

Lematizador para o Português do Brasil

[Digite o subtítulo do documento]

Este documento dá orientações de como utilizar o
lematizador

10/05/2012

Arquitetura



Como utilizar

No diretório do lematizador:

```
java -jar lematizador argumento1 argumento2
```

- *argumento1* é o caminho do texto a ser lematizado
- *argumento2*, se definido como “nf” oculta as palavras flexionadas da saída do lematizador

Saída é armazenada no arquivo com extensão “.out” e nome indicado por *argumento1*, ou seja, “*argumento1.out*”.

Formato de saída

Quando o argumento2 está definido como *nf* (*não-flexionada*) a saída é composta pelos lemas das palavras de entrada separados por espaços.

Quando o argumento2 não está definido ou definido com qualquer outro valor diferente de *nf*, a saída é da forma *palavra/lema*, por exemplo:

Os/o meninos/menino cantaram/cantar ./.

Quando argumento2 é definido como *nf*:

o menino cantar .

Verbos hifenizados (ênclise e mesóclise)

Para verbos com ênclise ou mesóclise retorna-se o verbo lematizado e o pronome oblíquo com a indicação se proveio de ênclise ou mesóclise.

Exemplos:

Disseram-se torna-se dizer (-se[enc])

Caso de ênclise

Tornar-me-ia torna-se tornar (-me[mes])

Caso de mesóclise

Dependências

Para executar o lematizador é necessária a instalação do Java.

O Java pode ser baixado em:

http://www.java.com/pt_BR/download/

Outras informações

Sempre que uma palavra não é encontrada na base lexical do Unitex-PB, sua própria forma flexionada é retornada.