

Évaluation des réalisations

Paul Zimmermann

Commission d'Évaluation  *INRIA*

(avec J.-P. Merlet, C. Schmid, A. Sulem, D. Margery)

Plan

- contexte : la commission d'évaluation
- groupe de travail "évaluation des réalisations"
- conclusions et propositions

La Commission d'Évaluation (CE)

- évaluation des équipes (séminaires d'évaluation)
- évaluation des chercheurs (titularisation, promotion) (*)
- recrutement des chercheurs (concours CR2, CR1, DR2, ...)

Composition de la CE

<http://www.inria.fr/inria/organigramme/ce.fr.html>

- **12 élus chercheurs** : Anne-Marie Kermarrec, Paul Zimmermann, Cordelia Schmid, Monique Thonnat, Fabrice Rouillier, Pierre Boulet, Luc Segoufin, Marianne Akian, Christian Laugier, Jean-Pierre Merlet, Yannick Toussaint, Jean-Pierre Talpin.
- **4 élus ITA** : Florian Dufour, Franck Yampolsky, Isabelle Allégret, Hugues Leroy
- **8 nommés internes** : Philippe Robert, Alain Viari, Thomas Jensen, Philippe De Groote, Jean Roman, Nicholas Ayache, Jean-Daniel Boissonnat (président), Agnès Sulem (vice-présidente)
- **8 nommés externes** : Vincent Levillain (EADS), Pierre Crégut (France Telecom R&D), Emmanuel Ledinot (Dassault-Aviation), Michel Schmitt (École des Mines de Paris), Ernst Biersack (Professeur Eurécom), Line Garnero (DR CNRS), Christine Gaspin (DR INRA), Rachid Guerraoui (Professeur EPFL).

Contexte

Lancement de 4 groupes de travail en octobre 2006 :

- Évaluation des publications (indices de citation, *h-number*, facteurs d'impact, liste des revues et conférences de classe A, ...) : J.-P. Merlet et Ph. Robert.
- Critères d'évaluation des chercheurs (J.-D. Boissonnat)
- Évaluation de la valorisation (E. Ledinot)
- **Évaluation des réalisations (P. Zimmermann)**

Le groupe de travail “évaluation des réalisations”

Jean-Pierre Merlet (Coprin, Sophia) : robots séries et parallèles (*)

Cordelia Schmid (Lear, Rhône-Alpes) : appariement d'image, reconnaissance d'objet, description vidéo

Agnès Sulem (Mathfi, Rocquencourt) : Premia (calcul d'options financières)

David Margery (Paris, Rennes) : gforge, Kerrighed, OAR(2)

Paul Zimmermann (Cacao, Lorraine) : MPFR, GMP-ECM

Objectifs

À court terme :

- aider les **candidats** à mieux décrire leurs réalisations
- aider les **rapporteurs** à mieux évaluer/comparer les réalisations (*)

Objectifs

À court terme :

- aider les **candidats** à mieux décrire leurs réalisations
- aider les **rapporteurs** à mieux évaluer/comparer les réalisations (*)

À plus long terme :

- réfléchir à la mise en valeur des réalisations à l'INRIA

`www.inria.fr/valorisation/logiciels`

Documents de référence

Développement de logiciels à l'INRIA. Objectifs et organisation, G. Giraudon, septembre 2001.

Contribution à la réflexion sur le développement logiciel à l'INRIA, Éric Gautrin, Bernard Martin, décembre 2002.

Évaluation des activités non académiques. Avis sur le texte « Recommandations aux évaluateurs » du 29/04/02, E. Ledinot, février 2004, 3 pages.

Sur l'évaluation, Paul-Louis George, janvier 2007.

Note de Giraudon

Développement de logiciels à l'INRIA. Objectifs et organisation, G. Giraudon (directeur du développement et des relations industrielles), septembre 2001, 7 pages.

<http://www.inria.fr/interne/dirdri/developpement/logiciel/developpement-logiciel.pdf>

« Dès qu'une personne quitte une équipe, le code qu'elle a développé peut rarement être maintenu ; les échanges de programmes entre projets sont rares ; le transfert de technologie vers l'industrie est difficile dès lors qu'il porte sur du logiciel et qu'aucune start-up n'a été créée pour exploiter le résultat des recherches ; si la diffusion d'un logiciel rencontre une audience certaine, le projet de recherche est vite dépassé par le succès, ... »

-
- Quels sont les types de logiciels développés dans l'institut et pour quels usages ?
 - Qui développe des logiciels et dans quel cadre ?
 - Quels bilans peut-on en tirer ?

[. . .] l'INRIA a des difficultés à rassembler et à mobiliser des forces de développement même avec son personnel permanent de type ingénieur

- Comment améliorer ce cadre, quelles retombées peut-on en attendre ?

[. . .] il est souhaitable de créer dans chaque Unité de Recherche un service identifié assurant le support au développement

Rapport Gautrin-Martin

Contribution à la réflexion sur le développement logiciel à l'INRIA. Rapport et recommandations, Éric Gautrin, Bernard Martin, décembre 2002, 36 pages,

`http://www.inria.fr/interne/dirdri/developpement/contributionsGM/mission-gautrin-martin.pdf`

*Les rencontres avec les projets de recherche ont montré que le développement de logiciel était considéré par la majorité d'entre eux comme un **élément clé** dans leur activité de recherche*

*[...] activité **peu valorisée** par les instances décisionnaires de l'institut*

*[...] Des **résultats concrets sont attendus** en termes de moyens et de politique scientifique*

*[...] il reste la question ouverte du « **comment** » **faire une évaluation des logiciels** au même titre que les publications*

-
- le développement logiciel dans les projets de recherche (**motivations**, processus, environnement)
 - les métiers du développements logiciel (fonctions, catégories, carrières, formation)
 - la diversité des besoins et attentes (moyens humains, organisation, environnements)

Motivations :

- valider une idée ou une théorie
- trouver de nouvelles problématiques
- réutiliser des résultats
- diffuser des résultats
- créer une start-up
- se faire évaluer à l'INRIA

les chercheurs [...] considèrent être essentiellement évalués sur des critères de publications, sans réelle prise en compte des activités de développement.

Recommandations : mettre en place . . .

1. des **structures de support** aux développements logiciels dans chaque UR
2. une coordination des structures de support avec une **animation au niveau national** (*)
3. les éléments de base d'un processus de **développement logiciel cohérent**
4. un **plan de formation** aux techniques et méthodes de développement logiciel
5. définir des **critères de mesure** du « succès » d'un développement logiciel
mise en place rapide d'un groupe de travail sous la responsabilité des directions scientifiques et du développement et des relations industrielles pour la définition de critères permettant d'évaluer l'intérêt, l'impact et le succès d'un logiciel
6. privilégier l'affectation des **ingénieurs permanents** pour les projets de recherche
7. veiller à la **qualité du recrutement** des ingénieurs permanents

Note Ledinot

Évaluation des activités non académiques. Avis sur le texte « Recommandations aux évaluateurs » du 29/04/02, E. Ledinot, février 2004, 3 pages.

Lettres de recommandation sur un transfert industriel devraient préciser :

- quelles personnes de l'entreprise utilisent le logiciel (ou la réalisation) transféré :
quelques personnes d'un service de R&D (stade 1), ou bien quelques personnes d'une **équipe opérationnelle** (stade 2), ou encore **plusieurs équipes opérationnelles** (stade 3) ?
- quels **investissements** l'entreprise a consentis autour du logiciel transféré ?
- quels **gains** (mesurés ou escomptés, qualitatifs ou quantitatifs) sont apportés à l'activité de l'entreprise par le logiciel transféré ?

« Un transfert attesté de type “stade 3” [...] doit pouvoir peser aussi lourd [...] que quelques très bonnes publications académiques ».

Note Ledinot (suite)

Difficulté d'évaluer la contribution d'un candidat dans une réalisation à **plusieurs auteurs** (spécification et architecture, parties algorithmiques sophistiquées mais peu volumineuses en lignes de code, ou au contraire parties d'interface très volumineuses mais sans rapport profond avec l'activité de recherche).

L'objectif principal est de détecter les « **faux positifs** » dans la section *réalisations* des dossiers, afin de ne pas pénaliser les chercheurs qui réalisent vraiment.

Note George

Sur l'évaluation, Paul-Louis George, janvier 2007, 2 pages.

Indique les limites de l'évaluation des publications, conférences et logiciels.

Note George

Sur l'évaluation, Paul-Louis George, janvier 2007, 2 pages.

Indique les limites de l'évaluation des publications, conférences et logiciels.

Qui utilise mon logiciel ?

Moi et moi seul, mon voisin (avec moi), mon voisin (sans moi), un groupe (avec moi), un groupe (sans moi), des gens que je ne connais pas et qui ne me connaissent pas. ()*

Note George

Sur l'évaluation, Paul-Louis George, janvier 2007, 2 pages.

Indique les limites de l'évaluation des publications, conférences et logiciels.

Qui utilise mon logiciel ?

Moi et moi seul, mon voisin (avec moi), mon voisin (sans moi), un groupe (avec moi), un groupe (sans moi), des gens que je ne connais pas et qui ne me connaissent pas. ()*

Quelle est la pérennité de mon logiciel ?

Ma thèse finie, il disparaît, mon contrat fini, il disparaît, mon papier fini (et publié), il disparaît ()*

Que fait mon logiciel ?

Il me permet de montrer un bel exemple (pour ma thèse, mon papier, une conf, ...). Pour tourner un autre exemple, je dois le tuner. Il tourne sur des cas balisés mais ne marche pas réellement sur des cas réels. Il tourne sur des cas que je ne connais pas.

Que fait mon logiciel ?

Il me permet de montrer un bel exemple (pour ma thèse, mon papier, une conf, ...). Pour tourner un autre exemple, je dois le tuner. Il tourne sur des cas balisés mais ne marche pas réellement sur des cas réels. Il tourne sur des cas que je ne connais pas.

Quelle est sa diffusion ?

*Pas de diffusion (moi et mon voisin), une diffusion dans un cercle réduit, une diffusion via le Net, une diffusion dans une boîte, une diffusion **via** une boîte, une diffusion via une intégration dans un “gros” code.*

Résultats

- version 1.0 (novembre 2006)
- version 2.0 (janvier 2007), présentée à la CE du 23 février
- rapport final (3.1) du groupe de travail (mars 2007)

<http://www.loria.fr/~zimmerma/ce/realisations.html>

Qu'est-ce qu'une réalisation ?

- une réalisation informatique (logiciel)
 - une plate-forme de réalité virtuelle
 - un CAVE
 - un robot
 - un volant de retour d'effort
 - un prototype destiné au domaine médical
 - une activité de standardisation (avec prototype, ex. démonstrateur IETF) ?
- ⇒ **pas uniquement des logiciels !**

Réalisations informatiques

- un logiciel
- une base de données (LEFFF, base d'images Caltech 101) (*)
- une *test suite*, un benchmark (*Frisco test suite*)
- un challenge (*RSA challenge, Visual Object Classes Challenge*)

Réalisations innovantes

S'inscrivent dans une activité de recherche.

Évaluées en tant que telles.

Robot de J.-P. Merlet (1988), bug du Pentium (1994) (*).

Objet de **recherche** ou de **transfert**

Cf note de Giraudon :

- **objet de recherche** (démonstrateur) : valider une théorie, prouver l'efficacité d'un algorithme, en général pas diffusé.
- **objet de transfert** (exploitation) : diffusé ou commercialisé. Durée de vie plus longue.

Évaluation

Paul Louis George :

- **que** fait mon logiciel ?
- **qui** l'utilise ?
- **quelle** est sa diffusion ?

Diffi culé vs impact

Ne vont pas forcément de pair.

- un soft très simple peut avoir un grand impact (seti@home)
- un logiciel de « haute technicité » peut avoir un faible impact (*)

Évaluation on-line vs off-line

Un logiciel peut (en général) être testé par l'évaluateur (compilation, reproduction des exemples du manuel, ...)

Pas toujours possible : exemple robot humanoïde de Toulouse.

⇒ importance des avis d'utilisateurs.

Autres critères

- volume de la réalisation (nombre de lignes, de composantes)
- caractère innovant : nouvelles fonctions, gain en temps et/ou mémoire, open source ?
- impact (pour moi, mon voisin, les autres)
- temps passé (intégration pour une réalisation matérielle)
- temps total dans le cas de plusieurs versions

Distinguer

- la **contribution individuelle**

⇒ évaluer la contribution du candidat, et non la réalisation elle-même (*)

idée, conception, création, documentation, maintenance, ...

- l'**évaluation globale** d'une réalisation

Contribution individuelle

Exemple : les dossiers de candidature CR.

4) RÉALISATION ET DIFFUSION DE LOGICIELS

Donner des informations les plus précises possibles : taille des logiciels, difficultés et durée du développement, co-auteurs éventuels, dépôts éventuels, mode de diffusion, impact, nombre d'utilisateurs, principales utilisations connues, etc. Pour les principales réalisations, il est demandé d'en donner une description succincte et le nom des auteurs (avec leurs contributions approximatives en pourcentages), et de préciser leur positionnement par rapport à l'état de l'art, leur diffusion et leur impact.

Propositions

- avoir de vraies **fiches détaillées** pour les **réalisations les plus importantes** dans les dossiers (comme pour les publications)
- exemple : maxi 2 fiches pour les CRs, 5 pour les DRs
- format commun à tous \Rightarrow comparaison plus facile
- pourrait être repris pour les **fiches de synthèse** (évaluation des équipes)

Exemple de fiche détaillée

- Description courte. Logiciel de recherche ou de transfert ? Prototype ou logiciel à vocation plus pérenne ? Logiciel généraliste (ex. MATLAB) ou spécialisé (boîte à outils MATLAB pour les équations diff. non linéaires).
- Positionnement vs état de l'art (nouveau, originalité)
- Auteurs (%age, temps consacré, conception, réalisation, maintenance)
- Difficulté (connaissances nécessaires, choix techniques, encadrement) et principaux résultats obtenus ou à obtenir
- Diffusion (publications, brevet, dépôt APP, page web, documentation, licence, tutoriaux)
- Validation (utilisation dans un projet, par d'autres, transfert industriel, ...)
- Impact sur sa propre recherche et sur les pairs (mailing lists, forum de discussion, ...).
Publications liées à cette réalisation.
- Divers (évolution dans le temps, ...)

Conclusions (évaluation individuelle)

- manque cruel d'outils **quantitatifs ou qualitatifs**
 - difficulté de connaître la contribution d'**une personne**
 - idem pour l'**impact à long terme**
- ⇒ les candidats doivent donner le plus d'information possible
- ⇒ mais attention à ne pas surévaluer les réalisations !

Évaluation globale d'une réalisation

Plus difficile à évaluer que les publications scientifiques :

- grande diversité
- pas d'évaluation par les pairs (en général)

⇒ publication logicielle ?

Publications logicielles : ce qui existe

- ACM Transactions on Mathematical Software, soumission ACM Algorithms (avec code)
- workshop satellite ALENEX (Algorithm Engineering & Experiments) de la conférence SODA (ACM Symposium on Discrete Algorithms)
- workshop satellite WAE (Workshop on Algorithm Engineering) de la conférence ESA (European Symposium on Algorithms)
- conférence TACAS (*Tools and Algorithms for the Construction and Analysis of Systems*) : *tool demonstration papers*
- *tool presentations* de la conférence CAV (*Computer Aided Verification*)
- logiciels ayant mis en place une évaluation propre : CGAL, GAP, distribution Fedora de Linux.

ACM Algorithms

*ACM Algorithms are published to make the fruits of software research readily available to as wide an audience as possible. An algorithm must either provide a capability **not readily available** or perform a task **better in some way than has been done before**. “Better” can mean anything from **improved reliability or efficiency to more attractive packaging**. In all cases, an algorithm must be of lasting value and must represent a **substantial contribution** in terms of the amount of work or the originality required for its creation.*

[...]

*An ACM algorithm must be **complete, portable, well documented, and well structured**. The meaning of these terms is clarified below. [...]*

Conférence TACAS

*Submitted **tool demonstration papers** must : [. . .] have an appendix [. . .] that provides a detailed description of : how the oral presentation will be conducted, e.g. illustrated by a number of **snapshots** ; the **availability of the tool**, the number and types of **users**, other information which may illustrate the **maturity and robustness** of the tool ; if applicable, a link to a web-page for the tool. (The appendix will not be included in the proceedings, but during the evaluation of the tool demonstration papers it will be equally important as the pages submitted for publication in the proceedings.) ; clearly describe the **enhancements and novel features** of the tool in case that one of its previous versions has already been presented at meetings or published in some form.*

CGAL

CGAL a un vrai **editorial board**.

*Submissions of new packages are handled similar to a submission to a **scientific journal**. The CGAL Editorial Board designates a **primary reviewer** who typically makes several iterations with the package author. (*)*

GAP

*Materials submitted for formal acceptance as part of the GAP distribution, typically packages or proposed additions to the GAP library, will undergo an additional review resembling the **refereeing process for journal articles**. As with submitted journal articles, authors will have the option during the review process of withholding their contributions from being made public in the /deposit directory. [...] The intent of the formal review process is [...] to provide authors with **recognition for their professional contributions** more or less equivalent to that for **publication of a journal article**.*

Fedora

*In order for a new package to be accepted into CVS, that package must first undertake a **formal review**. The purpose of this formal review is to try to ensure that the package meets the **quality control requirements** for Fedora. This does not mean that the package (or the software being packaged) is perfect, but it should meet baseline minimum requirements for quality.*

Formal software review

Cf Wikipedia :

Formal versus informal reviews : “Formality” identifies the degree to which an activity is governed by agreed (written) rules. Software review processes exist across a spectrum of formality, with relatively unstructured activities such as “buddy checking” towards one end of the spectrum, and more formal approaches such as walkthroughs, technical reviews, and software inspections, at the other. IEEE Std. 1028-1997 defines formal structures, roles, and processes for each of the last three (“formal peer reviews”), together with software audits.

Research studies tend to support the conclusion that formal reviews greatly outperform informal reviews in cost-effectiveness. Informal reviews may often be unnecessarily expensive (because of time-wasting through lack of focus), and frequently provide a sense of security which is quite unjustified by the relatively small number of real defects found and repaired.

Proposition

Création d'une **revue électronique internationale** dédiée aux logiciels (pas uniquement INRIA) :

- vrai comité éditorial
- arbitrages anonymes
- limiter à une thématique scientifique ?
- étendre aux autres réalisations informatiques (benchmarks, challenges) ? Aux réalisations non-informatiques ?