Planning and Executing Practice-Impactful Research

Tao Xie

Department of Computer Science
University of Illinois at Urbana-Champaign

taoxie@illinois.edu

http://taoxie.cs.illinois.edu/

Industry-Academia Collaboration

产学研结合：机会与挑战

谢　涛
美国伊利诺伊大学香槟分校

关键词：产学研结合　科研价值

　　科研人员，尤其是与技术关联紧密的学科科研人员，长期面临一个问题——如何让产学研相结合。产学研涉及产业界、教学和科研等不同领域，不同的主体对产学研的认识和期望也不相同，相关的研究和讨论也很多。本文从学术界科研人员的角度，

出了各类CCF奖项和荣誉，这些在一定阶段成为了科研努力的目标和方向。

　　科研评价是在不断发展变化的，如果我们一直停留在现有的评价体系中，我们可能就会在未来新的评价标准中落伍。近期科研评价出现了一些新的变化值得

科研人员带来启发和新思路，开创新的研究领域或大力推动某领域的学术发展。

　　这一点与以往基本相同，在美国计算技术研究协会的报告中，不同之处在于淡化或忽略了论文数量，而是侧重考查论文中科研想法的创新和启发价值。很

http://taoxie.cs.illinois.edu/publications/cccf15-industryacademia.pdf

FSE 2016 Panel: The State of Software Engineering Research

on 01 Dec 2016

Tao Xie
University of Illinois at Urbana-Champaign
Tao is an ACM distinguished researcher. His Research focuses on automated software testing, mobile security, and software analytics.

Laurie Williams
North Carolina State University
Laurie is a founder of the Extreme Programming / Agile Conference. Her research focuses on software security, testing, and agile programming.

Peri Tarr
IBM Research
Peri is a principal research staff member at IBM TJ Watson Lab and a technical lead for Cognitive Tools and Methods at IBM. Her research focuses on software composition and aspect oriented software development.

Prem Devanbu
University of California at Davis
Prem started his career as an industrial software developer, then worked at Bell Labs and AT&T Research before beginning to teach at University of California at Davis. His research focuses on empirical software engineering, naturalness of software, and social analytics.

Lionel Briand
University of Luxembourg
Lionel currently leads the Software Verification and

Evolution of Research Assessment

- #Papers ➜
- #International Venue Papers ➜
- #SCI/EI Papers ➜
- #CCF A (B/C) Category Papers ➜
- ???

CRA 2015 Report:
"Hiring Recommendation. Evaluate candidates on the basis of **the contributions in their top *one or two* publications**, …"

"Tenure and Promotion Recommendation. Evaluate candidates for tenure and promotion on the basis of **the contributions in their most important *three to five* publications** (where **systems** and other **artifacts** may be included)."

http://cra.org/resources/best-practice-memos/incentivizing-quality-and-impact-evaluating-scholarship-in-hiring-tenure-and-promotion/

Societal Impact

ACM Richard Tapia Celebration of Diversity in Computing

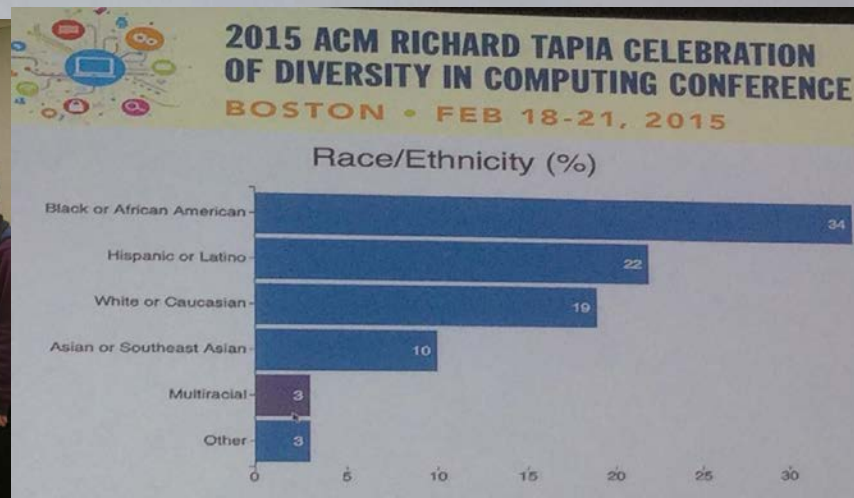http://tapiaconference.org/

Join us at the next Tapia Conference in Atlanta, GA on September 20-23, 2017!

Margaret Burnett: "Womenomics & Gender-Inclusive Software"
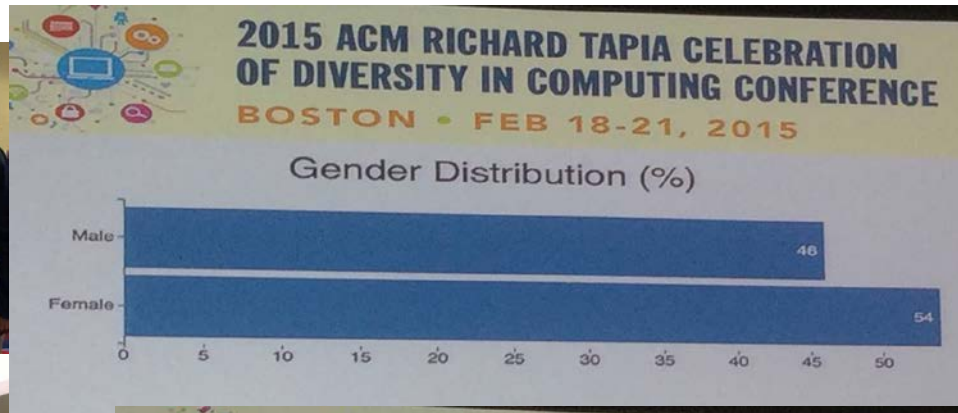
Andy Ko: "Why the Software Industry Needs Computing Education Research"

"Because anybody who thinks that we're just here because we're smart forgets that we're also privileged, and we have to extend that farther. So we've got to educate and help every generation and we all have to keep it up in lots of ways."
– David Notkin, 1955-2013

2015 ACM RICHARD TAPIA CELEBRATION OF DIVERSITY IN COMPUTING CONFERENCE
BOSTON • FEB 18-21, 2015

Gender Distribution (%)

Male	46
Female	54

2015 ACM RICHARD TAPIA CELEBRATION OF DIVERSITY IN COMPUTING CONFERENCE
BOSTON • FEB 18-21, 2015

Race/Ethnicity (%)

Black or African American	34
Hispanic or Latino	22
White or Caucasian	19
Asian or Southeast Asian	10
Multiracial	3
Other	3

Impact on Research Communities Beyond SE

Related to funding/head-count allocation, student recruitment, ...
→ community growth

Lori Clarke

Symbolic execution (1976)
also by James King, William
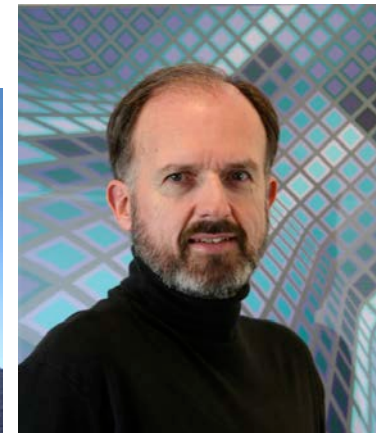Howden, Karl Levitt, et al.

Andreas Zeller

Delta debugging (1999)

Roy Fielding   Richard Taylor

Representational State Transfer
(**REST**) as a key architectural
principle of WWW (2000)

...

Practice Impact

- Diverse/balanced research styles shall/can be embraced
  - Our community already well appreciates **impact on other researchers**, e.g., SIGSOFT Impact Awards, ICSE MIP, paper citations
  - But often insufficient effort for **last mileage** or focus on **real problems**

- Strong need of ecosystem to incentivize practice impact pursued by researchers
  - Top down:

  Incentivizing Quality and Impact:
  Evaluating Scholarship in Hiring, Tenure, and Promotion
  Batya Friedman and Fred B. Schneider

  - Bottom up:
    - Conference PC for reviewing papers
      - Impact counterpart of "highly novel ideas"?
      - Impact counterpart of "artifact evaluation"?
    - Promote and recognize practice impact
      - Counterpart of ACM Software System Award?
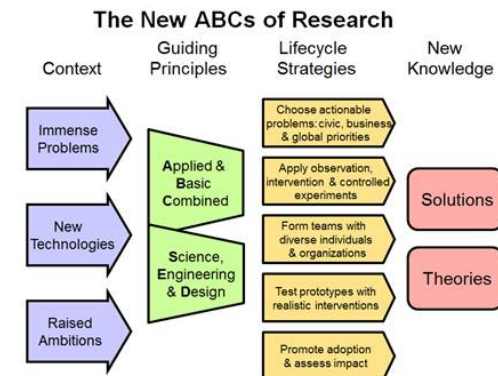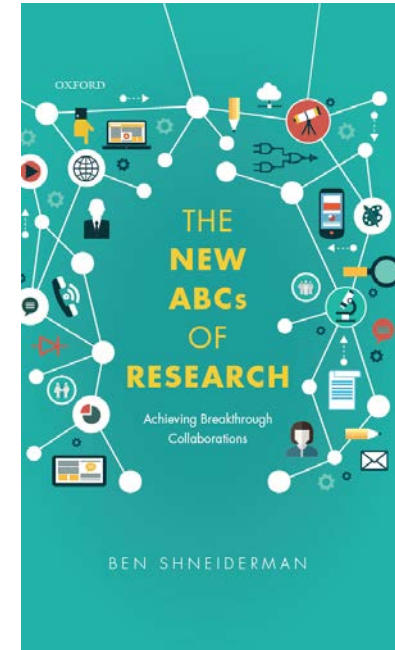
Practice-Impact Levels of Research

- Study/involve industrial data/subjects
  - Indeed, insights sometimes may benefit practitioners

- Hit (with a tool) and run
  - Authors hit and run (upon industrial data/subjects)
  - Practitioners hit and run

- **Continuous adoption by practitioners**
  - Importance of benefited domain/system (which can be just a single one)
    - Ex. WeChat test generation tool → WeChat with **> 900 million** users
    - Ex. MSRA SA on SAS → MS Online Service with **hundreds of million** users
    - Ex. Beihang U. on CarStream → Shenzhou Rental with > **30,000** vehicles over 60 cities
  - Scale of practitioner users
    - Ex. MSR Pex → Visual Studio 2015+ IntelliTest
    - Ex. MSR Code Hunt with close to **6 million** registered/anonymous/API accounts
    - Ex. MSRA SA XIAO → Visual Studio 2012+ Clone Analysis

Think about **>90%** startups fail! It is challenging to start from research and then single-handedly bring it to continuous adoption by target users; Academia-industry collaborations are often desirable.

Practice-Impact Levels of Research

- If there are practice impacts but no underlying research (e.g., published research), then there is no practice-impactful research
  - More like a startup's or a big company's product with business secrets

- Some industry-academia collaborations treat university researchers (students) like cheap(er) engineering labor → no or little research

Desirable Problems for Academia-Industry Collaborations

- Not all industrial problems are worth effort investment from university groups
  - High business/industry value
  - Allow research publications (not business secret) to advance the knowledge
  - Challenging problem (does it need highly intellectual university researchers?)
  - Desirably real man-power investment from both sides

- My recent examples
  - Tencent WeChat [FSE'16 Industry], [ICSE'17 SEIP]: Android app testing/analysis
  - Exploring collaborations with Baidu, Alibaba, Huawei
  - Exploring new collaborations with MSRA SA

Sustained Productive Academia-Industry Collaborations

- Careful selection of target problems/projects
  - Desirable to start with money-free collaborations(?)
  - If curiosity-driven nature is also from industry (lab) side, watch out.
- Each collaboration party needs to bring in something important and unique – win-win situation
  - High demand of abstraction/generalization skills on the academic collaborators to pursue research upon high-practice-impact work.
- Think more about the interest/benefit of the collaborating party
  - (Long-term) relationship/trust building
- Mutual understanding of expected contributions to the collaborations
  - Balancing research and "engineering"
  - Focus, commitment, deliverables, funding, …

Academia-Industry Collaboration

- **Academia** driven

- **Industry** driven

Academia-Driven: Research Dissemination

- Publishing research results → technologies there adopted by companies, e.g.,

ICSE 00 Daikon paper by Ernst et al. → Agitar Agitator
https://homes.cs.washington.edu/~mernst/pubs/invariants-relevance-icse2000.pdf
ASE 04 Rostra paper by Xie et al. → Parasoft Jtest improvement
http://taoxie.cs.illinois.edu/publications/ase04.pdf
PLDI/FSE 05 DART/CUTE papers by Sen et al. → MSR SAGE, Pex
http://srl.cs.berkeley.edu/~ksen/papers/dart.pdf
http://srl.cs.berkeley.edu/~ksen/papers/C159-sen.pdf

Academia-Driven: Research Commercialization

- Commercializing research results in startup →
  tools/products used by companies, e.g.,

Having a startup !→ leading to huge adoption

Academia-Driven: Tool Community Building

- Release open source infrastructures or libraries to engage academic/industry communities to **use and contribute**, e.g.,
  - MPI/PETSc by Bill Gropp et al.
  - Charm++ by Laxmikant (Sanjay) Kale et al.

  - LLVM by Vikram Adve, Chris Lattner, et al.

"The openness of the LLVM technology and the quality of its architecture and engineering design are key factors in understanding the success it has had both in academia and industry."

Industry-Driven: Infrastructure

- Making infrastructure available for academia to build upon, e.g.,
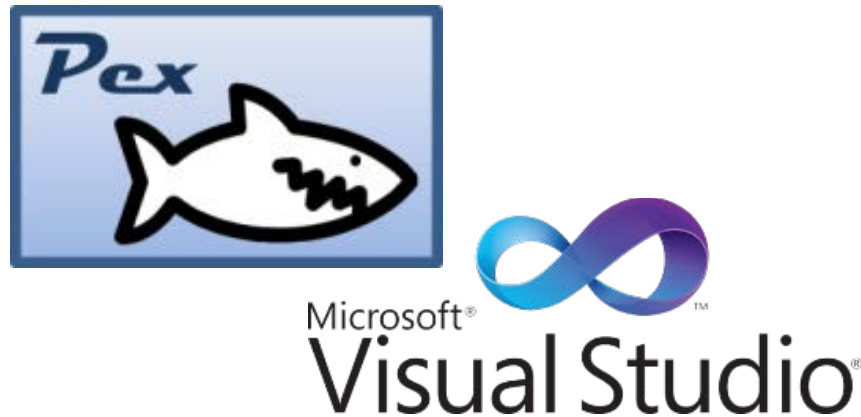
http://research.microsoft.com/pex/

Nikolai Tillmann, Jonathan de Halleux, and Tao Xie. Transferring an Automated Test Generation Tool to Practice: From Pex to Fakes and Code Digger. In Proceedings of the 29th IEEE/ACM International Conference on Automated Software Engineering (ASE 2014), Experience Papers
http://taoxie.cs.illinois.edu/publications/ase14-pexexperiences.pdf

Industry-Driven: Data

- Making data available
  - inside the company (visiting professors, student interns)

- to academia

Kika Emoji Keyboard

http://research.microsoft.com/codehunt/
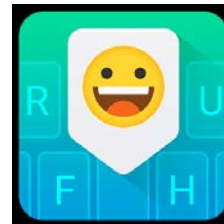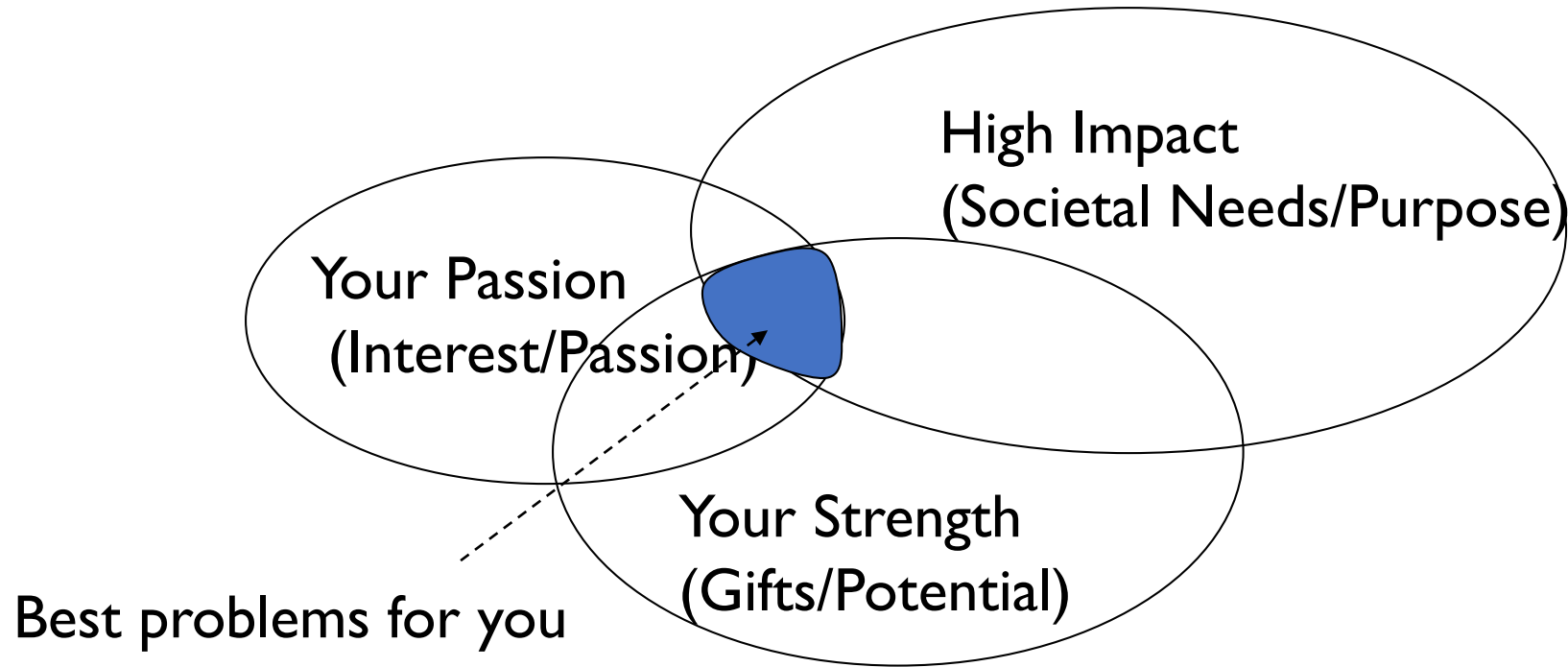
Industry Academia "Tensions"

- Academia (research recognitions, e.g., papers) vs. Industry (company revenues)
- Academia (research innovations) vs. Industry (likely involving engineering efforts)
- Academia (long-term/fundamental research or out of box thinking) vs. Industry (short-term research or work)
- Industry: problems, infrastructures, data, evaluation testbeds, …
- Academia: educating students, …

Summary: (How) Can A University Group Do It?

- Start a startup
  - but desirable to have right people (e.g., former students) to start
- Release free tools/libraries to aim for adoption
  - but a lot of efforts to be invested on "non-researchy" stuffs
- Collaborate with industrial research labs
  - but many research lab projects may look like univ. projects
- Collaborate with industrial product groups
  - but many problems faced by product groups may not be "researchy"

Optimizing "Research Return":
Pick a Problem Best for You

High Impact
(Societal Needs/Purpose)

Your Passion
(Interest/Passion)

Your Strength
(Gifts/Potential)

Best problems for you

Find your passion: If you don't have to work/study for money, what would you do?

Test of impact: If you are given $1M to fund a research project, what would you fund?

Find your strength/Avoid your weakness: What are you (not) good at?

Find what interests you that you can do well, and is needed by the people

Experience Reports on Successful Tool Transfer

- Yingnong Dang, Dongmei Zhang, Song Ge, Ray Huang, Chengyun Chu, and Tao Xie. **Transferring Code-Clone Detection and Analysis to Practice**. In *Proceedings of ICSE 2017*, SEIP.
  http://taoxie.cs.illinois.edu/publications/icse17seip-xiao.pdf

- Nikolai Tillmann, Jonathan de Halleux, and Tao Xie. **Transferring an Automated Test Generation Tool to Practice: From Pex to Fakes and Code Digger**. In *Proceedings of ASE 2014*, Experience Papers.
  http://taoxie.cs.illinois.edu/publications/ase14-pexexperiences.pdf

- Jian-Guang Lou, Qingwei Lin, Rui Ding, Qiang Fu, Dongmei Zhang, and Tao Xie. **Software Analytics for Incident Management of Online Services: An Experience Report**. In *Proceedings ASE 2013*, Experience Paper.
  http://taoxie.cs.illinois.edu/publications/ase13-sas.pdf

- Dongmei Zhang, Shi Han, Yingnong Dang, Jian-Guang Lou, Haidong Zhang, and Tao Xie. **Software Analytics in Practice**. IEEE Software, Special Issue on the Many Faces of Software Analytics, 2013.
  http://taoxie.cs.illinois.edu/publications/ieeesoft13-softanalytics.pdf

- Yingnong Dang, Dongmei Zhang, Song Ge, Chengyun Chu, Yingjun Qiu, and Tao Xie. **XIAO: Tuning Code Clones at Hands of Engineers in Practice**. In *Proceedings of ACSAC 2012*.
  http://taoxie.cs.illinois.edu/publications/acsac12-xiao.pdf

Microsoft® **Research**

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Questions & Discussion?