Evolution of Corporate CS Research

Attila Takacs, Mukund Sundararajan, Karnik Shah

This work was created in an open classroom environment as part of a program within the Sutardja Center for Entrepreneurship & Technology and led by Prof. Iklaq Sidhu at UC Berkeley. There should be no proprietary information contained in this work. No information contained in this work is intended to affect or influence public relations with any firm affiliated with any of the authors. The views represented are those of the authors alone.
The Seeds of Innovation and Entrepreneurship

Innovation is often symbolized by the image of a few friends toiling in a garage at night to create something new and disruptive. Many of today's industry giants were built from modest beginnings – literally in someone's garage.
Corporate research is dead, long live *Le Garage*!
Study change in the nature of corporate CS research

We identify and compare two eras
- Golden Age - pre ~1980
  – Characterized by large centralized research labs
- New Age - post 2000
  – Characterized by integrated R&D
- Note transition period in between

Where is this heading?
We Apply Three Frameworks

• **PEST & Horizons Framework**
  – Studies changing context of research
  – Political, Economic, Social and Technological

• **Pasteur Quadrant Analysis**
  – Studies theory v/s practice
PEST Framework

**Political**
- company leadership

**Economic**
- financial strength of the company

**Social**
- attitude towards research of the society at large

**Technological**
- technology maturity
<table>
<thead>
<tr>
<th></th>
<th>Golden Age</th>
<th>New Age</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P</strong></td>
<td>technologists at the top</td>
<td>financial background, shareholders, cost, risk</td>
</tr>
<tr>
<td><strong>E</strong></td>
<td>monopolistic position, surplus</td>
<td>fierce competition, cost control</td>
</tr>
<tr>
<td><strong>S</strong></td>
<td>high esteem of basic research</td>
<td>startup, maker movement</td>
</tr>
<tr>
<td><strong>T</strong></td>
<td>new industries, research directly drove societal and financial benefit</td>
<td>maturating technologies in the ICT domain</td>
</tr>
<tr>
<td><strong>H</strong></td>
<td>Horizon 1-2-3</td>
<td>1-2</td>
</tr>
</tbody>
</table>
Pasteur Quadrant

- Bohr: Foundational Yes, Use-oriented No
- Pasteur: Foundational Yes, Use-oriented Yes
- Edison: Foundational No, Use-oriented Yes

The Pasteur Quadrant is characterized by longer term, higher risk, and cost. The Bohr Quadrant is associated with shorter term, lower risk, and cost.
Mapping Papers into Quadrants

• Use DBLP + ACM data-set
  – year of publication, citations, conference/journal of publication

• Basic Idea:
  – Curated list of theory conferences and journals
  – Paper category = Theory, if published in one of these
Theory Papers, Practice Papers

Take Away #1:
Practice papers growing exponentially.
Theory papers growing (slow) linearly.
Citation Ratio

“Pasteurs” influenced by both theory and practice papers

- influence = citation
- citation ratio = #citation of theory papers/#citations of practice

**Take Away #2:**
From 1995 onwards, this ratio is closer to 1:1.

Theory is more disconnected from practice today
Stratified Sampling

J. ACM is the premier CS theory journal
- Sample articles from 1964(4), 1984(4), 2014(6)
- Score manually for fundamentalness, intent of use

![Graph showing Intent of Use over Year](image1)

![Graph showing Fundamental over Year](image2)
Theory Papers

- **Bohr**
  - Foundational: Yes
  - Use-oriented: No
  - Years: 1960’s

- **Pasteur**
  - Foundational: Yes
  - Use-oriented: Yes
  - Years: 1980’s

- **Edison**
  - Foundational: No
  - Use-oriented: Yes
  - Years: 2000’s

- **?**
  - Foundational: No
  - Use-oriented: No

The diagram illustrates the evolution of ideas from foundational to use-oriented over different decades.
Practice Papers

<table>
<thead>
<tr>
<th>Foundational</th>
<th>Use-oriented</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Bohr</td>
</tr>
<tr>
<td>No</td>
<td>?</td>
</tr>
</tbody>
</table>
| No           | Edison       | Yes
Who is a Pasteur?

Excellent Technical Skills + Excellent understanding of the world’s problems
Amplifying Technical Skills

• For many problems in NLP, data supplants theoretical models
  – The unreasonable effectiveness of Data
• Deep Learning extends this to image processing, speech
  • e.g. Identifying dog breeds automatically
  • e.g. Identifying street numbers
• Experience this yourself: Torch tutorial
  – 60 mins
  – recognize images in objects!!
Amplify Problem Understanding

• **Analyze data**
• Get quick feedback
  – **Validating Learning** (Lean Startup)
  – **Experiment frameworks**
    • large-scale traffic experiments on FB, Google, Bing
• Focus on feedback, not process!
Discussion

Image credit:
- Slide 2: http://www.telcomhistory.org/vm/heroesBell.shtml; Bell's workshop (AT&T file photo)
- Slide 3: https://lebbeuswoods.wordpress.com/2012/02/27/saarinens-last-experiment/