ECRTS 2012

WATERS

Panel: Software and Academic Research: are we going in the right direction?

G. Lipari

École Normal Supérieure de Cachan, France

July 10 2012

WATERS

Outline

WATERS

How to succeed in your academic career

A few advices to young PhD students:

1. Write papers

How to succeed in your academic career

A few advices to young PhD students:

1. Write papers
2. (*lots* of them!)

How to succeed in your academic career

A few advices to young PhD students:

1. Write papers
2. (*lots* of them!)
3. Have your papers accepted in *good* journals with *high impact factors*

How to succeed in your academic career

A few advices to young PhD students:

1. Write papers
2. (*lots* of them!)
3. Have your papers accepted in *good* journals with *high impact factors*
4. Have your papers accepted in *important* conferences

Writing papers

How to write a successful paper?

- It should contain <u>at least</u> one equation
- It should be *new* and *original* research
- It should be understandable by the reviewers
  - This means the equations should be simple
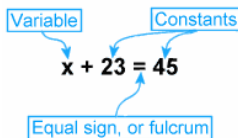 - better if straightforward

$$\sigma_{ed}(\mu,E) = \sigma_{ed}(\mu,E)$$
$$- \frac{2\eta}{1-\mu}\text{Re}\left\{\exp\left(i\eta\ln\frac{1-\mu}{2}\right)\sum_{\ell=0}^{\text{NL}}\frac{2\ell+1}{2}a_\ell(E)P_\ell(\mu)\right\}$$
$$+ \sum_{\ell=0}^{\text{NL}}\frac{2\ell+1}{2}b_\ell(E)P_\ell(\mu)$$

$$\sigma_{ei}(\mu,E) = \sigma_{ei}(\mu,E)$$
$$- \frac{2\eta}{1-\mu^2}\text{Re}\left\{\sum_{\ell=0}^{\text{NL}}\left[(1+\mu)\exp\left(i\eta\ln\frac{1-\mu}{2}\right)\right.\right.$$
$$+ (-1)^\ell(1-\mu)\exp\left(i\eta\ln\frac{1+\mu}{2}\right)\left.\left.\right]\frac{2\ell+1}{2}a_\ell(E)P_\ell(\mu)\right\}$$
$$+ \sum_{\ell=0}^{\text{NL}}\frac{4\ell+1}{2}b_\ell(E)P_{2\ell}(\mu)$$

Writing papers

How to write a successful paper?

- It should contain <u>at least</u> one equation
- It should be *new* and *original* research
- It should be understandable by the reviewers
    - This means the equations should be simple
    - better if straightforward

Theoretical vs. System research

Q: To increase the chances of being a successful researcher, should I be a theoretician or a *system* researcher?

Theoretical vs. System research

Q: To increase the chances of being a successful researcher, should I be a theoretician or a *system* researcher?

My personal impression: theory may be more difficult, but it is easier to build an academic career with theoretical papers

- In our conferences and journals, there is sometimes too much emphasis on theory and too much emphasis on "new" and "original" research
- Unfortunately, too often such "new" and "original" research turns out to be "not useful"
- in other words, the *impact* of our theoretical research is minimal
- Why?

Defining impact

- The impact of a paper is not limited to "real products"
- the paper may positively influence other papers
    - A small progress in a complex theoretical problem can open the way to other theoretical and system papers
    - An implementation of a theoretical algorithm can show strengths and weaknesses
- the paper just present negative results
    - For example, a negative proof, a proof of impossibility
    - A simulation can show that the gains are minimal, or implementation is impractical
- However, we do not often see comparative analysis or implementation papers
    - "It's just implementation! It's just a comparison!"
    - They look like "second class" papers, good for second choice conferences and journals

Other research areas

- Many of the problems described above are common to many scientific research areas
- However, consider one of the most cited papers in Nature with 3,077 citations:

  Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls

- Especially in biology and medicine, many paper are reports of experimental studies, with interpretation of the results
    - Nothing "new" and "original", apparently
    - However, few people question the usefulness of these studies
- Toward experimental Computer Science?

Scientific papers on the rise

*The scale of the academic universe is dizzying [...].
Our most recent estimate is that there are over 24,000
academic journals in existence, 1.3 million academic
papers published every year, and over 50 million
papers published since scholarship began. And for
every one of these 50 million papers there will be
unknowable quantities of blind alleys, abandoned
experiments, conference presentations, work in
progress seminars, and more.*

Ben Goldacre, "Bad Science" blog

Summary

- We are witnessing:
    - a steady increase in the number of published papers
    - a steady increase in the number of conferences and journals
- Many of these paper have minimal impact
- Sometimes too much emphasis on theory and too much emphasis on "original" research
    - Too often, research is made just to publish yet another paper
- No emphasis on good engineering and design

WATERS

Outline

WATERS

Writing software in academy

Today, scientists who write and release code often get little recognition for their work. Someone who has created a terrific open source software program that's used by thousands of other scientists is likely to get little credit from peers.

"It's just software!" is the response many scientists have to such work. From a career point of view, the authors of the code would have been better off spending their time writing a few minor papers that no one reads. This is crazy: a lot of scientific knowledge is far better expressed as code than in the form of a scientific paper

Michael Nielsen, "Reinventing Discovery"

Reusing software

- In order to write a paper, it is often needed to perform some simulation comparing different algorithms
- The most common approach is to *start from scratch every single time*
  - Every PhD student writes its own simulator or little analysis program
  - The code is forgotten and "thrown away" after the simulation results have been obtained
  - Nobody looks at the code (except for the student who wrote it), so nobody can *validate* the results of the simulation
  - Nobody can reuse the code
- Basically, this is "disposable code", that it is used once and then thrown away
- A very sad state of affairs for Computer Scientists!

Goals

- The goal of the PhD student is to get a job
  - Hopefully an academic job!

WATERS

Goals

- The goal of the PhD student is to get a job
  - Hopefully an academic job!

- And to succeed, she/he must write papers and be cited

WATERS

Goals

- The goal of the PhD student is to get a job
  - Hopefully an academic job!

- And to succeed, she/he must write papers and be cited
- No wonder it does not have time to maintain software

Consequences

- Difficult to interact with real engineers (we cannot pretend they spend so much time reading papers)
    - Too little connection to industry
    - Low impact
- If we (professors) do got give value to software design, neither our PhD students will do

Consequences

- Difficult to interact with real engineers (we cannot pretend they spend so much time reading papers)
    - Too little connection to industry
    - Low impact
- If we (professors) do got give value to software design, neither our PhD students will do

- For "applied" computer scientists this is terrible!
    - We are educating generations of future professionals that do not give value to software
    - Also, there is less possibility for them to find a job in industry

COUNSELING CENTER

I HAVE THIS FEAR OF THE REAL WORLD...

© Original Artist
Reproduction rights obtainable from
www.CartoonStock.com

search ID: hscn1985

WATERS

Paper oriented research

- Another problem is concerned with the way scientific publication is done today

Paper oriented research

- Another problem is concerned with the way scientific publication is done today
  - "Papers" are still mainly printed in paper, or released as PDF files

Paper oriented research

- Another problem is concerned with the way scientific publication is done today
  - "Papers" are still mainly printed in paper, or released as PDF files
  - PDF is a nice format for printing, but not for on-line fruition!

Paper oriented research

- Another problem is concerned with the way scientific publication is done today
  - "Papers" are still mainly printed in paper, or released as PDF files
  - PDF is a nice format for printing, but not for on-line fruition!
- If any code is used in the paper, it must be provided separately in a web site

WATERS

Paper oriented research

- Another problem is concerned with the way scientific publication is done today
    - "Papers" are still mainly printed in paper, or released as PDF files
    - PDF is a nice format for printing, but not for on-line fruition!
- If any code is used in the paper, it must be provided separately in a web site
    - It would be much better to distribute the code along with the paper

Paper oriented research

- Another problem is concerned with the way scientific publication is done today
  - "Papers" are still mainly printed in paper, or released as PDF files
  - PDF is a nice format for printing, but not for on-line fruition!
- If any code is used in the paper, it must be provided separately in a web site
  - It would be much better to distribute the code along with the paper
- Same thing happens for data: figures cannot be analysed further, because it is not possible to get the raw data from the PDF

Paper oriented research

- Another problem is concerned with the way scientific publication is done today
  - "Papers" are still mainly printed in paper, or released as PDF files
  - PDF is a nice format for printing, but not for on-line fruition!
- If any code is used in the paper, it must be provided separately in a web site
  - It would be much better to distribute the code along with the paper
- Same thing happens for data: figures cannot be analysed further, because it is not possible to get the raw data from the PDF
  - It would be better to distribute the data that produced the figure along with the paper

Paper oriented research

- Another problem is concerned with the way scientific publication is done today
  - "Papers" are still mainly printed in paper, or released as PDF files
  - PDF is a nice format for printing, but not for on-line fruition!
- If any code is used in the paper, it must be provided separately in a web site
  - It would be much better to distribute the code along with the paper
- Same thing happens for data: figures cannot be analysed further, because it is not possible to get the raw data from the PDF
  - It would be better to distribute the data that produced the figure along with the paper

WATERS

Paper oriented research

- Another problem is concerned with the way scientific publication is done today
  - "Papers" are still mainly printed in paper, or released as PDF files
  - PDF is a nice format for printing, but not for on-line fruition!
- If any code is used in the paper, it must be provided separately in a web site
  - It would be much better to distribute the code along with the paper
- Same thing happens for data: figures cannot be analysed further, because it is not possible to get the raw data from the PDF
  - It would be better to distribute the data that produced the figure along with the paper
- The scientific publication process has fallen behind the Internet revolution

Summary

- Researchers are professionals that try to maximize their ROI (Return On Investment)

    -
    - Revenues depends on publications, h-index, etc.
    - Thus, researchers seek to maximize the number and the quality of their publications
- Any other activity is a waste of time (from the researcher point of view)
    - Nobody wants to waste time on "programming"
    - because programming is not recognised by our community, something that brings no immediate return on investments
- The publication process does not help

WATERS

Outline

WATERS

A dream

- We would like to change the academic culture (at least a little)
- We would like academics to give value to good SOFTWARE
  - PhD students write software for simulation, analysis, comparison, reusing and improving existing code
  - Their work should be acknowledged (when it's good, of course)
- We would like academics to give value to not-so-new research that is well done, well designed and well implemented
- The publication process should change
  - (I'm not even talking about Open Access, just the technical process of publication and distribution)

WATERS

Outline

WATERS

Experiences

- Internet, Web 2.0, and collaborative tools open new perspectives to science
- A few initiatives
  - Open collaborative research
  - Citizen science
  - Open data projects
- CS research has not been involved in this kind or experiments, yet
- Regarding publication:
  - Many open access journals
  - (but still distributing mostly pdf files)

WATERS

Experiences

- The PolyMath project web site:
  `http://polymathprojects.org/`

Citizen science – astronomy
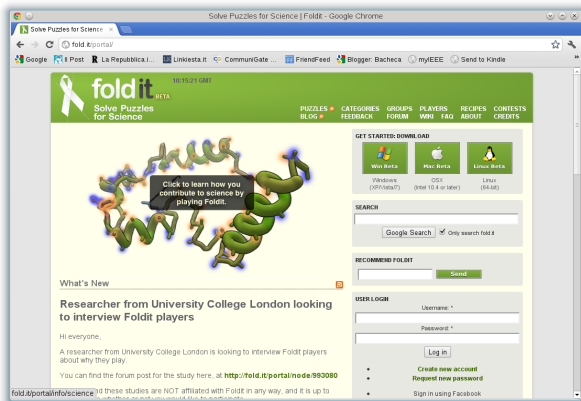
- Citizens helps scientists `http://www.galaxyzoo.org/`
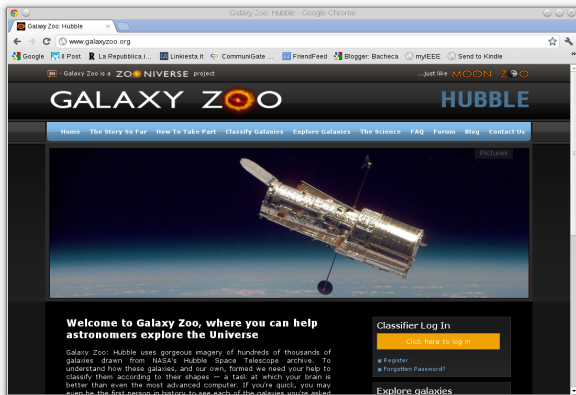
Citizen science – molecular biology

- Fold proteins by playing a game
  `http://fold.it/portal`

Outline

WATERS

Discussion

- Comments?
  - Other problems that come to your mind?
  - What's your experience with the publication process?
  - What's your experience with writing and re-using software?
  - What can be learnt from other research domains?
- Ideas?
  - How to make the dream come true?
  - How to make our community grow larger?
  - How to *publish software* and *data* in the academic community?