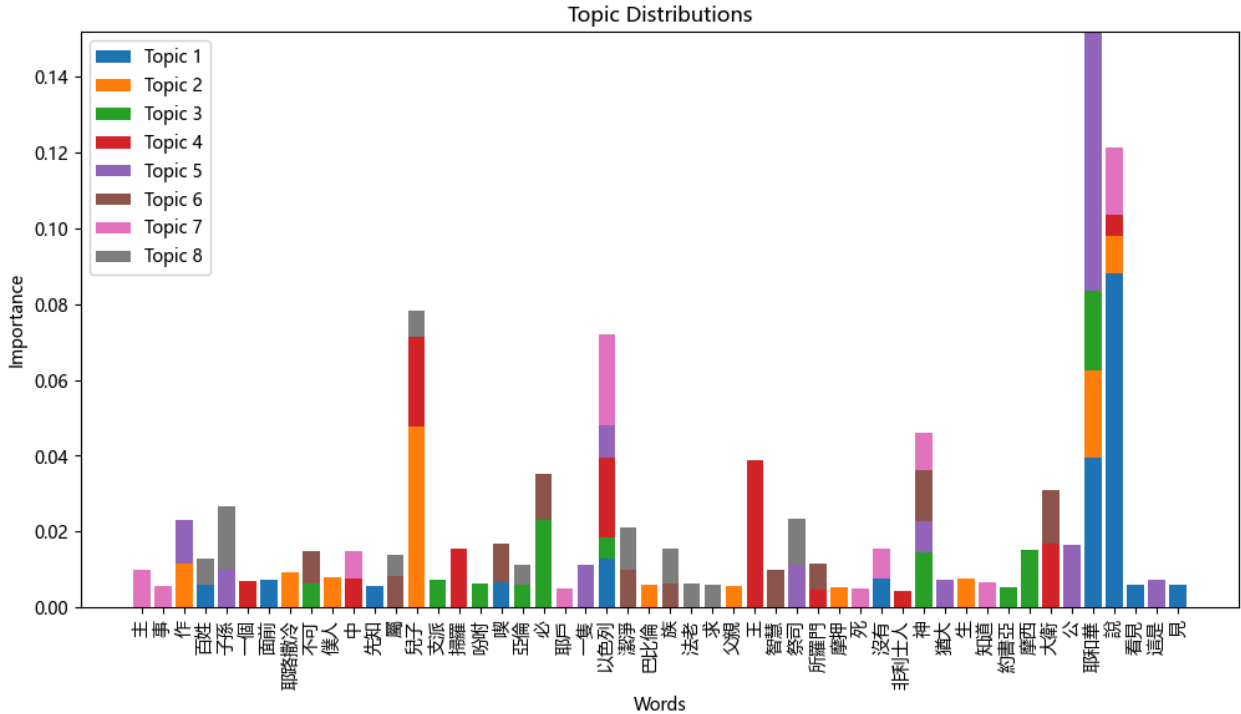
Topic 6:

大衛 (0.013946)  
神 (0.013238)  
必 (0.012048)  
喫 (0.010017)  
潔淨 (0.009914)  
智慧 (0.009857)  
不可 (0.008521)  
屬 (0.008256)  
所羅門 (0.007096)  
族 (0.006195)

Topic 7:

以色列 (0.023954)  
說 (0.017891)  
神 (0.009990)  
主 (0.009987)  
沒有 (0.007934)  
中 (0.007325)  
知道 (0.006654)  
事 (0.005543)  
死 (0.004957)  
耶戶 (0.004867)

Topic 8:  
子孫 (0.017001)  
祭司 (0.012394)  
潔淨 (0.011287)  
族 (0.009211)  
百姓 (0.007058)  
兒子 (0.006958)  
法老 (0.006163)  
求 (0.006089)  
屬 (0.005543)  
亞倫 (0.005256)



Model Evaluation Scores:  
Coherence: 0.3160562514672284  
Topic diversity: 0.0011409047343307346  
Topic size distribution: 0.0017813404586951682

Finished 'nmf\_process' in 19.9662 secs

### Save NMF Model

```
In [22]: nmf.save("models/nmf_cuv")
```

NMF model has been saved: 'models/nmf\_cuv\_model.gensim' and 'models/nmf\_cuv\_dictionary.gensim'

### Load NMF Model

```
In [23]: nmf2 = tm.NMF("", nmf.num_topics)
nmf2.model, nmf2.dictionary = nmf2.load("models/nmf_cuv")
nmf2.show_topics_words()
```

Topics-Words from NMF Model:

Topic 1:

說 (0.088083)  
耶和華 (0.039516)  
以色列 (0.012834)  
沒有 (0.007518)  
面前 (0.007325)  
喫 (0.006688)  
見 (0.006139)  
百姓 (0.005975)  
看見 (0.005931)  
先知 (0.005677)

Topic 2:

兒子 (0.047777)  
耶和華 (0.023131)  
作 (0.011530)  
說 (0.009809)  
耶路撒冷 (0.009204)  
僕人 (0.007863)  
生 (0.007701)  
巴比倫 (0.005934)  
父親 (0.005816)  
摩押 (0.005335)

Topic 3:

必 (0.023225)  
耶和華 (0.020787)  
摩西 (0.015048)  
神 (0.014480)  
支派 (0.007139)  
不可 (0.006407)  
吩咐 (0.006254)  
亞倫 (0.006091)  
以色列 (0.005665)  
約書亞 (0.005176)

Topic 4:

王 (0.038957)  
兒子 (0.023522)  
以色列 (0.021094)  
大衛 (0.016886)  
掃羅 (0.015428)  
中 (0.007632)  
一個 (0.006893)  
說 (0.005689)  
所羅門 (0.004547)  
非利士人 (0.004371)

Topic 5:

耶和華 (0.068373)  
公 (0.016609)  
作 (0.011502)  
一隻 (0.011399)  
祭司 (0.011106)  
子孫 (0.009801)  
以色列 (0.008523)  
神 (0.008372)  
這是 (0.007268)  
猶大 (0.007225)

Topic 6:

大衛 (0.013946)  
神 (0.013238)  
必 (0.012048)  
喫 (0.010017)  
潔淨 (0.009914)  
智慧 (0.009857)  
不可 (0.008521)  
屬 (0.008256)  
所羅門 (0.007096)  
族 (0.006195)

Topic 7:

以色列 (0.023954)  
說 (0.017891)  
神 (0.009990)  
主 (0.009987)  
沒有 (0.007934)  
中 (0.007325)  
知道 (0.006654)  
事 (0.005543)  
死 (0.004957)  
耶戶 (0.004867)

Topic 8:

子孫 (0.017001)  
祭司 (0.012394)  
潔淨 (0.011287)  
族 (0.009211)  
百姓 (0.007058)  
兒子 (0.006958)  
法老 (0.006163)  
求 (0.006089)  
屬 (0.005543)  
亞倫 (0.005256)