

Acknowledging Yade: citation of an open source software in scientific publications.

Bruno Chareyre

February 14, 2011

Abstract

Software development is not well recognized by the academic world, where publications and citations are the universally accepted indicators of researchers activities, even though a software and its documentation can be valuable contributions to research. It can be explained by the fact that softwares are not considered by bibliometry tools. The model proposed by developers of the CGAL project to address the issue and make development and documentation more attractive is presented. The adaptation of this model to YADE platform is proposed and discussed.

Keywords: Yade-DEM, Google Scholar, Citation analysis, Publish or perish, h-index, g-index, Journal impact factor.

1 Introduction

To start with, we could quote the “Acknowledging CGAL” section of CGAL [1], a geometric algorithms library that used in some of Yade’s classes:

“CGAL is implemented for the most part by researchers. The academic world evaluates and rewards researchers for a good part by the analysis of the number of published papers and the number of citations of their papers, which measures their impact.”

Like CGAL, YADE is developed by researchers interested by the Discrete Element Method (DEM). The large majority of them are from the community of engineering sciences and physics. They will usually not consider software development as a primary objective from an academic point of view, as long as this work wouldn’t be recognized by peers.

Therefore, situations where some developments are partly committed, badly documented, or not maintained, are likely to occur: net losses for the project itself, missed opportunity for the developer to show what he did and eventually initiate collaborations with others on this basis. A publication model that implemented by CGAL project recently could serve as

an example for allowing accurate citations of Yade code itself (in fact the documented part) in scientific papers, giving it more value in the academic sense.

Since the documentation includes the description of original algorithms and derivations, and gives access to implementation details of the DEM that remain unrevealed too often in published papers. The proposed model would also let one cite accurately an algorithm that he used but that would not be explained in a journal article.

The CGAL citation model is briefly reviewed below, then an implementation in YADE is detailed and discussed.

2 Citation model

2.1 CGAL citation model

Again, we can quote CGAL[1]:

“In order to make the CGAL project attractive for researchers to contribute their work (which allows users to benefit from new contributions), we are pushing a model where the CGAL manual chapters are considered like publications, and can be cited in articles as such. We therefore kindly ask users to cite CGAL as appropriately as possible in their papers.”

The opportunity of citing (i.e. make more visible) a development that couldn't be in itself the subject of a journal article is not to be neglected, especially in the case of PhD and young researchers. In CGAL model, the idea is that citations are pointing to the documentation of the software or of a precise package.

Different levels of citations are found for CGAL. The highest level is for citing the CGAL project as a whole[1]. Citing the documentation of CGAL[2] is using a different reference, where the documentation is seen as a book, and the author is CGAL Editorial Board. Precise authors names are defined for individual chapters of the book, like the chapter on 3D triangulation [3] (23 citations in Google Scholar - mostly in journals and conferences).

2.2 Implementation in YADE

This model is well adapted to the documentation system implemented by Smilauer [4], since a good part of the documentation is typed in the c++ code directly. In this situation, there can be a good correlation between the author of an implementation and the author of the corresponding documentation ¹. CGAL is defining a total of more than sixty chapters. Such a

¹Here, we are assuming that developers are using the documentation system, else the undocumented code can't be considered as a real contribution

fine granularity would be difficult to apply to Yade presently. It is possible, however, to divide the documentation in three self-contained chapters.

The part of the documentation that is generated on the basis of source code is the concatenation of texts from different authors. This part is referenced as “Yade’s reference documentation“ [5]. It contains a significant part of the documentation, including equations and algorithms descriptions.

Other parts have been written in independant files, using the reStructuredText format. They are regrouped in two chapters. The first one, ”Using and Programing“ [4], explains software design and provides guidance for users and programmers. The other, “Dem Formulation“ [6], is describing the most basic concepts and algorithms of the DEM implemented in Yade.

All three documents are forming Yade documentation [7], published online at yade-dem.org by Yade project, and edited by Smilauer.

Like in CGAL, a higher level of citation can be defined for citing the Yade project as a whole [8]. This last reference is well suited for mentioning Yade project in some occasions (e.g. research proposals). It is perhaps less adapted for citations in scientific papers since it doesn’t define a specific document.

All references defined for Yade are listed as bibtex entries in appendices. The reference documentation, for instance, is defined like this :

```
@InCollection{ yade:reference,
  AUTHOR = "V. Šmilauer and E. Catalano and B. Chareyre and
  S. Dorofenko and J. Duriez and A. Gladky and J. Kozicki and
  C. Modenese and L. Scholtès and L. Sibille and J. Stránský and K. Thoeni",
  BOOKTITLE = "{Y}ade {D}ocumentation",
  EDITOR = "V. Šmilauer",
  PUBLISHER = "The {Y}ade {P}roject",
  TITLE = "{Y}ade {R}eference {D}ocumentation",
  YEAR = "2010",
  EDITION = "1st",
  NOTE = "http://yade-dem.org/doc/",
  ALTNOTE = "\url{http://yade-dem.org/doc/}"
}
```

3 Discussion

3.1 They are not ISI references

The fact that bibliographic references are defined for Yade documentation is not turning it into an ISI-referenced publication, but it should not be the source of excessive pessimism. With the generalization of automated citations counts (ISI Web Of Sciences (ISI), Scopus, Google Scholar (GS),...), new bibliometric indicators based on citations are becoming popular (H factor, G factor,...), while the prestige or the impact factor of the journal is less and less considered a good a priori estimate of the quality and impact

of an article. Recommendations to evaluators are made in that sense in a recent report of the French Academy of Sciences (ref), for instance. This report mentions that 50% of the papers published in *Nature* have a very low citation count. Sometimes, articles with decent citations count are seen as non-scientific by the academic world, like a famous "memory of water" article [9] (GS:368,ISI:227). On the other hand, extremely high citations counts can be found on papers published in highly specialized journals, or with modest impact factors. An example close to us is the 1979 Cundall's paper [10] (GS:2876,ISI:2469).

The fact that the number of citations is considered a better estimate of the importance of a paper is in one sense breaking the popular "publish ISI or perish" model (a model based on the idea that the journals referenced by ISI Web of Knowledge have more value than others). After all, if a paper with a large citation count is not in an ISI-referenced journal, it only denotes the failure of the database to realistically account for the influence of this paper.

At the same time, alternatives like Google Scholar are being developed. GS is referencing a lot more articles than ISI [11], and is used as a database for more advanced bibliometric tools like Harzing's² "Publish or Perish" [12]. Academics are getting used to such alternatives. In reaction probably, ISI started including conference proceedings in WoS database. Clearly, the fact that a reference is not ISI does not mean it will never be considered by peers, if it is cited in conferences or journals.

Google referencing being quite liberal, reviewers tend to give less weight to GS or Harzing PoP counts in comparison with ISI counts. If GS reports citations of a paper that is not found in ISI, it will only gain consideration if the citation count is high enough. Hopefully, with the number of Yade users, it is not unrealistic to expect a rapid increase.

3.2 How would it work?

Google referencing system is based on scans of online contents. It is searching in publishers websites, PhD thesis, and pdf files available online. For instance, citations of a visionary article on Yade design, published online by Šmilauer, are being reported correctly (3 citations).

The content of pdf files is analyzed in order to find, namely, title, authors and affiliations, abstract, and list of references. In order to define valid documents, these items will have to be defined for each of the three documentation volumes. It should let citations in thesis, conferences, and papers be accounted for correctly.

It is probably preferable to keep static pdf files available to help the referencing process and favor uniqueness in reference counts, in the beginning

²Distributed as a small MS Windows executable

at least. It means the documentation referenced in bibliometric databases will not point to the frequently updated versions that one can find at yade-dem.org. However, GS is officially supposed to handle different editions of a same book, so that incrementing the edition number for introducing substantial changes or additional chapters could be considered (with care) in the future. Another advantage of generating static pdf versions is that it will let editorial decisions be taken, regarding for instance the addition of new independent chapters, without being bound by a purely automatized document generation process.

4 Conclusion

The citation of Yade’s documentation is introduced as an incentive to commit in house developments and document them seriously, thus providing an enhanced modelling tool for users. It also gives a way for users to cite accurately algorithms that would not be published in journal papers.

We encourage contributions in the form of self-contained reStructured-Text documents that can be included as new chapters of the documentation, reflecting for instance developments made in the context of PhD thesis.

References

- [1] “CGAL, Computational Geometry Algorithms Library.” <http://www.cgal.org>.
- [2] The CGAL Project, *CGAL User and Reference Manual*. CGAL Editorial Board, 3.7 ed., 2010. http://www.cgal.org/Manual/3.7/doc.html/cgal_manual/packages.html.
- [3] S. Pion and M. Teillaud, “3D triangulations,” in *CGAL User and Reference Manual*, CGAL Editorial Board, 3.7 ed., 2010. http://www.cgal.org/Manual/3.7/doc.html/cgal_manual/packages.html#Pkg:Triangulation3.
- [4] V. Šmilauer, A. Gladky, J. Kozicki, C. Modenese, and J. Stránský, “Yade Using and Programming,” in *Yade Documentation* (V. Šmilauer, ed.), The Yade Project, 1st ed., 2010. <http://yade-dem.org/doc/>.
- [5] V. Šmilauer, E. Catalano, B. Chareyre, S. Dorofenko, J. Duriez, A. Gladky, J. Kozicki, C. Modenese, L. Scholtès, L. Sibille, J. Stránský, and K. Thoeni, “Yade Reference Documentation,” in *Yade Documentation* (V. Šmilauer, ed.), The Yade Project, 1st ed., 2010. <http://yade-dem.org/doc/>.
- [6] V. Šmilauer and B. Chareyre, “Yade DEM Formulation,” in *Yade Documentation* (V. Šmilauer, ed.), The Yade Project, 1st ed., 2010. <http://yade-dem.org/doc/>.
- [7] V. Šmilauer, E. Catalano, B. Chareyre, S. Dorofenko, J. Duriez, A. Gladky, J. Kozicki, C. Modenese, L. Scholtès, L. Sibille, J. Stránský, and K. Thoeni, *Yade Documentation*. The Yade Project, 1st ed., 2010. <http://yade-dem.org/doc/>.
- [8] “Yade: Open source discrete element method.” <http://yade-dem.org>.
- [9] E. Davenas, F. Beauvais, J. Amara, M. Oberbaum, B. Robinzon, A. Miadonna, A. Tedeschi, B. Pomeranz, P. Fortner, P. Belon, *et al.*, “Human basophil degranulation triggered by very dilute antiserum against IgE,” *Nature*, vol. 333, no. 6176, pp. 816–818, 1988.

- [10] P. Cundall and O. Strack, "A discrete numerical model for granular assemblies," *Géotechnique*, vol. 29, no. 1, pp. 47–65, 1979.
- [11] A. Harzing and R. van der Wai, "Google Scholar as a new source for citation analysis," *Ethics in Science and Environmental Politics(ESEP)*, vol. 8, no. 1, pp. 61–73, 2008.
- [12] A. Harzing, "Publish or perish, version 3.1," 2011. available at www.harzing.com/pop.htm.