

Instruções para instalação do Aelius

25/02/13

Pré-requisitos

Básicos

- 1) Python (versão 2.7, mas não a versão 3.0)

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

- 2) Bibliotecas exigidas pelo NLTK compatíveis com a sua versão de Python, a serem instaladas nesta sequência:

1. PyYAML

<http://pyyaml.org/>

2. NumPy

<http://numpy.scipy.org/>

3. Matplotlib

<http://matplotlib.sourceforge.net/>

- 3) NLTK na versão 2.0.1rc1 (recomendada):

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

Recomendamos, tanto para usuários do Linux e do Mac OS X quanto do Windows, a instalação a partir do código-fonte, disponível em um destes arquivos compactados:

<http://code.google.com/p/nltk/downloads/detail?name=nltk-2.0.1rc1.tar.gz>

<http://code.google.com/p/nltk/downloads/detail?name=nltk-2.0.1rc1.zip>

Informações detalhadas sobre a instalação da versão 2.0.1rc1 do NLTK encontram-se no antigo site do projeto:

<https://sites.google.com/site/naturallanguagetoolkit/Home>

Link direto para as instruções de instalação (ver seção *NLTK Source Installation*):

<https://sites.google.com/site/naturallanguagetoolkit/download>

Basicamente, para instalar o NLTK na versão 2.0.1rc1 a partir do código-fonte, basta entrar, no Terminal, na pasta em que descompactou o arquivo tar.gz ou .zip e executar no Linux ou Mac

OS X:

```
sudo python setup.py install
```

No Windows, execute no interpretador de comandos [cmd.exe](#):

```
python setup.py install
```

Para que o comando acima funcione, é preciso que a pasta onde se encontra o binário de Python, geralmente C:\Python27, esteja na sua [variável PATH](#).

Caso já tenha instalado o NLTK e não saiba qual é a sua versão, execute:

```
>>> import nltk
>>> nltk.__version__
'2.0.1rc1'
>>>
```

O Aelius não funciona plenamente com as versões 2.0.4 e 3.0a1 do NLTK (esta última, versão ainda experimental para Python 3), as mais recentes em 25 de fevereiro de 2013, cujas instruções de instalação estão disponíveis no site atual do projeto:

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

4) Toqueizador Punkt do NLTK, que pode ser instalado por meio do NLTK Downloader:

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

Adicionais

1) Etiketador Hunpos

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

2) Etiketador LX-Tagger e todas as bibliotecas exigidas. Consultar:

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

3) Etiketador Stanford Tagger e todos os pré-requisitos exigidos. Consultar:

<http://nlp.stanford.edu/software/tagger.shtml>

4) Apache OpenNLP e todos os pré-requisitos exigidos. Consultar:

<http://opennlp.apache.org/>

Passos para a instalação

Nas instruções abaixo, assumimos que você baixou a versão February-25-2013 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-February-25-2013.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 -m $HOME/aelius_data
```

5. O diretório `$HOME/aelius_data` deverá conter, entre outros, os seguintes arquivos:

`AeliusBRUBT.pkl`, `AeliusHunPos`, `AeliusHunPosMM`, `AeliusMaxEnt`,
`AeliusRUBT.pkl`, `AeliusStanfordMM`,
`LICENSE.txt`, `README.html`, `README.txt`, `exemplo.nltk.gold.txt`,
`exemplo.txt`, `freq_tycho_a.pkl`

O uso dos etiquetadores treinados em versão modificada do Corpus Histórico do Português Tycho Brahe, doravante CHPTB, tais como `AeliusBRUBT.pkl`, `AeliusRUBT.pkl`, `AeliusHunPos` e `AeliusMaxEnt`, bem como do arquivo `freq_tycho_a.pkl`, uma lista das palavras mais frequentes de amostra do CHPTB, está sujeito às condições estabelecidas no arquivo de nome-base `README` que os acompanha na pasta `aelius_data`, onde se encontram informações de *copyright*.

Do mesmo modo, o uso dos arquivos `AeliusHunPosMM` e `AeliusStanfordMM`, entre outros etiquetadores treinados no Corpus MAC-Morpho, implica observância das restrições de utilização desse corpus. Para mais detalhes, consultar o referido arquivo de nome-base `README`.

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>

Em trabalhos que utilizem algum desses materiais baseados no CHPTB, deve-se citar esse corpus por meio da seguinte referência:

GALVES, C.; FARIA, P. *Tycho Brahe Parsed Corpus of Historical Portuguese*. [S.l.: s.n.], 2010. Disponível em: <<http://www.tycho.iel.unicamp.br/~tycho/corpus/en/index.html>> Acesso em: 13 dez. 2012.

Visite o sítio do Projeto Lácio-Web para obter mais informações sobre o Corpus MAC-Morpho:

<http://www.nilc.icmc.usp.br/lacioweb/index.htm>

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-POST/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.

Quanto ao Stanford Tagger, a instalação é análoga à do MXPOST, com a diferença de que o primeiro etiquetador exige Java 1.6 ou superior. Descarregue o arquivo compactado disponível a partir de link na URL

<http://nlp.stanford.edu/software/tagger.shtml>

Descompacte o arquivo na sua pasta `$HOME`. Supondo que se trata da versão disponível em <http://nlp.stanford.edu/software/stanford-postagger-2012-11-11.zip>, você deverá ter

```
$HOME/stanford-postagger-2012-11-11
```

Coloque os seguintes comandos no seu arquivo de configuração `.profile` ou `.bash_profile`:

```
CLASSPATH="${CLASSPATH} : $HOME/stanford-postagger-2012-11-11/stanford-
postagger.jar"
export CLASSPATH
```

Atente para a versão do Stanford Tagger que você descarregou. Precisarás alterar o comando acima se

está usando versão diferente da 2012-11-11.

A instalação do Apache OpenNLP segue os mesmos princípios, uma vez que se trata de programa em Java.

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
```

Ver arquivo `profile.txt` com todas as alterações a serem feitas no seu arquivo de configuração!

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__
'0.9.7 february-25-2013'
```

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 `analises`, onde deverá armazenar os arquivos a serem analisados. Verifique na *shell*:

```
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.

Instruções de Instalação do Aelius by [Leonel F. de Alencar](#) is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License](#).