

# Instruções para instalação do Aelius

---

## Pré-requisitos

---

1) Python (versão 2.6 ou superior, mas não a versão 3.0), NLTK (versão 2.0.1 ou superior) e todas as bibliotecas exigidas pelo NLTK. Consultar:

<http://www.nltk.org/>

Caso não saiba qual é a sua versão do NLTK, execute:

```
>>> import nltk
>>> nltk.__version__
```

2) Tokenizador Punkt do NLTK, que pode ser instalado por meio do NLTK Downloader:

```
>>> import nltk
>>> nltk.download("punkt")
```

3) Etiquetador Hunpos

<http://code.google.com/p/hunpos/>

4) Etiquetador LX-Tagger e todas as bibliotecas exigidas por ele. Consultar:

<http://lxcenter.di.fc.ul.pt/tools/pt/conteudo/LXTagger.html>

---

## Passos para a instalação

---

Nas instruções abaixo, assumimos que você baixou a versão 23-12-2011 do Aelius e vai instalar o programa no Linux ou Mac OS X.

Para outras versões do Aelius ou para instalação em outros sistemas operacionais, faça as adaptações necessárias.

1. Descompacte o arquivo Aelius-23-12-2011.zip
2. No diretório \$HOME, crie a pasta Applications, se ela ainda não existir.
3. Copie a pasta Aelius para o diretório \$HOME/Applications; verifique isso na shell:

```
ls $HOME/Applications/Aelius
```

4. Copie a pasta aelius\_data para o diretório \$HOME; verifique isso na shell:

```
ls $HOME/aelius_data
```

5. O diretório \$HOME/aelius\_data deverá conter os seguintes arquivos:

AeliusBRUBT.pkl	LICENSE.txt	exemplo.txt
AeliusHunPos	README.pdf	freq_tycho_a.pkl
AeliusRUBT.pkl	README.txt	

O uso dos arquivos AeliusBRUBT.pkl, AeliusRUBT.pkl e AeliusHunPos (etiquetadores treinados em versão modificada do Corpus Histórico do Português Tycho Brahe, doravante CHPTB) e freq\_tycho\_a.pkl (lista de palavras mais frequentes de amostra do CHPTB) está sujeito às condições estabelecidas no arquivo README.txt ou README.pdf que os acompanha na pasta aelius\_data. Informações de copyright encontram-se também nessa pasta.

Visite as seguintes páginas na WWW para obter informações sobre o Corpus Histórico do Português Tycho Brahe:

Homepage:

<http://www.tycho.iel.unicamp.br/~tycho/corpus/>

Licença de uso do corpus:

<http://www.tycho.iel.unicamp.br/~tycho/corpus/termos.html>

6. Além desses arquivos, é necessário criar, no diretório \$HOME/aelius\_data, um link simbólico para o modelo do português MXPOST do LX-Tagger, disponível em:

<http://lxcenter.di.fc.ul.pt/tools/pt/conteudo/LXTagger.html>

Siga as instruções de instalação do LX-Tagger fornecidas no respectivo site ou na pasta do etiquetador descarregada da Internet. Você precisa do Java 5 (JDK1.5) ou superior, que já é pré-instalado nas máquinas com sistema operacional MAC OS X, mas que pode não constar de sua instalação do Linux. Verifique se Java está instalado no Terminal:

```
java -version
```

Além disso, é necessária a instalação do aplicativo JAVA MXPOST, desenvolvido por Adwait Ratnaparkhi, disponível em:

<ftp://ftp.cis.upenn.edu/pub/adwait/jmx/jmx.tar.gz>

ou a partir de link em

<http://sites.google.com/site/adwaitratnaparkhi/publications>

Siga as instruções fornecidas pelo autor do MXPOST. Você precisa incluir as seguintes linhas no seu arquivo de configuração \$HOME/.profile ou \$HOME/.bash\_profile (conforme a sua máquina):

```
CLASSPATH="$HOME/jmx/mxpost.jar"
export CLASSPATH
PATH="${PATH} : $HOME/jmx"
export PATH
```

Supondo que você descompactou o arquivo do LX-Tagger POSTagger.zip no diretório \$HOME, resguardadas alterações na distribuição do LX-Tagger, proceda da seguinte forma para criar link simbólico para o modelo do português desse etiquetador:

```
cd $HOME/aelius_data
ln -s $HOME/POSTagger/Tagger/MX-PoS/UTF-8_Model_Cintil_Written \
lxtagger
```

Verifique se o link está correto:

```
ls $HOME/aelius_data/lxtagger
```

Quanto ao Hunpos, para que funcione com o NLTK, é preciso colocar os dois arquivos binários

```
hunpos-tag
hunpos-train
```

numa das pasta pré-definidas pela interface do NLTK para o Hunpos, por exemplo:

```
$HOME/Applications/bin
```

Inclua as seguintes linhas no seu arquivo de configuração \$HOME/.profile ou \$HOME/.bash\_profile (conforme a sua máquina):

```
PATH="${PATH} : $HOME/Applications/bin"
export PATH
```

Consultar módulo hunpos.py do pacote "tag" do NLTK para outros lugares onde instalar HunPos.

7. Inclua \$HOME/Applications na lista de caminhos da variável de ambiente PYTHONPATH no seu arquivo de configuração (.profile ou .bash\_profile):

```
PYTHONPATH="${PYTHONPATH} : $HOME/Applications"
export PYTHONPATH
```

8. Abra o IDLE a partir do Terminal e verifique se o Aelius aparece em File/Path Browser, na pasta \$HOME/Applications

9. Execute

```
>>> import Aelius
>>> Aelius.__versao__
'0.9.4 December-23-2011'
```

10. Para obter ajuda, execute

```
>>> help(Aelius)
```

11. Faça o mesmo para os módulos individuais do pacote ou para funções individuais desses módulos.

12. Sugerimos que crie em \$HOME uma pasta analyses, onde deverá armazenar os arquivos a serem analisados. Verifique:

```
ls $HOME/analises
```

Inicie o IDLE a partir do terminal:

```
cd $HOME/analises  
idle
```

Com isso, a pasta \$HOME/analises será o diretório de trabalho de Python, no qual o Aelius salva os arquivos que anota.

13. Caso tenha dúvida, registre-se no SourceForge e coloque a sua pergunta na seção Forum/Open Discussion do Aelius.