

Bermuda Hyphenator

Table of Contents

1	BASIC INFORMATION.....	2
1.1	Web Service name	2
1.2	Overview and purpose of the Web Service.....	2
1.3	A short description of the service	2
2	TECHNICAL INFORMATION	2
2.1	WSDL code for invoking the web-service	2
2.2	Software dependencies and system requirements (if any)	2
2.3	Execution instructions.....	2
2.4	Input / Output data formats	2
2.5	Integration with external tools	3
3	CONTENT INFORMATION	3
3.1	An usage example with associated data.....	3
3.2	An example of the output data	3
3.3	Approximation of the time necessary to process the test input file	3
4	ADMINISTRATIVE INFORMATION	3
4.1	Contact	3
5	REFERENCES	3

1 BASIC INFORMATION

1.1 Web Service name

The Web Service name is Bermuda Hyphenator (part of the VoiceForge deliverable in Batch 3). It is a substitute for the ro-Hyphen tool in Batch 2.

1.2 Overview and purpose of the Web Service

Bermuda Hyphenator is used to split words into syllables and add stress information at syllable level.

1.3 A short description of the service

Bermuda Hyphenator uses a MaxEnt classifier for splitting the word into syllables based on lexical features and then applies a custom algorithm (also based on MaxEnt) in order to assign stress to a syllable. The modification of the original MaxEnt classifier for stress assignment consists in the fact that the MaxEnt classifier is *forced* to assign exactly one lexical stress. The tool uses different stress patterns based on the part-of-speech of the word.

2 TECHNICAL INFORMATION

2.1 WSDL code for invoking the web-service

The WSDL code is located at <http://khufu.racai.ro:8081/Hyphenator.asmx?WSDL>.

2.2 Software dependencies and system requirements (if any)

There is no need for local installations and there are no software dependencies or system requirements for this tool, since it is hosted by a remote machine (khufu.racai.ro).

2.3 Execution instructions

The GetHyphenation method is used to hyphenate a given word (parameter 1) and assign stress to a syllable based on the part-of-speech (POS) (parameter 2) of the given word. Use “N” for nouns, “V” for verbs, “A” for adjectives and “R” for adverbs. For other or unknown use “OTHER” as POS. Also use “OTHER” if you don’t have enough information about the word’s POS.

2.4 Input / Output data formats

It returns the word’s syllables separated with hyphens and it marks the stressed syllable with the “” character. An example output for the word “testare” (en. testing) is: “tes-’ta-re”

2.5 Integration with external tools

Bermuda Hyphenator is fully self-contained.

3 CONTENT INFORMATION

3.1 An usage example with associated data

For the request: word:'testare' PartOfSpeech: 'N' the Web Service returns the following response:

```
<string>tes-'ta-re</string>
```

3.2 An example of the output data

An example of the output data is given in section 3.1.

3.3 Approximation of the time necessary to process the test input file

Bermuda Hyphenator Web Service is hosted on a 8-core Intel(R) Xeon(R) x64 CPU E5504 @ 2.00GHz and 8 GB of RAM. It takes about 1-2ms to process a request (if the resources are pre-loaded).

4 ADMINISTRATIVE INFORMATION

4.1 Contact

For further information, please contact Tiberiu BOROȘ (tibi@racai.ro).

5 REFERENCES