

# The Texai Lexicon RDF File Explained

Stephen L. Reed  
December 12, 2007

The texai lexicon [RDF](#) (Resource Description Framework) file contains information about these six English lexical entities:

- word (e.g. rose)
- word form (e.g. roses, the plural form of rose)
- word sense (e.g. a [shrub](#) of the genus *Rosa*, with red, pink, white or yellow [flowers](#))
- sample phrase (e.g. "I pricked my hand on a rose.")
- ARPABET pronunciation (e.g. R OW1 Z)
- ARPABET phoneme (e.g. R)

The Texai lexicon N3 file contains the same content in an easy to parse, non-XML format described at the end of this document.

## Word

RDF is based upon XML. Below is the RDF entry for the word *run*, with the salient sections indexed. The order of the sub-elements within the rdf:Description element is arbitrarily determined by the Sesame export software that creates this file.

```
[1] <rdf:Description
rdf:about="http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWord_6f4d0bcc-f19a-482c-
ba0f-ca237714568c">

[2] <rdf:type rdf:resource="http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWord"/>

[3] <rdf:type rdf:resource="http://texai.org/texai/TexaiEnglishWord-IsNotOpenCycPrettyString"/>

[4] <domainEntityClassName
xmlns="http://texai.org/texai/">org.texai.lexicon.domainEntity.TexaiEnglishWord</domainEntityClassName>

[5] <texaiWordNetEnglishWord xmlns="http://texai.org/texai/"
rdf:resource="http://texai.org/texai/WordNetEnglishWord_4338803"/>

[6] <texaiEnglishWordSenseListForWord xmlns="http://texai.org/texai/"
rdf:nodeID="node12mnjq6erx211978"/>

[7] <texaiLemma xmlns="http://texai.org/texai/">run</texaiLemma>

[8] <texaiWiktionaryEnglishWord xmlns="http://texai.org/texai/"
rdf:resource="http://texai.org/texai/org.texai.wiktionary.domainEntity.WiktionaryEnglish
Word_42439c6d-c755-4abb-95cb-e4ff2232f1e8"/>

[9] </rdf:Description>
```

The RDF Description element [1] identifies this entry. In RDF, each concept is represented by a URI (Uniform Resource Identifier). This entry's URI is:

[http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWord\\_6f4d0bcc-f19a-482c-ba0f-ca237714568c](http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWord_6f4d0bcc-f19a-482c-ba0f-ca237714568c)

Texai URIs do not point to an actual web page. As allowed by RDF, they simply uniquely name a resource within the texai.org namespace. Some other Texai RDF entry can refer to this one via this URI.

Element [2] of the entry indicates that this entry is a TexaiEnglishWord.

Element [3] of the entry indicates that this entry is not an OpenCyc pretty string, which means that this entry is not solely derived from the OpenCyc lexicon.

Element [4] gives the Java class from which an instance is constructed when this entry is retrieved from an RDF store into a Java application:

```
org.texai.lexicon.domainEntity.TexaiEnglishWord
```

Element [5] identifies the corresponding entry in the WordNet lexicon RDF file, which is available from the Texai SourceForge project site as a separate download.

Element [6] points to the head of an RDF list containing the word sense entries for this word. The node at the head of the list has the id: node12mnjq6erx211978.

```
<rdf:Description rdf:nodeID="node12mnjq6erx211978">
  [A] <rdf:first
    rdf:resource="http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWordSense_a5cde1
    3b-e8e8-43e8-88de-48980edblada"/>
  [B] <rdf:rest rdf:nodeID="node12mnjq6erx211979"/>
</rdf:Description>

<rdf:Description rdf:nodeID="node12mnjq6erx211979">
<rdf:first
  rdf:resource="http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWordSense_c42dd1
  04-b3bb-466b-8db1-aa92295ff02c"/>
<rdf:rest rdf:nodeID="node12mnjq6erx211980"/>
</rdf:Description>

. . .

<rdf:Description rdf:nodeID="node12mnjq6erx212038">
<rdf:first
  rdf:resource="http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWordSense_b6e12f
  3e-3cb5-4d20-8f8d-f220c5f2aa36"/>
  [C] <rdf:rest rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#nil"/>
</rdf:Description>
```

The rdf:first element [A] indicates the URI of the next word sense entry. The rdf:rest element [B] indicates the URI of the next list entry. The end of the list is indicated by the rdf:rest element [C] value of nil.

Element [7] gives the lemma, or stem, of this word. In the example the lemma is *run*.

Element [8] identifies the corresponding entry in the Wiktionary lexicon RDF file, which is available from the Texai SourceForge project site as a separate download.

XML tag [9] closes this entry.

## Word Sense

Below is a word sense entry for the verb *run* whose gloss is “direct or control; projects, businesses, etc.”.

```
[1] <rdf:Description
  rdf:about="http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWordSense_562bba32-98bc-
  483e-94cb-b0f848e3e200">

[2] <rdf:type
  rdf:resource="http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWordSense"/>

[3] <rdf:type rdf:resource="http://texai.org/texai/TexaiEnglishWordSense-IsNotOpenCyc"/>

[4] <domainEntityClassName
  xmlns="http://texai.org/texai/">org.texai.lexicon.domainEntity.TexaiEnglishWordSense</domainEntityClass
  Name>

[5] <texaiWordSenseWordNetSynset xmlns="http://texai.org/texai/"
  rdf:resource="http://texai.org/texai/WordNetSynset_2606353"/>

[6] <texaiWordSenseGloss xmlns="http://texai.org/texai/">direct or control; projects, businesses,
  etc.</texaiWordSenseGloss>

[7] <texaiWordSenseNumber xmlns="http://texai.org/texai/"
  rdf:datatype="http://www.w3.org/2001/XMLSchema#int">38</texaiWordSenseNumber>

[8] <texaiWordSenseSamplePhrase xmlns="http://texai.org/texai/"
  rdf:resource="http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiSamplePhrase_1a
  cd3d08-46a3-4396-8e59-3971d871a39e"/>

[9] <texaiWordSenseSpeechPart xmlns="http://texai.org/texai/"
```

```

rdf:resource="http://sw.cyc.com/2006/07/27/cyc/Verb"/>

[10] <texaiEnglishWordSenseForWord xmlns="http://texai.org/texai/"
rdf:resource="http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWord_6
f4d0bcc-f19a-482c-ba0f-ca237714568c"/>

[11] <wordNetCategory xmlns="http://texai.org/texai/"
rdf:resource="http://texai.org/texai/org.texai.wordnet.domain.entity.WordNetCategory_2261253"/>

[12] <texaiWordSenseBasicSpeechPart xmlns="http://texai.org/texai/"
rdf:resource="http://sw.cyc.com/2006/07/27/cyc/Verb"/>

[13] </rdf:Description>

```

Element [1] identifies this word sense entry with the URI:

[http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWordSense\\_562bba32-98bc-483e-94cb-b0f848e3e200](http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWordSense_562bba32-98bc-483e-94cb-b0f848e3e200)

Element [2] of the entry indicates that this entry is a TexaiEnglishWordSense.

Element [3] of the entry indicates whether this is an OpenCyc word sense.

Element [4] gives the Java class from which an instance is constructed when this entry is retrieved from an RDF store into a Java application:

```
org.texai.lexicon.domainEntity.TexaiEnglishWordSense
```

Optional element [5] of this entry indicates the corresponding WordNet 2.1 synset number.

Element [6] gives the Java class from which an instance is constructed when this entry is retrieved from an RDF store into a Java application:

```
org.texai.lexicon.domainEntity.TexaiEnglishWordSense
```

Element [7] indicates the word sense number, which indexes this word sense within the lemma.

Optional element [8] gives the URI of an associated sample phrase for this word sense.

Element [9] indicates the part-of-speech URI for this word sense. These URIs are either OpenCyc lexical concepts or elaborated Texai lexical concepts (e.g. Verb, Noun, etc.). A complete list of the speech parts can be found in: `/doc/inflection-speech-parts.txt`.

Element [10] gives the URI identifying the Texai English word entry for which this entry is a word sense.

Optional element [11] gives the URI for an associated WordNet 2.1 category for this word sense. The WordNet category entry is stored in the separate WordNet RDF file.

Element [12] indicates the basic part-of-speech URI for this word sense. These parts of speech are those used by WordNet (i.e. Noun, Verb, Adjective and Adverb).

XML tag [13] closes this word sense entry.

## Word Form

Below is a word form entry for the word *runs*.

```

[1] <rdf:Description
rdf:about="http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWordForm_f7c23c53-075d-
4726-be9d-f4c7d01fa92d">

[2] <rdf:type
rdf:resource="http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWordForm"/>

[3] <rdf:type rdf:resource="http://texai.org/texai/TexaiEnglishWordForm-IsNotOpenCycPrettyString"/>

[4] <domainEntityClassName

```

```

xmlns="http://texai.org/texai/">org.texai.lexicon.domainEntity.TexaiEnglishWordForm</domainEntityClassName>

[5] <texaiEnglishWordFormForWord xmlns="http://texai.org/texai/"
rdf:resource="http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWord_6f
4d0bcc-f19a-482c-ba0f-ca237714568c"/>

[6] <texaiWordForm xmlns="http://texai.org/texai/">runs</texaiWordForm>

[7] <texaiARPABETPronunciation xmlns="http://texai.org/texai/"
rdf:resource="http://texai.org/texai/org.texai.cmudict.domainEntity.ARPABETPronunciation
_e2d403f8-1a43-4d5e-9edd-44804467be7f"/>

[8] <texaiWordFormInflection xmlns="http://texai.org/texai/"
rdf:resource="http://sw.cyc.com/2006/07/27/cyc/Verb"/>

[9] <texaiWiktionaryEnglishWordForm xmlns="http://texai.org/texai/"
rdf:resource="http://texai.org/texai/org.texai.wiktionary.domainEntity.WiktionaryEng
lishWordForm_93781b37-ef2e-4888-9ad1-2502226095e5"/>

[10] <texaiCMUDictionaryEnglishWordForm xmlns="http://texai.org/texai/"
rdf:resource="http://texai.org/texai/org.texai.cmudict.domainEntity.CMUDictionary
EnglishWordForm_15752902"/>

[11] </rdf:Description>

```

Element [1] identifies this word form entry with the URI:

[http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWordForm\\_f7c23c53-075d-4726-be9d-f4c7d01fa92d](http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWordForm_f7c23c53-075d-4726-be9d-f4c7d01fa92d)

Element [2] of the entry indicates that this entry is a TexaiEnglishWordForm.

Element [3] indicates whether this entry was derived from the OpenCyc lexicon.

Element [4] gives the Java class from which an instance is constructed when this entry is retrieved from an RDF store into a Java application:

```
org.texai.lexicon.domainEntity.TexaiEnglishWordForm
```

Element [5] gives the URI identifying the Texai English word entry for which this entry is a word form.

Element [6] gives the word form text.

Optional element [7] gives the URI identifying the ARPABET pronunciation entry for this word form.

Element [8] indicates the part-of-speech URI for this word form.

Optional element [9] gives the URI identifying the Wiktionary RDF entry for this word form. This entry is stored in the separate Wiktionary RDF file.

Optional element [10] gives the URI identifying the CMU Pronouncing Dictionary entry for this word form. This entry is stored in the separate CMU Pronouncing Dictionary RDF file.

XML tag [11] closes this entry.

### Sample Phrase

Below is a sample phrase for the word sense *runs*, as in *direct or control, projects, businesses, etc.*:

```

[1] <rdf:Description
rdf:about="http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiSamplePhrase_1acd3d08-46a3-4396-
8e59-3971d871a39e">

[2] <rdf:type rdf:resource="http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiSamplePhrase"/>

[3] <domainEntityClassName
xmlns="http://texai.org/texai/">org.texai.lexicon.domainEntity.TexaiSamplePhrase</domainEntityClassName
>

```

[4] `<texaiSamplePhraseText xmlns="http://texai.org/texai/">she is running a relief operation in the sudan</texaiSamplePhraseText>`

[5] `</rdf:Description>`

Element [1] identifies this word form entry with the URI:

[http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiSamplePhrase\\_1acd3d08-46a3-4396-8e59-3971d871a39e](http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiSamplePhrase_1acd3d08-46a3-4396-8e59-3971d871a39e)

Element [2] of the entry indicates that this entry is a `TexaiSamplePhrase`.

Element [3] gives the Java class from which an instance is constructed when this entry is retrieved from an RDF store into a Java application:

```
org.texai.lexicon.domainEntity.TexaiSamplePhrase
```

Element [4] gives the sample phrase text.

XML tag [5] closes this entry.

### ARPABET Pronunciation

Below is the ARPABET pronunciation entry for the word *runs*. Most of these are derived from [The CMU Pronouncing Dictionary](#). The remainder have been automatically translated from Wiktionary IPA format to ARPABET format. The [Sphinx](#) automatic speech recognition tool and the [FreeTTS](#) text-to-speech synthesis tool both employ pronunciations encoded in the ARPABET format. The Texai dialog project plans to incorporate these tools for voice applications.

[1] `<rdf:Description`

`rdf:about="http://texai.org/texai/org.texai.cmudict.domainEntity.ARPABETPronunciation_e2d403f8-1a43-4d5e-9edd-44804467be7f">`

[2] `<rdf:type`

`rdf:resource="http://texai.org/texai/org.texai.cmudict.domainEntity.ARPABETPronunciation"/>`

[3] `<rdf:type rdf:resource="http://texai.org/texai/ARPABETPronunciation_Primary"/>`

[4] `<domainEntityClassName`

`xmlns="http://texai.org/texai/">org.texai.cmudict.domainEntity.ARPABETPronunciation</domainEntityClassName>`

[5] `<cxgARPABETPhonemeList xmlns="http://texai.org/texai/" rdf:nodeID="node12mnjq6erx948128"/>`

[6] `</rdf:Description>`

Element [1] identifies this pronunciation entry with the URI:

[http://texai.org/texai/org.texai.cmudict.domainEntity.ARPABETPronunciation\\_e2d403f8-1a43-4d5e-9edd-44804467be7f](http://texai.org/texai/org.texai.cmudict.domainEntity.ARPABETPronunciation_e2d403f8-1a43-4d5e-9edd-44804467be7f)

Element [2] of the entry indicates that this entry is a `ARPABETPronunciation`.

Element [3] gives the Java class from which an instance is constructed when this entry is retrieved from an RDF store into a Java application:

```
org.texai.cmudict.domainEntity.ARPABETPronunciation
```

Element [4] gives URI which is the head of the list of phonemes that constitute this pronunciation.

XML tag [6] closes this entry.

### ARPABET Phoneme

Below is the ARPABET phoneme entry for *R*.

[1] `<rdf:Description`

`rdf:about="http://texai.org/texai/org.texai.cmudict.domainEntity.ARPABETPhoneme_35b55654-2962-4a0c-`

```
b2e1-3ce67d19e6b6">
```

```
[2] <rdf:type rdf:resource="http://texai.org/texai/org.texai.cmudict.domainEntity.ARPABETPhoneme"/>
```

```
[3] <domainEntityClassName  
xmlns="http://texai.org/texai/">org.texai.cmudict.domainEntity.ARPABETPhoneme</domainEntityClassName>
```

```
[4] <cxgPhonemeName xmlns="http://texai.org/texai/">R</cxgPhonemeName>
```

```
[5] </rdf:Description>
```

Element [1] identifies this phoneme entry with the URI:

[http://texai.org/texai/org.texai.cmudict.domainEntity.ARPABETPhoneme\\_35b55654-2962-4a0c-b2e1-3ce67d19e6b6](http://texai.org/texai/org.texai.cmudict.domainEntity.ARPABETPhoneme_35b55654-2962-4a0c-b2e1-3ce67d19e6b6)

Element [2] of the entry indicates that this entry is an ARPABETPhoneme.

Element [3] gives the Java class from which an instance is constructed when this entry is retrieved from an RDF store into a Java application:

```
org.texai.cmudict.domainEntity.ARPABETPhoneme
```

Element [4] gives the phoneme code text

XML tag [5] closes this entry.

### N3 Format

[Notation 3](#) is a non-XML format designed for human readability. For example, below are most of the N3 statements in the `texai-lexicon.n3` file that correspond to the *runs* RDF word form entry described previously.

```
[1] <http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWordForm_f7c23c53-075d-4726-be9d-f4c7d01fa92d> a <http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWordForm> , <http://texai.org/texai/TexaiEnglishWordForm-IsNotOpenCycPrettyString> ;
```

```
[2] <http://texai.org/texai/domainEntityClassName>  
"org.texai.lexicon.domainEntity.TexaiEnglishWordForm" ;
```

```
[3] <http://texai.org/texai/texaiEnglishWordFormForWord>  
<http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWord_6f4d0bcc-f19a-482c-ba0f-ca237714568c> ;
```

```
[4] <http://texai.org/texai/texaiWordForm> "runs" ;
```

```
[5] <http://texai.org/texai/texaiARPABETPronunciation>  
<http://texai.org/texai/org.texai.cmudict.domainEntity.ARPABETPronunciation_e2d403f8-1a43-4d5e-9edd-44804467be7f> ;
```

```
[6] <http://texai.org/texai/texaiWordFormInflection> <http://sw.cyc.com/2006/07/27/cyc/Verb> ;
```

```
[7] <http://texai.org/texai/texaiWiktionaryEnglishWordForm>  
<http://texai.org/texai/org.texai.wiktionary.domainEntity.WiktionaryEnglishWordForm_93781b37-ef2e-4888-9ad1-2502226095e5> ;
```

```
[8] <http://texai.org/texai/texaiCMUDictionaryEnglishWordForm>  
<http://texai.org/texai/org.texai.cmudict.domainEntity.CMUDictionaryEnglishWordForm_15752902> .
```

Element [1] identifies this word form entry with the URI:

[http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWordForm\\_f7c23c53-075d-4726-be9d-f4c7d01fa92d](http://texai.org/texai/org.texai.lexicon.domainEntity.TexaiEnglishWordForm_f7c23c53-075d-4726-be9d-f4c7d01fa92d)

Element [2] of the entry indicates that this entry is a `TexaiEnglishWordForm`.

Element [3] gives the URI identifying the `Texai` English word entry for which this entry is a word form.

Element [4] gives the word form text.

Optional element [5] gives the URI identifying the ARPABET pronunciation entry for this word form.

Element [6] indicates the part-of-speech URI for this word form.

Optional element [7] gives the URI identifying the Wiktionary RDF entry for this word form. This entry is stored in the separate Wiktionary RDF (or n3) file.

Optional element [8] gives the URI identifying the CMU Pronouncing Dictionary entry for this word form. This entry is stored in the separate CMU Pronouncing Dictionary RDF file.