

Guia para instalação e utilização do CorpusDraw/CorpusSearch

Autor: Pablo Faria (pablofaria@gmail.com)

Observações gerais

O objetivo deste documento é descrever como baixar, instalar e utilizar o pacote compactado, contendo as ferramentas CorpusDraw-CorpusSearch, scripts para sua utilização e a máquina virtual *Java* necessária para sua execução. Espera-se um conhecimento básico de utilização do **Prompt do DOS** (o terminal do Windows). Caso não saiba bem como usar, pesquise por tutoriais no Google, há vários pela internet. Também é necessário ter um descompactador ZIP instalado na sua máquina (normalmente, o Windows já vem com essa funcionalidade).

O pacote disponibilizado no site do Projeto Tycho Brahe (CTB) tem o intuito de tornar quase trivial a utilização dessas ferramentas, ao fornecer um ambiente pré-configurado. Assim, acreditamos que será possível executar as ferramentas em praticamente qualquer computador Windows, inclusive a partir de um *pendrive*, se for o caso.

Instalação

1. Baixe o arquivo `csearch-dos.zip` contendo todos os arquivos em <http://www.tycho.iel.unicamp.br/~tycho/tmp/>. O download pode levar algum tempo, pois o arquivo tem aproximadamente **80MB**.
2. Grave o arquivo no local desejado e descompacte. Será criada uma pasta chamada “csearch”, que contém:
 - Scripts (.BAT) para execução das ferramentas: `CDraw` (versão “69”), `CDraw57` (versão “57”) e `CSearch` (versão “69”).
 - Scripts (.BAT) para automatização de algumas tarefas: `update-new` e `update-out`.
 - Scripts (.BAT) para execução de um terminal de texto (Prompt do DOS), *pré-configurado para executar o Java*. O arquivo `Prompt-DOS` será utilizado quando você quiser utilizar as ferramentas numa só máquina, de modo “fixo”. O outro arquivo `Prompt-DOS_pendrive` será utilizado para executar as ferramentas em qualquer máquina Windows, por exemplo, a partir de um pendrive (*neste caso, o arquivo ZIP deve ser descompactado dentro do pendrive*).
 - Uma pasta contendo as ferramentas: `csearch\jar\`
 - Uma pasta contendo a máquina virtual Java, versão 1.6: `csearch\java\`
 - O arquivo `port.def` necessário para executar as ferramentas.
 - A pasta `port` que contém o arquivo `tags.tag` (necessário para correção/consulta de textos anotados do CTB) e em que são colocados os textos anotados sintaticamente.
 - A pasta `queries` onde são colocadas as buscas (arquivos “.q”). Por padrão, esta pasta contém o arquivo `port.q`, para visualização de todas as sentenças de um arquivo, no CorpusDraw.
3. A partir daqui, é apenas para utilização “fixa” das ferramentas:
 1. Grave a pasta `csearch` na raiz do computador (por exemplo, em `c:\csearch`).
 2. Copie as pastas `port` e `queries` e o arquivo `port.def` para sua pasta de documentos (por exemplo, para “Meus documentos”).
 3. Copie o arquivo `Prompt-DOS` para sua área de trabalho (ou crie um atalho para ele, se souber fazer isso).

Execução

Agora, a utilização das ferramentas é simples. Para a utilização não-“fixa” das ferramentas, basta dar um duplo clique em “Prompt-DOS_pendrive”, do contrário, faça o mesmo sobre “Prompt-DOS”. Uma janela de terminal será aberta, já na pasta (diretório) correta. A execução para cada ferramenta é (repare nos espaços em branco, na sintaxe dos comandos):

- CorpusDraw, versão 69:
 - `cdraw queries\<nome-do-arquivo-de-busca> port\<nome-do-arquivo-annotado>`, ou apenas
 - `cdraw port\<nome-do-arquivo-annotado>`, quando você for usar a busca “port.q”.
- CorpusDraw, versão 57:
 - `cdraw57 queries\<nome-do-arquivo-de-busca> port\<nome-do-arquivo-annotado>`, ou apenas
 - `cdraw57 port\<nome-do-arquivo-annotado>`, quando você for usar a busca “port.q”.
- CorpusSearch:
 - `csearch queries\<nome-do-arquivo-de-busca> port\<nome-do-arquivo-annotado>`

Exemplos:

- `cdraw port\A003.psd`
- `cdraw queries\port.q port\A003.psd`
- `cdraw57 port\A003.psd`
- `csearch queries\relativas.q port\A003.psd`

Importante:

Ao final de cada sessão de revisão com o CorpusDraw é criado um arquivo “<nome-do-arquivo-annotado>.**new**”, ficando o original *intacto*. Assim, é necessário executar (no mesmo local de execução do comando “cdraw”): `update-new port\<nome-do-arquivo-psd>`, para fazer valer as alterações. Este script faz uma cópia de segurança do original, chamada “<nome-do-arquivo-annotado>.**old**”.

O mesmo vale para revisões utilizando o CorpusSearch, que cria o arquivo “<nome-do-arquivo-annotado>.**out**”. Assim, deve-se executar: `update-out port\<nome-do-arquivo-psd>`, para fazer valer as alterações. Este script faz uma cópia de segurança do original, chamada “<nome-do-arquivo-annotado>.**bak**”.

Sobre as ferramentas

Caso você ainda não saiba utilizar as ferramentas CorpusDraw/CorpusSearch, sugerimos recorrer ao manual disponibilizado online em <http://corpussearch.sourceforge.net/CS-manual/Contents.html> e em PDF, na pasta `csearch\guias\`.