

Notas do Curso por
Sanderson C. Ribeiro



TOPCAT

Sumário

1	Curso de TOPCAT	3
2	Principais Características	4
2.1	Capacidades avançadas	4
2.2	Limitações e pontos de atenção	5
3	Instalação do TOPCAT	6
3.1	Instalando o Java no <i>Windows</i>	6
3.2	Executando o TOPCAT	8

1 | Curso de TOPCAT

O TOPCAT (*Tool for OPerations on Catalogues And Tables*) é uma **ferramenta interativa projetada para manipulação, análise e visualização de catálogos astronômicos e tabelas de dados**. É amplamente utilizada por astrônomos e cientistas de dados em astronomia devido à sua capacidade de lidar com grandes conjuntos de dados com eficiência. Nesse sentido o TOPCAT é voltado para dados astronômicos, mas fornece a possibilidade de trabalhar também com outros tipos de dados.

Foi desenvolvido inicialmente por Mark Taylor, astrofísico da Universidade de Bristol do Reino Unido e lançado em 2003. Sua logo, apresentada na Figura 1.1, foi desenhada por Phil Hall, amigo de Taylor, e faz alusão ao personagem do desenho animado *TOP CAT* (Manda-Chuva em português-BR) da Hannah-Barbera no qual temos um gato que é carismático, engenhoso e sempre em busca de uma solução criativa para os problemas — um paralelismo com as funções do TOPCAT no processamento e análise de dados complexos de catálogos astronômicos. Inicialmente Taylor usava o próprio personagem do desenho animado como logo, apesar de não ter os direitos legais para uso do desenho, como o mesmo relata na documentação do *software*. Apenas a partir da versão 4 do TOPCAT que a logo desenhada por Hall começa a ser vinculada ao *software*.



Tool for Operations on Catalogues And Tables

Figura 1.1: Logo do TOPCAT: um gato na cor amarela desenhado por Phil Hall.

O *software* foi escrito na linguagem de programação Java. Entretanto é uma ferramenta interativa GUI (do inglês, *Graphical User Interface*) sendo uma aplicação que fornece uma interface gráfica para o usuário interagir com o *software*. Assim, diferentemente de interfaces baseadas em texto ou comandos de terminal (CLI - *Command Line Interface*), as ferramentas GUI utilizam elementos visuais, como botões, menus, janelas e ícones, para tornar a interação mais intuitiva e amigável, especialmente para usuários que não têm experiência com programação ou comandos de linha. Dessa forma o TOPCAT não exige um conhecimento específico de determinada linguagem de programação para ser utilizado. Mas deve-se compreender bem as suas ferramentas para manipulação dos dados de acordo com os seus objetivos. Atualmente, em 2024, o TOPCAT é *open source* de acordo com a política da GNU *General Public Licence* e é mantido pela comunidade, sendo que qualquer pessoa pode contribuir com o seu desenvolvimento.

2 Principais Características

O TOPCAT é uma ferramenta com diversas características envolvendo conjuntos de dados. A seguir listamos algumas delas.

- **Leitura de formatos de dados variados:** suporte a múltiplos formatos de arquivo: FITS (*Flexible Image Transport System*), VOTable (XML para tabelas astronômicas), CSV (*Comma-Separated Values*), ASCII (textos formatados), dentre outros muito utilizados em Astronomia;
- **Manipulação de tabelas e dados:** filtragem, seleção e ordenação de linhas de tabelas. Junção (*crossmatch*) de diferentes catálogos e tabelas. Possibilidade de criar novas colunas com expressões personalizadas. Operações matemáticas e estatísticas (cálculo da média, desvio padrão, etc) em colunas de dados;
- **Visualização de dados:** geração de gráficos interativos como gráficos de dispersão 2D e 3D, histogramas e mapas celestes permitindo a personalização de estilos e atributos gráficos;
- **Integração com o Virtual Observatory (VO):** permite acessar diretamente bancos de dados astronômicos através de protocolos VO. Suporta buscas e downloads de dados de catálogos distribuídos em servidores ao redor do mundo;
- **Interoperabilidade:** pode ser integrado a fluxos de trabalho com outras ferramentas analíticas, como o Aladin Sky Atlas, o STILTS (versão linha de comando do TOPCAT) ou *scripts* em Python que usam bibliotecas como o Astropy;
- **Interface amigável:** apesar de não apresentar uma interface moderna, o TOPCAT oferece uma interface gráfica simples, mas poderosa, com suporte a operações interativas. Além disso inclui uma linha de comando para usuários avançados que preferem automação ou processamento em lote.

2.1 Capacidades avançadas

Algumas capacidades avançadas do TOPCAT são:

- **Crossmatching eficiente:** foi projetado para lidar eficientemente com tabelas muito grandes, contendo milhões de linhas;
- **Visualizações complexas:** possibilidade de sobrepor dados em gráficos, ajustar parâmetros de visualização, alterar a escala e definir sub conjuntos de dados para análise;
- **Análise interativa:** filtragem dinâmica dos dados com atualização instantânea dos gráficos.

2.2 Limitações e pontos de atenção

Como todo *software* de análise de dados ele tem algumas limitações e pontos de atenção que devemos tomar certos cuidados.

- **Curva de aprendizado:** pode exigir um tempo para se dominar completamente as funcionalidades mais avançadas do TOPCAT;
- **Memória e desempenho:** apesar dele ser projetado para trabalhar com tabelas contendo milhões de linhas, quando passamos para a casa de bilhões o TOPCAT pode exigir máquinas com mais recursos de execução;
- **Interface:** como o TOPCAT tem o foco em GUI, usuários que preferem trabalhar com *scripts* por vezes vão achar mais prático o STILTS para automatizar certas tarefas de fluxo de trabalho.

3 Instalação do TOPCAT

Acesse o site oficial do TOPCAT em: www.star.bris.ac.uk (hospedado pela Universidade de Bristol). Role a página até a seção relacionada a Downloads que descreve como ter acesso ao TOPCAT em seu computador. Se você tiver o Java instalado em seu computador basta clicar sobre o link referente ao arquivo `topcat-extra.jar`¹ do TOPCAT em Standalone Jar File como mostrado na Figura 3.1. A versão `topcat-extra.jar` inclui suporte ao formato de arquivo Parquet que é um formato de arquivo colunar otimizado para armazenar e processar grandes volumes de dados, especialmente em ambientes distribuídos, como aqueles que utilizam o Apache Hadoop, Apache Spark ou outros frameworks de processamento de dados. É interessante baixarmos essa versão, até por que se for analisar ocupa uma quantidade razoável de memória do HD, e nunca se sabe quando estaremos trabalhando com esses tipos de arquivos.

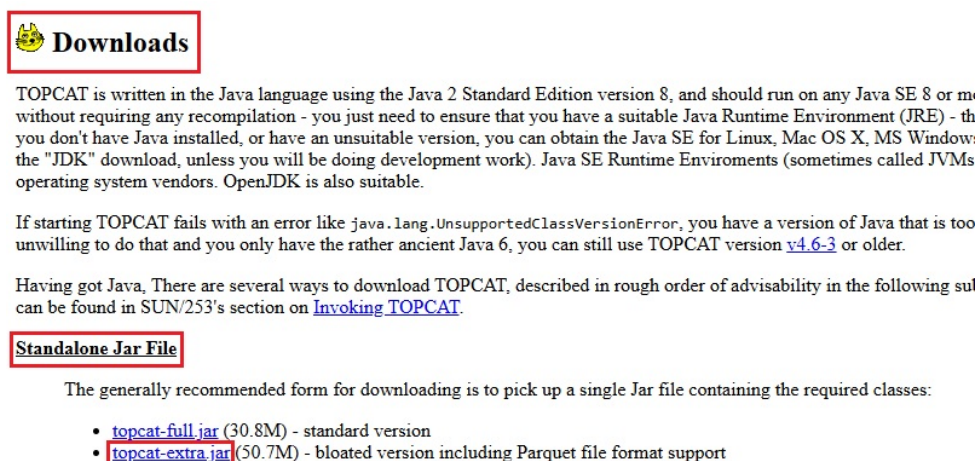


Figura 3.1: *Download* da versão *Standalone* do TOPCAT.

Se você não tiver o Java instalado em seu computador siga os passos da seção a seguir antes de prosseguir.

3.1 Instalando o Java no *Windows*

Acesse o site oficial do Java em java.com. Na aba Fazer Download clique em [Fazer Download do Java](#). Muito provavelmente aparecerá um botão na tela inicial no site do Java com essa última opção, pode clicar sobre ela sem precisar acessar a aba Fazer Download. Então clique sobre o botão [Download Java](#) que será iniciado o download do arquivo Java na pasta padrão Downloads do seu sistema.

¹Quando clicamos sobre esse arquivo em particular o TOPCAT é executado automaticamente sem a necessidade de instalação do mesmo em seu computador. Daí o nome *standalone* pois ele funciona por ele mesmo.

Acesse o diretório Downloads e com o botão esquerdo do *mouse* dê um duplo clique sobre o arquivo baixado, e permita que este aplicativo faça alterações em seu dispositivo clicando na opção Sim. Se aparecer uma tela como a mostrada na Figura 3.2, apenas clique em Instalar para que a instalação do Java em seu sistema se inicie. Caso a tela em questão não apareça, apenas clique em Next afim de iniciar a instalação. Se a mesma ocorreu como planejado, então uma tela de Instalação concluída com êxito deve aparecer, parecida com a mostrada na Figura 3.3. Clique em Fechar. Você pode excluir o arquivo executável do Java da pasta Downloads pois não vai precisar mais dele.

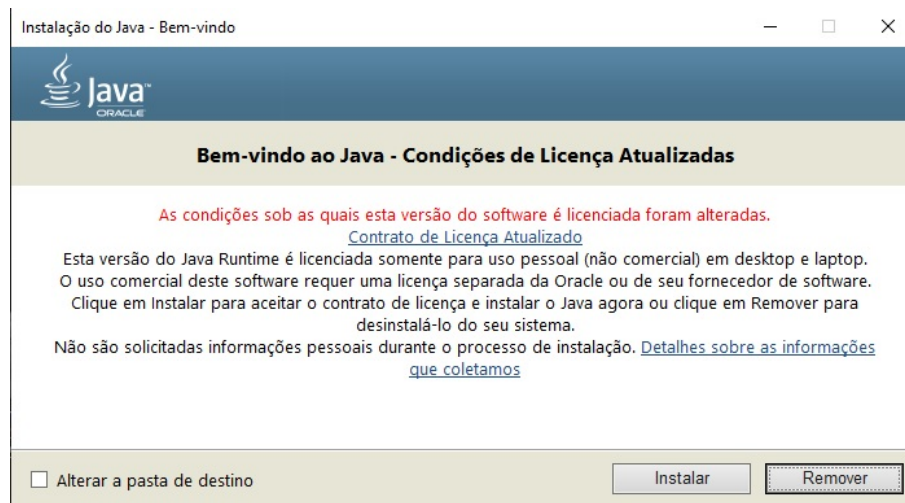


Figura 3.2: Tela de Bem-vindo e condições de licença do instalador Java.

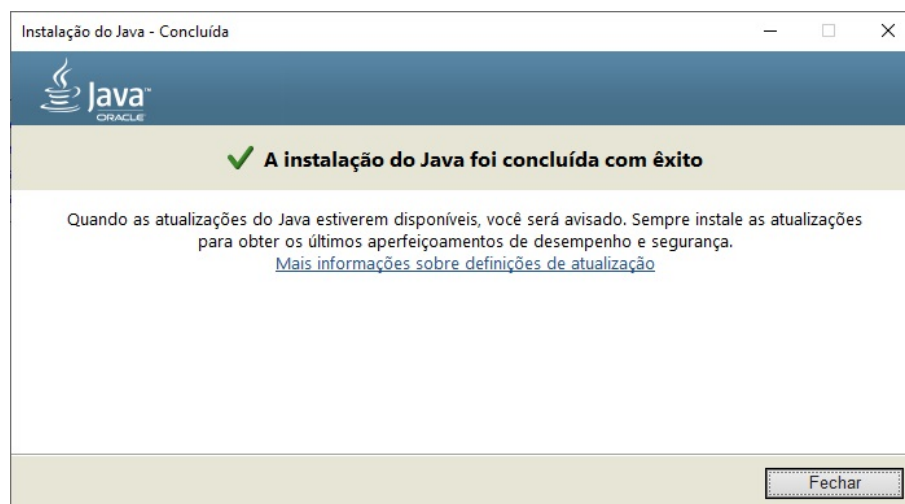


Figura 3.3: Tela de conclusão da instalação do Java no Windows.

Se você acessar novamente o diretório Arquivos de Programas do seu Disco Local verá que tem uma pasta chamada Java. Isso nos diz que o Java está devidamente instalado em seu sistema. Outro modo de verificar isso é com o comando `java --version`

ou `java -version` no terminal que será imprimido na tela a versão do java instalada no seu sistema, como mostrado a seguir, no terminal emulado, para a versão 1.8.0_431 do Java.

```
C:\Users\UserName>java -version
java version "1.8.0_431"
Java(TM) SE Runtime Environment (build 1.8.0_431-b10)
Java HotSpot(TM) 64-Bit Server VM (build 25.431-b10, mixed mode)
```

3.2 Executando o TOPCAT

Com o arquivo baixado clique sobre ele duas vezes com o botão esquerdo do *mouse* se você estiver no *Windows* ou execute o seguinte comando no Terminal:

```
C:\Users\UserName>java -jar topcat-extra.jar
```

Lembrando que você deve primeiro acessar a pasta em que se encontra o arquivo que foi baixado. Se for a pasta Downloads, basta executar no *prompt*: `cd Downloads` e só então executar o comando anterior para que o TOPCAT seja executado.

Em sistemas do tipo Unix você deve seguir um protocolo de linha de comando de acordo com o que é descrito na aba Standalone Jar File para Unix da Seção Downloads.

Prosseguindo com o uso do TOPCAT no *Windows*: se abrir uma janela do *Windows Defender Firewall* apenas clique em Permitir acesso com a opção de Redes públicas marcada como mostra a Figura 3.4.

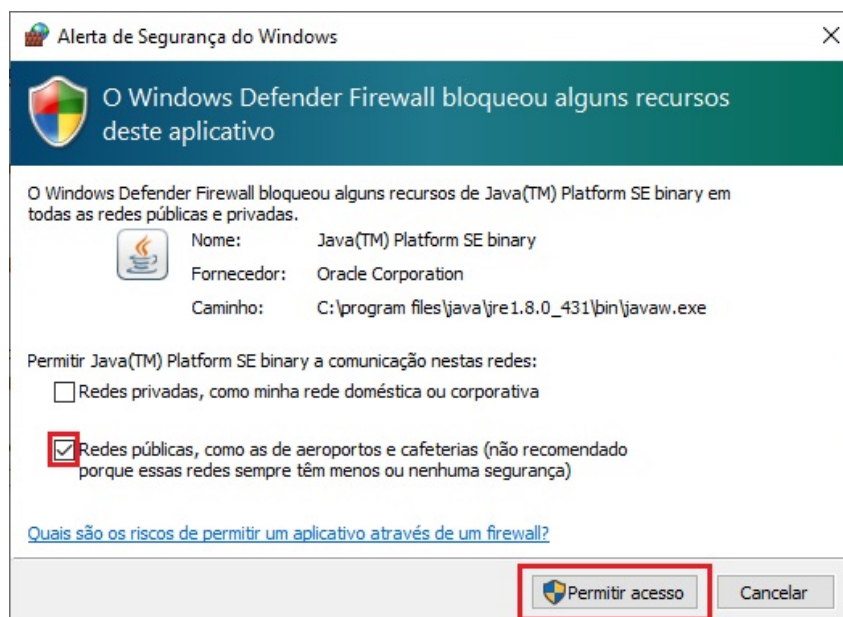


Figura 3.4: Janela do *Windows Defender Firewall* para a execução do TOPCAT pela primeira vez.

Então você verá a Interface inicial do TOPCAT com as suas ferramentas, menus e abas disponíveis para trabalhos com tabelas de dados como mostra a Figura 3.5.

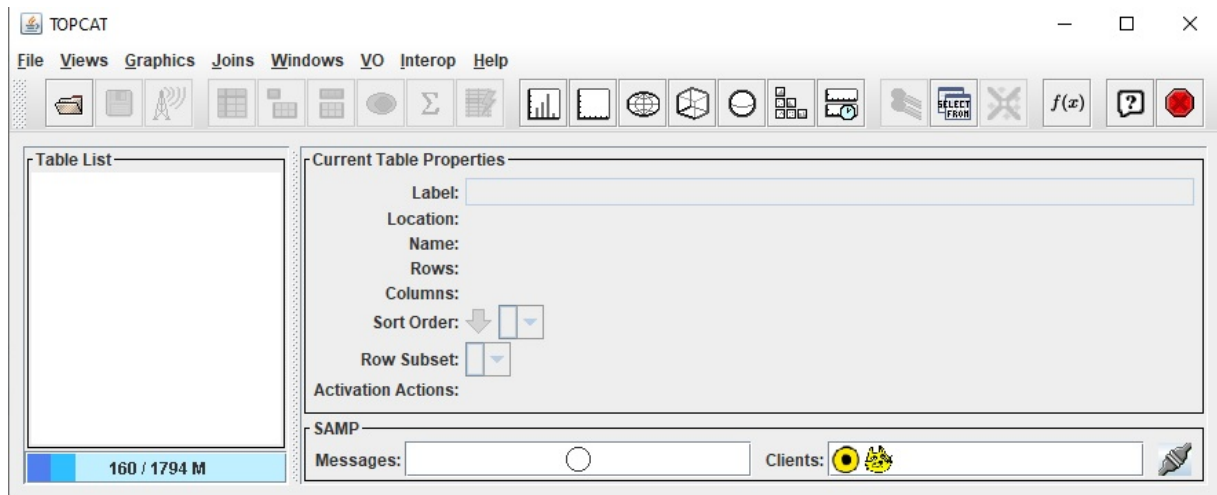


Figura 3.5: Interface gráfica inicial do TOPCAT. O valor apresentado de 160/1794 no canto esquerdo inferior vai depender de cada sistema, pois refere-se ao uso de memória (RAM) pelo programa em relação à memória máxima disponível para ele.

Para a parte prática veja, a partir dos 32min, a video-aula referente ao curso de TOPCAT realizado pelo LIneA em [\[Curso\] TOPCAT](#).