

Melhorando a citabilidade de programas de computador para pesquisa com o Citation File Format

Stephan Druskat

Department of German Studies and Linguistics, Humboldt–Universität zu Berlin, Berlin, Germany

stephan.druskat@hu-berlin.de

Raniere Silva

Software Sustainability, School of Computer Science, University of Manchester, Manchester, United Kingdom

raniere.silva@software.ac.uk

Resumo

Programas de computador são parte integral da maioria das pesquisas científicas e seus desenvolvedores, assim como outros atores no processo de investigação científica, contam com citações para mensurar a relevância do seu trabalho. Infelizmente, a prática de citar programas de computador não é comum. Nesse trabalho, os autores apresentam o Citation File Format que tenta facilitar a prática de citar programas de computador utilizados na pesquisa científica.

Palavras-chave: Programa de computador, Crédito, Referência bibliográfica

Improving citability of research software with Citation File Format

Abstract

Software is part of most of scientific research and their developers, like other actors involved in scientific investigation, have their job performance measured by the number of citation their work receive. Unfortunately, cite research software is uncommon. In this work, the authors introduce the Citation File Format which tries to facilitate the practice of cite research software used in scientific investigations.

Key-words: Software, Credit, Reference

Introdução

Hoje, programas de computador são parte integral da maioria das pesquisas científicas, e o seu desenvolvimento tornou-se parte do processo científico. Como um rico produto da pesquisa, programas de computador deveria receber crédito como outros produtos, por exemplo, artigos, livros, conjunto de dados, etc. Entretanto, é comum a omissão ou falta de crédito para programas de computador em publicações, cf., e.g., Howison and Bullard (2016). Isso deve-se parcialmente à falta de reconhecimento ao programa de computador para pesquisa como um componente central no processo de pesquisa, e parcialmente devido à forma estabelecida de creditar pesquisa – citação – que não é ideal para programas de computador. Programas de computador possuem características diferentes dos tradicionais produtos de uma pesquisa. Eles são, por exemplo, dinâmicos e interativos, ao contrário de conjuntos de dados e artigos. Citar um recurso dinâmico é difícil, mas é necessário na ausência de um melhor modelo de distribuição dos créditos. O que torna a citação de programas de computador mais difícil é o fato da informação necessária para sua citação ser menos acessível que, por exemplo, para artigos. Um artigo já inclui a informação relevante para sua citação, como título, autores, data, local e o contexto da publicação. Programas de computar normalmente não fornecem, exceto por seu nome e talvez a versão, semelhante informação de forma fácil ao usuário final.

Princípios específicos para a citação de programas de computador foram definidos e publicados (Smith et al. (2016)) recentemente, e a comunidade acadêmica começou a trabalhar em uma implementação, no FORCE11 Software Citation Working Group (<https://www.force11.org/group/software-citation-implementation-working-group>). Melhorar o acesso à informação relevante para a citação de programas de computador é parte desse esforço. Research Software Engineers e outros desenvolvedores de programas de computador para pesquisa devem ser capazes de fornecer tal informação de maneira que a) garanta a aplicação dos princípios da citação de programas de computador, b) torne a informação acessível para todos seus possíveis usuários, c) não crie barreiras técnicas ou trabalho adicional desnecessário, e d) seja compreendido por máquinas para possibilitar processos automatizados.

Os requerimentos são satisfeitos quando desenvolvedores fornecem um arquivo CITATION, utilizando o Citation File Format (o Autor (2018)), junto com o programa de computador sendo distribuído.

O Citation File Format (CFF) é um formato focado em humanos para o armazenamento da informação necessária para a citação de programas de computador. Ele é compreendido por máquinas – implementado em YAML – como também por ser lido e escrito por humanos.

CFF deriva da ideia de Robin Wilson que para tornar programas de computador citáveis, desenvolvedores deveriam simplesmente, na tradução dos autores, “adicionar um arquivo de texto chamado CITATION no diretório raiz do código fonte, e inserir informação nele cobrindo

como o programa de computador deveria ser citado” (<https://www.software.ac.uk/blog/2013-09-02-encouraging-citation-software-introducing-citation-files>), em analogia com os arquivos README e LICENSE, etc. Com o objetivo de tornar esses arquivos acessíveis ao processamento por máquinas, e alinhá-lo com os princípios de citação de programas de computador, um grupo formado durante o “Workshop on Sustainable Software for Science: Practice and Experiences (WSSSE5.1)” propôs a criação de um formato padrão para os arquivos CITATION (<https://www.software.ac.uk/blog/2017-12-12-standard-format-citation-files>). O Citation File Format é a implementação dessa proposta. Desde seu nascimento em 2017, ele vem sendo adotado pela comunidade de programas de computador para pesquisa, e está sendo utilizado em instituições, projetos e ferramentas. O Netherlands eScience Center, por exemplo, utiliza-o para prover informação para citação no seu Research Software Directory (<https://research-software.nl/>). O formato também é suportado pelo CiteAs.org (<http://citeas.org/>), um serviço fornecendo referências formatadas de identificadores para um diverso grupo de ferramentas de pesquisa, como por exemplo DOIs, URLs, and ArXiv IDs.

O Citation File Format impõe a aplicação dos princípios da citação dos programas de computador por requisitar que a informação necessária seja incluída no arquivo CITATION, que é denominado CITATION.cff.

Abaixo encontra-se um exemplo minimalista de um arquivo CITATION.cff, que em princípio é um mapa de chave-valor.

```
cff-version: 1.0.3
message: If you use this software, please cite it as below.
authors:
  - family-names: Druskat
    given-names: Stephan
    orcid: https://orcid.org/0000-0003-4925-7248
title: My Research Tool
version: 1.0.4
doi: 10.5281/zenodo.1234
date-released: 2017-12-18
```

CFF também fornece chaves para várias outras informações específicas de programas de computador, tais como licença, repositórios, palavras-chave, contato, etc.

O conjunto de ferramentas para a criação e processamento de arquivos CITATION.cff está crescendo, e atualmente inclui geradores, conversores, leitores, e validadores em linguagens de programação como Python, Ruby e Java. Uma lista das ferramentas disponíveis encontra-se disponível no repositório principal do projeto no GitHub (<https://github.com/citation-file-format/citation-file-format>). Todas as ferramentas são

disponibilizadas como programas de computador de código livre/aberto, e – como CFF é um esforço comunitário – contribuidores são bem-vindos em todos os projetos.

Enquanto isso, o processo de citação de programas de computador não termina com o fornecimento da informação para citação junto com o programa. Essa informação deve ser reutilizada por diferentes atores, como, por exemplo, repositórios de pesquisa, índices, resolvedores, e, por último, mas não menos importante, as plataformas de publicação científica. Para garantir que a citação e outras informações sobre um programa de computador estará disponível para todos esses serviços, ela pode ser convertida em um formato próprio para troca: CodeMeta (<https://codemeta.github.io/>). CodeMeta provê um modelo de informação para programa de computador (de forma geral), e um formato de troca como linked data em JSON-LD. O Citation File Format é totalmente compatível com CodeMeta, e uma extensão ao CFF provendo todos os campos definidos pelo CodeMeta está sendo desenvolvida. Desta forma, fornecer a informação para a citação do seu programa de computador em um arquivo CITATION.cff garante que a informação pode ser reutilizada pelos usuários para tornar seu programa de computador visível, e citável, de forma que você receba os créditos que merece.

Referências bibliográficas

DRUSKAT, Stephan. (2017) – Track 2 Lightning Talk: Should CITATION Files Be Standardized? figshare. [Consult. 23 Abr. 2018]. Disponível na Internet: <https://doi.org/10.6084/m9.figshare.3827058.v4>.

HOWISON, James.; BULLARD, Julia. (2016) – Software in the scientific literature: Problems with seeing, finding, and using software mentioned in the biology literature. *J Assn Inf Sci Tec*, 67: 2137–2155. [Consult. 23 Abr. 2018]. Disponível na Internet: <https://doi.org/10.1002/asi.23538>.

SMITH, Arfon M.; KATZ, Daniel S.; NIEMEYER, Kyle E. (2016) – FORCE11 Software Citation Working Group. Software citation principles. *PeerJ Computer Science* 2:e86 [Consult.