

Guia rápido para o Text File Splitter 2.5.1

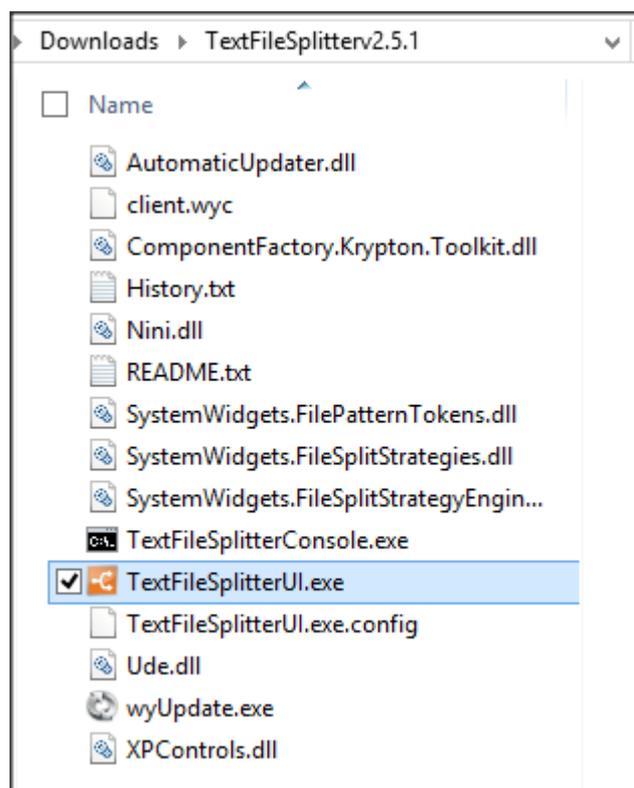
O Text File Splitter é um software gratuito para separar textos (Muito útil depois de compilar um *corpus* com o BootCat e mais fácil de usar que o TextWedge.)

1. Para baixar o programa, entre em: <http://www.systemwidgets.com/>



No menu, clique em Downloads >> Freeware Downloads >> Text File Splitter. O arquivo virá zipado. O Windows 7 e 8 já têm um descompactador. Caso você queira baixar um, procure pelo WinRAR no Baixaki.

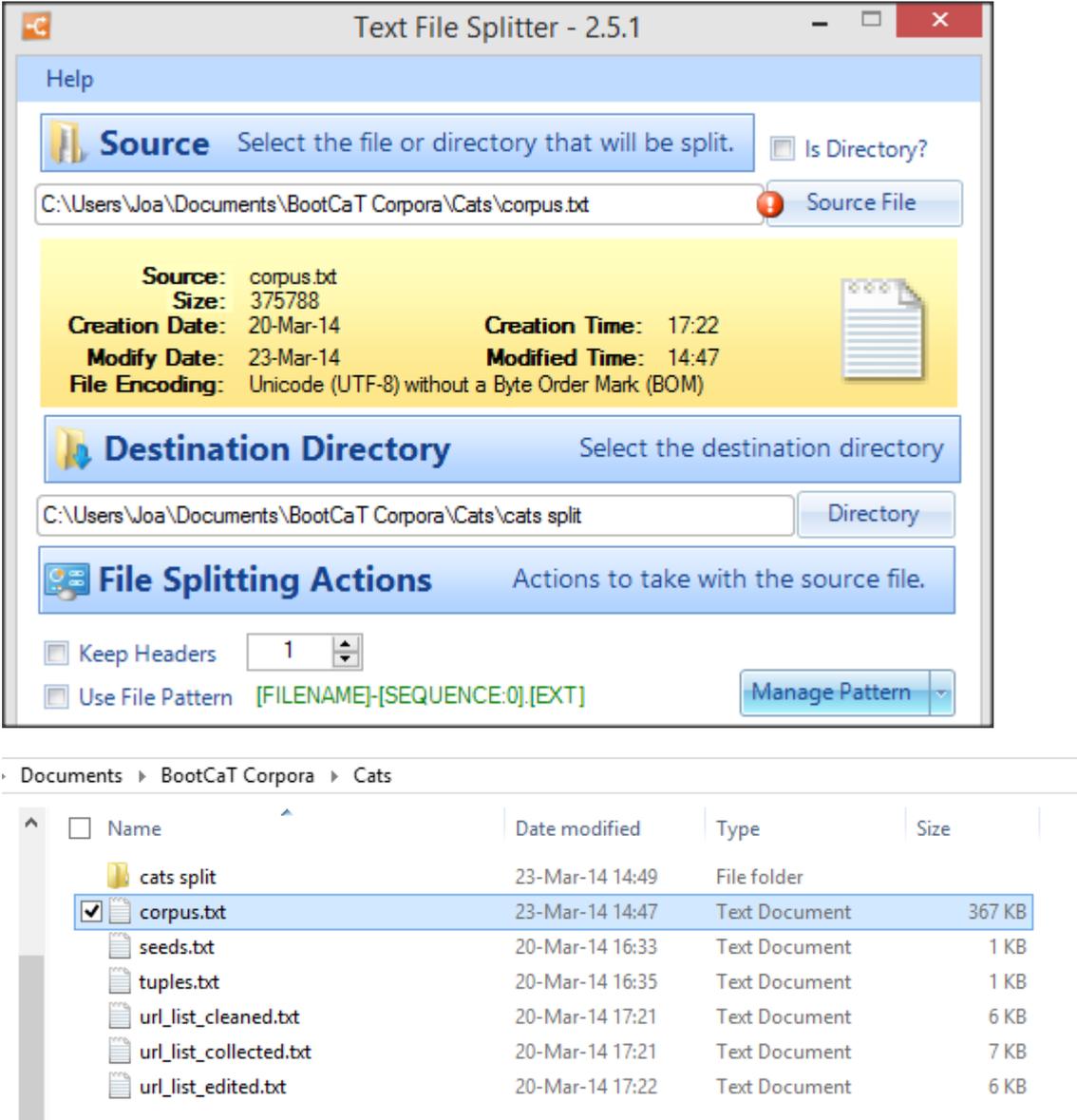
Uma vez descompactado, na pasta você encontrará vários arquivos. Selecione o **TextFileSplitterUI.exe** para iniciar o programa. Se quiser colocar um atalho na área de trabalho, clique com o botão direito do mouse sobre o arquivo e selecione “Enviar para” e depois “Área de trabalho”.



2. Uso com funções básicas

O BootCat reúne todos os textos coletados da Internet em um só arquivo chamado *corpus.txt*. Esse arquivo fica em uma pasta dentro do diretório do BootCat (geralmente na pasta Documentos) e tem o mesmo nome que você deu ao *corpus* logo no início do processo de compilação. Ao abrir esse arquivo, você perceberá que cada texto retirado de um site da Internet se inicia com a palavra CURRENT. Essa palavra nos ajudará a dividir o arquivo *corpus.txt* em arquivos individuais.

a. Com o programa aberto, carregue o arquivo **corpus.txt** para ser dividido.

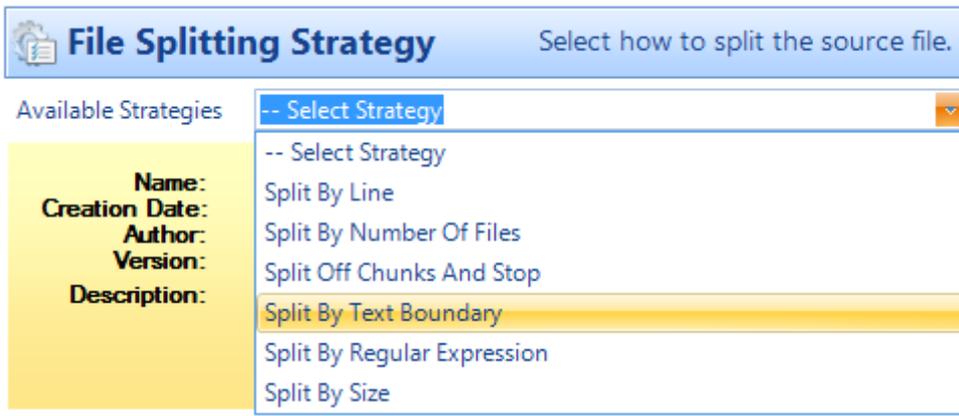


The image shows the Text File Splitter - 2.5.1 application window. The 'Source' field is set to 'C:\Users\Joa\Documents\BootCaT Corpora\Cats\corpus.txt'. The 'Destination Directory' is 'C:\Users\Joa\Documents\BootCa T Corpora\Cats\cats split'. The 'File Splitting Actions' section has 'Keep Headers' checked and 'Use File Pattern' set to '[FILENAME]-[SEQUENCE:0].[EXT]'. Below the application window, a Windows File Explorer window shows the contents of the 'cats split' directory. The files listed are:

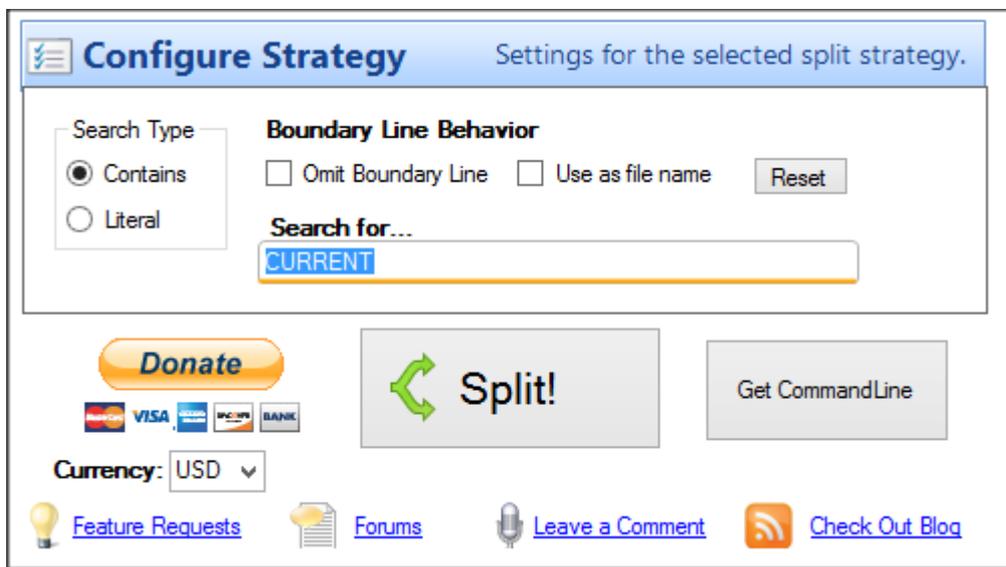
Name	Date modified	Type	Size
cats split	23-Mar-14 14:49	File folder	
corpus.txt	23-Mar-14 14:47	Text Document	367 KB
seeds.txt	20-Mar-14 16:33	Text Document	1 KB
tuples.txt	20-Mar-14 16:35	Text Document	1 KB
url_list_cleaned.txt	20-Mar-14 17:21	Text Document	6 KB
url_list_collected.txt	20-Mar-14 17:21	Text Document	7 KB
url_list_edited.txt	20-Mar-14 17:22	Text Document	6 KB

Em seguida, selecione o diretório de destino. Para manter o diretório organizado, crie uma pasta que receberá os arquivos resultantes da divisão. No campo *Destination Directory*, selecione esta pasta ou a pasta onde você quer que os arquivos sejam salvos.

b. No campo *File Splitting Strategy* selecione “Split By Text Coundary”.



c. No campo *Configure Strategy* selecione “Contains” em *Search for ...* escreva “CURRENT”.



d. Por fim, clique em SPLIT!

Os textos, agora separados, estão na pasta que foi designada em *Destination Directory*.

As ferramentas não são perfeitas e sempre precisam da supervisão humana. Se o programa não dividiu um ou mais textos, tente descobrir a razão ou separe-os manualmente.

Eu não sou técnico em informática; sou um linguista que gosta de tecnologia. Se você tiver algum problema com o programa, posso tentar ajudar, mas a melhor coisa é ir ao fórum de discussões e perguntar diretamente aos desenvolvedores no site do programa em:

<http://www.systemwidgets.com/Products/TextFileSplitter/Overview/tabid/119/Default.aspx>.

Joacyr T Oliveira
USP 2014
joacyr@outlook.com