Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Libresoft tools and main research

David Arroyo Menéndez

CVSAnaly

- Remote analysis and measurement of libre software systems by means of the CVSAnalYtool
  - This paper proposes a methodology for measuring and analyzing re-motely big libresoftware projects using publicly-available data from their version control repositories
- $ git clone https://github.com/jsichi/cvsanaly.git

Sloccount

- Counting potatoes: the size of Debian 2.2
  - Debian is the largest Free Software distribution, with well over 2,800 source packages in the latest stable release (Debian 2.2) and more than 4,000 source packages in the release currently in preparation. But, how large is “the largest”? In this paper, we use David Wheeler’s sloccount system to determine the number of physical source lines of code (SLOC) of Debian 2.2 (aka Potato). We show that Debian 2.2 includes over 56,000,000 physical SLOC (almost twice than Red Hat 7.1, released about 8 months later), showing that the Debian development model (based on the work of a large group of voluntary developers spread around the world) is at least as capable as other development methods (like the more centralized one, based on the work of employees, used by Red Hat or Microsoft) to manage distributions of this size.
-

Perceval

- Send Sir Perceval on a quest to retrieve and gather data from software repositories.
  - $ perceval git 'https://github.com/grimoirelab/perceval.git' --from-date '2016-01-01'
-

Applying Social Network Analysis to the Information in CVS Repositories

- The huge quantities of data available in the CVS repositories of large, long-lived libre (free, open source) software projects, and the many interrelationships among those data offer opportunities for extracting large amounts of valuable information about their structure, evolution and internal processes. Unfortunately, the sheer volume of that information renders it almost unusable without applying methodologies which highlight the relevant information for a given aspect of the project. In this paper, we propose the use of a well known set of methodologies (social network analysis) for characterizing libre software projects, their evolution over time and their internal structure. In addition, we show how we have applied such methodologies to real cases, and extract some preliminary conclusions from that experience

On the Inequality of Contributions to Wikipedia

Wikipedia is one of the most successful examples of massive collaborative content development. However, many of the mechanisms and procedures that it uses are still unknown in detail. For instance, how equal (or unequal) are the contributions to it has been discussed in the last years, with no conclusive results. In this paper, we study exactly that aspect by using Lorenz curves and Gini coefficients, very well known instruments to economists. We analyze the trends in the inequality of distributions for the ten biggest language editions of Wikipedia, and their evolution over time. As a result, we have found large differences in the number of contributions by different authors (something also observed in free, open source software development), and a trend to stable patterns of inequality in the long run.

Applying Social Network Analysis to the Information in CVS Repositories

- The huge quantities of data available in the CVS repositories of large, long-lived libre (free, open source) software projects, and the many interrelationships among those data offer opportunities for extracting large amounts of valuable information about their structure, evolution and internal processes. Unfortunately, the sheer volume of that information renders it almost unusable without applying methodologies which highlight the relevant information for a given aspect of the project. In this paper, we propose the use of a well known set of methodologies (social network analysis) for characterizing libre software projects, their evolution over time and their internal structure. In addition, we show how we have applied such methodologies to real cases, and extract some preliminary conclusions from that experience

Quantitative analysis of the wikipedia community of users

- Many activities of editors in Wikipedia can be traced using its database dumps, which register detailed information about every single change to every article. Several researchers have used this information to gain knowledge about the production process of articles, and about activity patterns of authors. In this analysis, we have focused on one of those previous works, by Kittur et al. First, we have followed the same methodology with more recent and comprehensive data. Then, we have extended this methodology to precisely identify which fraction of authors are producing most of the changes in Wikipedia's articles, and how the behaviour of these authors evolves over time. This enabled us not only to validate some of the previous results, but also to find new interesting evidences. We have found that the analysis of sysops is not a good method for estimating different levels of contributions, since it is dependent on the policy for electing them (which changes over time and for each language). Moreover, we have found new activity patterns classifying authors by their contributions during specific periods of time, instead of using their total number of contributions over the whole life of Wikipedia. Finally, we present a tool that automates this extended methodology, implementing a quick and complete quantitative analysis of every language edition in Wikipedia.

Applying Social Network Analysis to the Information in CVS Repositories

- The huge quantities of data available in the CVS repositories of large, long-lived libre (free, open source) software projects, and the many interrelationships among those data offer opportunities for extracting large amounts of valuable information about their structure, evolution and internal processes. Unfortunately, the sheer volume of that information renders it almost unusable without applying methodologies which highlight the relevant information for a given aspect of the project. In this paper, we propose the use of a well known set of methodologies (social network analysis) for characterizing libre software projects, their evolution over time and their internal structure. In addition, we show how we have applied such methodologies to real cases, and extract some preliminary conclusions from that experience

Contributions at UNED

- ALFANET: Active learning in aLF through Virtual Communities and CSCL. CEN/ISSS Workshop Learning Technologies meeting, Madrid, Spain (2004)
- An active collaborative framework for improving e-learning practices. Proceedings of the 21st world conference on open learning and distance education, lifelong learning in the networked world, International council for open and distance education (ICDE), Hong Kong (2004).
- ALF: Un Entorno Abierto para el Desarrollo de Comunidades Virtuales de Trabajo y Cursos Adaptados a la Educación Superior. Virtual Educa 2004, Barcelona, Spain (2004)
- ALF: Aprendizaje efectivo mediante comunidades virtuales y espacios estructurados de trabajo colaborativo. IV Conferencia Internacional sobre Educación, Formación y Nuevas Tecnologías (Virtual Educa 2003) (2003)
- The INNOVA approach. Eadtu Anual Conference: E-Bologna, 2003., Madrid, Spain, 11/2003.
- ALF: Mejora del aprendizaje a través de un entorno colaborativo basado en comunidades virtuales. 3a Conferencia Internacional de la Educación y de la Formación Basada en las Tecnologías (Online Educa 2003), Barcelona, Spain (2003).
- A dynamic web platform for distance learning research and training. The future of learning - learning for the future. 20th world conference on open learning and distance education, Germany (2001).
- TutorT - UNED: Tutor telemático de la UNED.

More Contributions

- Articles Mundo Linux
  - Lisp (I). Mundo Linux
  - Lisp (II). Mundo Linux
  - Aplicaciones prácticas de Inteligencia Artificial. Mundo Linux
  - Aumenta tu productividad con GNU/Emacs
  - Biblioteca Digital y Web Semántica
- Books translations
  - An Introduction to Programming in Emacs Lisp
  - The compact Org-mode guide
  - An Introduction to GCC