

Archimedes Project Morphology Service. XML-RPC Interface documentation

Description

The [Archimedes Project](#) Morphology Service exposes the Archimedes Project's morphological tools using the [XML-RPC specification](#) for syndication of server procedures.

The purpose of the service is to find lemmata -- dictionary head words -- for entered words in a given language. A list of supported languages can be found below.

The server can be contacted at

`<http://archimedes.mpiwg-berlin.mpg.de:8098/RPC2>`

The format of an XML-RPC call consists of a method name, followed by a data structure. For a detailed how-to on implementation, see <http://xmlrpc-c.sourceforge.net/xmlrpc-howto/xmlrpc-howto.html>

Methods

lemma

lemma [\$switch,[word1,word2,...],\$stype]

Send a list of words for lemmatization. Returns the wordlist as a list with each member pointing to a list of possible lemmata, or to nothing if no lemmata are found. If optional \$stype is supplied, each lemma will point to a list of possible grammatical analyses of the supplied form, when this feature is available for a language (Latin,Greek)

Input is a switch where \$switch is one of

- "-LA" (Latin)
- "-IT" (Italian)
- "-NL" (Dutch)
- "-DE" (German)
- "-FR" (French)
- "-GRC" (Ancient Greek)

a reference to a list of words, such as

- auferri,circulo,distat

and optional \$stype, where \$stype is one of

- full

Some typical calls to the "lemma" method would thus be:

- lemma,"-GRC",['ESTI/N','LU/W'],full

- lemma,"-DE",[Fuchs,Igel]

Return values are:

```
{word1 =>[lemma1,...], word2=>[lemma1, ...],...}
```

or, if \$type=full is supplied,

```
{word1 =>[ lemma1 =>[ grammAnal1, grammAnal2], ...], word2 => [ lemma1 => [ grammAnal1, grammAnal2 ] ...], ...}
```

lemma.supported

lemma.supported

No parameters. Returns list of supported languages with switches and encodings to be used for each language with the "lemma" method.

Languages are indicated by language codes as defined in the [iso639-2](#) standard.

The following data structure will be returned

```
'la' => {
  'enc' => 'unicode',
  'switch' => '-LA'
},
'nl' => {
  'enc' => 'unicode',
  'switch' => '-NL'
},
'it' => {
  'enc' => 'Perseus',
  'switch' => '-FR'
},
'grc' => {
  'enc' => 'beta code',
  'switch' => '-GRC'
},
'de' => {
  'enc' => 'unicode',
  'switch' => '-DE'
},
'fr' => {
  'enc' => 'Latin1',
  'switch' => '-FR'
}
};
```

lemma.help

lemma.help

The uri of this document.

Examples

A sample client script written in perl is available [here](#). It uses the Frontier module for XML-RPC available from [CPAN](#).

A sample client script written in python is available [here](#). You need python 2.2+ to run it.

Encoding

The following encodings are used for each of the supported languages

Ancient Greek	Beta Code
Latin	Unicode
Dutch	Unicode
German	Celex encoding
French	Latin1
Italian	Perseus Project encoding (undocumented)

Perseus Project Encoding for Italian

- All letters should be lowercase
- Accents are represented by signs following the accented letter, as indicated in the table below

acute	following /	piu/
grave	following \	e\

Dictionary Server Lookup Tool

The XML-RPC interface to the Archimedes Project's morphology services serves as the backend to the [Dictionary Server Lookup Tool](#).

Datasets

The table below gives the datasets that have been used to provide the morphological analysis for each language.

Greek	Perseus Project morphological analyzer (Morpheus)
Latin	Perseus Project morphological analyzer (Morpheus)
Italian	Perseus Project morphological analyzer (Morpheus)

Dutch	Celex Database
German	Celex Database
French	http://www.limsi.fr/Individu/anne/OLDlexique.txt

This documentation was last edited by [Brian Fuchs](#) on Thu Nov 11 13:18:50 CET 2004