



UNIVERSIDADE ESTADUAL DE CAMPINAS

FACULDADE DE ENGENHARIA ELÉTRICA E DE
COMPUTAÇÃO

IA369Z - REPRODUTIBILITY IN COMPUTATIONAL
RESEARCH

Best Practices for Reproducible Research

PROFA.: LETICIA RITTNER

CARLOS RONY RECALDE VITONERA, RA:229973

CAMPINAS,SP - 1S/2019

POR ONDE COMEÇAR ?

Caro leitor

Se você chegou até aqui é porque como eu você também está interessado e percebe que tem aprendido mais sobre a reprodutibilidade no campo de pesquisa.

Neste diário meu objetivo é dividir com você minha experiência e iniciar a produção do meu artigo, trazendo alguns conselhos que lhes podem ser úteis quando você for produzir o seu próprio artigo. O primeiro passo é ser criterioso ao selecionar um artigo para ser reproduzido. Nesta disciplina, chegamos a uma conclusão unânime de que não existe uma maneira única e correta de se reproduzir um artigo. O que podemos fazer é seguir as regras existentes(a risca), para que possamos nos aproximar do nosso objetivo final.

A ciência busca resultados que possam ser reproduzidos por terceiros. As pesquisas avançam com mais rapidez, quando se torna viável a verificação(prova real) e a disponibilização de resultados obtidos em um artigo, evitando assim, o gasto desnecessário de tempo que se levará para descobrir como chegar em determinado resultado, pois para atingir o mesmo, buscamos caminhos distintos que acabam por resultar em dados superficiais ou não proveitosos.

Sabe-se que, na prática nem todos os artigos produzidos se tornam reproduzíveis, Sendo assim, selecionamos 5 elementos importantes para trabalharmos a reprodutibilidades.

1.Dados

Vão existir muitos momentos em que não teremos acesso a informação de forma pública. E por isso, devemos ser cuidadosos ao escolher o artigo base que desejamos reproduzir.

Dicas:

1. Faça uso de repositórios online, como por exemplo, o Github ou o Bitbucket.
2. Se for possível, escolha dados que possam ser corroborados com os verdadeiros (originais)
3. Verifique se os dados estão completos ou se somente uma parte deles encontra-se disponível.
4. Entenda os tipos de dados que estão disponíveis.
5. Classifique os tipos de dados.

Links de interesse:

<https://github.com>

<https://desktop.github.com/>

2. Código

Outro ponto importante sem dúvida é o código, vem como os problemas ue o uso do código de terceiros pode possuir. Por isso te trago algumas dicas que podem evitar dores de cabeça futuras.

Dicas

1. Use ferramentas de controle de versão.
2. Use a mesma versão em que o código foi escrito.
3. Faça uso de repositórios para armazenar o código.
4. Tente escrever seu código com comentários e atribuir nomes relacionados para facilitar o entendimento.
5. É bom criar um arquivo README que forneça informações importantes para outras pessoas.

3. Documentação

Durante o semestre analisamos diferentes ferramentas que podem ser usadas na reprodutibilidade. Como sugestão, te digo que todos nós interagimos com Jupyter Notebook, uma ferramenta de Literate Computing, que nos permite a combinação de texto e código ao mesmo tempo. Mas é sempre bom ter umas dicas.

Dicas no Jupyter:

1. É mais fácil iniciar o Jupyter a partir do terminal.
2. É melhor criar vários cadernos para cada parte importante do trabalho e, no final, colocá-los juntos.
3. Preste atenção aos nomes atribuídos a cada caderno criado para manter o pedido.
4. Use bibliotecas que podem ser úteis para o seu projeto e que simplifiquem seu trabalho.
5. Seja organizado.

4. Ambiente

Um dos meus principais problemas ao tentar reproduzir um trabalho foi como o fato das versões mudarem. As novas versões que saem nem sempre são compatíveis com as versões antigas utilizadas nos artigos. Como, por exemplo, ao tentar reproduzir um trabalho feito há dois anos utilizando a versão nova dos programas utilizados na mesma, pode-se não obter os resultados esperados, ou na pior das hipóteses, você vai perceber que não há suporte para a versão que foi utilizada no trabalho que você escolheu. Minha dica é.

Dicas

1. Agrupe todas as dependências necessárias para facilitar o trabalho de produção.
2. Quando possível, trabalhe em mais de um ambiente (windows,linux,mac os).
3. Disponibilize o trabalho em um container ou em uma máquina virtual.

4. Tenha um arquivo README contendo as informações importantes.
5. Crie um minitutorial com todos os passos sem se esquecer de nenhum.

Links

<https://www.docker.com/>

<https://www.virtualbox.org/>

5. Workflow

Este último elemento é de grande valia para o entendimento do funcionamento do trabalho de uma outra pessoa. Aqui, de uma maneira gráfica, o processo que é seguido através da reprodução de um trabalho, é explicado. A ideia é poder oferecer um mapa de todo o trabalho, para que outra pessoa possa ser guiada da maneira correta, e assim possa alcançar o resultado esperado. Coloquei no final, por causa que aqui você já vai ter uma ideia clara do funcionamento do seu trabalho e agora vai poder criar um caminho para alguém mais.

Dicas

1. Tenha ideias claras sobre como o seu trabalho funciona.
2. Seja coerente em cada passo.
3. Use ferramentas de desenho.
4. Seja claro e criativo
5. Inclua apenas os pontos importantes, não o código.

Links

<https://www.docker.com/>

<https://www.virtualbox.org/>