www.er.ethz.ch www.smi.ethz.ch

Master Project (Spring Semester 2009):
Factors of Success and Failure in Open Source Projects

Supervision:
D. Sornette (Professor)
T. Maillart (PhD Student)

Co-Supervision:
G. Von Krogh (Professor)
S. Spaeth (Post-Doc)

This work will be performed in close collaboration with two research groups focusing on Open Source and Software Development:
1. Chair of Strategic Management and Innovation, Prof. Von Krogh, D-MTEC, ETHZ
2. Chair of Software Design, Prof Gall, University of Zurich (data provider)

Introduction:
This work is at the edge of theory and research in economics and management, and emergence of cooperation in complex systems. This project is a unique opportunity to understand quantitatively the emergence of success in new ventures. This work is integrated in a joint research program jointly supervised by Profs. Sornette and von Krogh.

Problem Setting:
One way of measuring the “performance” of any Package in an Open Source Linux Distribution can be measured by the number of other packages, which call it in its routines. It is now established that average connectivity of a given package is increasing following a multiplicative random stochastic process called "proportional growth". This basically means that the more a package is connected the more likely it will gain new connections [2].
However this view of the evolution of Linux Packages is coarse and provides only a general benchmark model for Package behavior. For example it does not explain factors that make a package become essential or on the contrary completely useless.
Our hypothesis is that some external factors (market, utility) as well as internal factors (organization, distribution of work) contribute to the success of an Open Source project. We call this framework Endo/Exo; it is well explained in [1].

Goal:
Uncover factors of success in Open Source Projects by applying the endo/exo framework

Modus Operandi:
The only reliable and quantitative data we have on Open Source Project is CVS (or

Subversion) to track changes in the code and Bugzilla to track debugging process. For the time series, we can rely on versioning on the one side, and success Evolution in Debian Linux on the other side.
These four aspects must be explored within the endo/exos framework as well as a large bench of statistical methods to analyze project dynamics (for instance we would like to verify the existence of long memory processes in time series).

Duration of the Project: 6 months (Master Project), for exceptionally motivated and skilled student it could be partially addressed during a semester project.

Necessary Know How:
1. Ease in statistical analysis
2. Programming skills in SQL, Matlab, knowledge (or will to learn) of Python is a plus
3. General knowledge on Open Source Projects and OSS Community

Bibliography:
[1] R. Crane and D. Sornette, Robust dynamic classes revealed by measuring the response function of a social system, Proc. Nat. Acad. Sci. USA 105 (41), 15649-15653 (2008).

[2] T. Maillart, D. Sornette, S. Spaeth and G. von Krogh, Empirical Tests of Zipf's Power Law Mechanism in Open Source Linux Distribution, accepted in Physical Review Letters, 2008 (http://arXiv.org/abs/0807.0014).

[3] von Krogh et al. Community, joining, and specialization in open source software innovation: a case study. Research Policy (2003)

[4] Haefliger et al. Code Reuse IN Open Source Software. strategy.ch

[5] von Krogh et al. Knowledge reuse in open source software: An exploratory study of 15 open source projects. HICSS'05. Proceedings of the 38th Annual Hawaii …

[6] von Krogh et al. Collective Action and Communal Resources in Open Source Software Development: The Case of Freenet. Academy of Management (2003)

[7] Challet et al. Fat tails, long memory, maturity and ageing in open-source software projects. arxiv.org (2008)

[8] Valverde et al. Logarithmic growth dynamics in software networks. Europhys. Lett (2005)