

TextCritical.net - Bug #1247

Word lookup fails to return all forms

03/07/2016 05:27 AM - Luke Murphey

Status:	Closed	Start date:	03/07/2016
Priority:	Normal	Due date:	
Assignee:	Luke Murphey	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:	1.3		
Description			
Here are some cases that don't work			
<ul style="list-style-type: none">• νομος (νομον works)• έργα (doesn't return all forms though έργον returns many of them)			
This means the search engine isn't expanding words into all of the possible forms.			

History

#1 - 03/09/2016 05:25 AM - Luke Murphey

- Subject changed from Word lookup fails to Word lookup fails to return all forms

- Description updated

#2 - 03/09/2016 08:19 PM - Luke Murphey

Call stack:

- GreekVariations::get_variations()
- reader.utils.get_all_related_forms()
- reader.utils.get_word_descriptions()

#3 - 03/09/2016 08:25 PM - Luke Murphey

νομος has a word form:

<http://127.0.0.1:8080/admin/reader/wordform/?q=%CE%BD%CF%8C%CE%BC%CE%BF%CF%82>

```
from reader.models import *
descriptions = WordDescription.objects.filter( word_form__form="νόμος")
```

```
lemmas = []
```

```
for d in descriptions:
    if d.lemma not in lemmas:
        lemmas.append(d.lemma)
```

```
for l in lemmas:
    matching_descs = WordDescription.objects.filter(lemma=l)
    for m in matching_descs:
        if m.word_form not in word_forms:
            word_forms.append( m.word_form )
```

#4 - 03/09/2016 08:27 PM - Luke Murphey

I'm seeing uppercase characters in the forms:

```
[<WordForm: Νόμε>, <WordForm: Νόμοι>, <WordForm: Νόμοιν>, <WordForm: Νόμοιο>, <WordForm: Νόμοις>, <WordForm: Νόμοισι>, <WordForm: Νόμοισιν>, <WordForm: Νόμον>, <WordForm: Νόμος>, <WordForm: Νόμου>, <WordForm: Νόμους>, <WordForm: Νόμω>, <WordForm: Νόμωι>, <WordForm: Νόμων>, <WordForm: Νόμως>, <WordForm: Νόμω>, <WordForm: νόμε>, <WordForm: νόμοι>, <WordForm: νόμοιν>, <WordForm: νόμοιο>, <WordForm: νόμοις>, <WordForm: νόμοισι>, <WordForm: νόμοισιν>, <WordForm: νόμον>, <WordForm: νόμος>, <WordForm: νόμου>, <WordForm: νόμους>, <WordForm: νόμω>, <WordForm: νόμωι>, <WordForm: νόμων>, <WordForm: νόμως>, <WordForm: νόμω>]
```

#5 - 03/09/2016 08:27 PM - Luke Murphey

- Status changed from New to In Progress

- % Done changed from 0 to 20

#6 - 03/09/2016 11:20 PM - Luke Murphey

I'm finding that a search doesn't return any word descriptions: http://127.0.0.1:8080/api/word_forms/%CE%BD%CF%8C%CE%BC%CE%BF%CF%82

This works for some reason though:

```
from reader.utils import *
get_word_descriptions(u"νόμος")
```

#7 - 03/09/2016 11:31 PM - Luke Murphey

I think this has to do with the two forms of sigmas.

#8 - 03/09/2016 11:36 PM - Luke Murphey

Removing `Greek.beta_code_to_unicode()` from line 426 in `get_variations()` makes this work.

#9 - 03/10/2016 04:42 AM - Luke Murphey

- Status changed from In Progress to Closed

- % Done changed from 20 to 100