Google's Hybrid Approach to Research

Alfred Spector
Google Inc.
azs@google.com

Peter Norvig
Google Inc.
pnorvig@google.com

Slav Petrov
Google Inc.
slav@google.com

1 Introduction

In this paper, we describe how we organize Computer Science (CS) research at Google. We focus on how we integrate research and development (R&D) and discuss the benefits and risks of our approach.

The challenge in organizing R&D is great because CS is an increasingly broad and diverse field. It combines aspects of mathematical reasoning, engineering methodology, and the empirical approaches of the scientific method. The empirical components are clearly on the upswing, in part because the computer systems we construct have become so large that analytic techniques cannot properly describe their properties, because the systems now dynamically adjust to the hard-to-predict needs of a diverse user community, and because the systems can learn from vast data sets and large numbers of interactive sessions that provide continuous feedback.

We have also noted that CS is an expanding sphere, where the core of the field (Theory, Operating Systems, etc.) continues to grow in depth, while the field keeps expanding into neighboring application areas. Research results come not only from universities, but also from companies, large and small. The way that research results are disseminated is also evolving and the peer-reviewed paper is under threat as the dominant dissemination method. Open source releases, standards specifications, data releases, and novel commercial systems that set new standards upon which others then build, are increasingly important.

To compare our approach to research with that of other companies is beyond the scope of this paper. But, for reference, we note that in the terminology of Pasteur's Quadrant [1], we do "use-inspired basic" and "pure applied" (CS) research. [2] and [3] discuss information technology research generally, pointing out the movement in industrial labs towards research that strongly considers product needs. Recent articles, such as [4] and [5], illustrate related issues on how firms do research and catalyze innovation.

2 Research in Computer Science at Google

The goal of research at Google is to bring significant, practical benefits to our users, and to do so rapidly, within a few years at most. Research happens throughout Google, exploring technical innovations whose implementation is risky, and may well fail. Sometimes, research at Google operates in entirely new spaces, but most frequently, the goals are major advances in areas where the bar is already high, but there is still potential for new methods. In these cases, simply establishing the feasibility of a research idea may be a substantial task, but even greater effort is required to create a true success or useful negative result.

Because of the time-frame and effort involved, Google's approach to research is iterative and usually involves writing production, or near-production, code from day one. Elaborate research prototypes are rarely created, since their development delays the launch of improved end-user services. Typically, a single team iteratively explores fundamental research ideas, develops and maintains the software, and helps operate the resulting Google services – all driven by real-world experience and concrete data. This long-term engagement serves to eliminate most risk to technology transfer from research to engineering. This approach also helps ensure that research efforts produce results that benefit Google's users, by allowing research ideas and implementations to be honed on empirical data and real-world constraints, and by utilizing even failed efforts to gather valuable data and statistics for further attempts.

2.1 Implications of Google's Mission and Capabilities

Google's mission "To organize the world's information and make it universally accessible and useful," both supports and requires innovation in almost all CS disciplines. For example, we aim to "understand" user intent and the meaning of documents, to translate between languages with ever higher fidelity, and to be able to transform content in one modality (say, image) into relevant content in all others (say, text). Google's entire organization is focused on rapid innovation, and three aspects of Google's technology and business model support this:

Organizing all of the worlds information requires large amounts of resources. By providing a rich set of computing abstractions and powerful processors, storage, and networking capabilities in our data centers, Google has been able to gain economies of scale and to sidestep some of the complexity of heterogeneous computing environments.

The services-based delivery model brings significant benefits to research and development. Even a small team has at its disposal the power of many internal services, allowing the team to quickly create complex and powerful products and services. Design, testing, production and maintenance processes are simplified. Additionally, the services model, particularly one where there is significant consumer engagement, facilitates empirical research.

Finally, Google has been able to hire a talented team across the entire engineering operation. This gives us the opportunity to innovate everywhere, and for people to move between projects, whether they be primarily research or primarily engineering.

2.2 Hybrid Research at Google

Google's focus on innovation, its services model, its large user community, its talented team, and the evolutionary nature of CS research has led Google to a "Hybrid Research Model." In this model, we blur the line between research and engineering activities and encourage teams to pursue the right balance of each, knowing that this balance varies greatly. We also maintain considerable fluidity in terms of moving both people and projects as needs change. As such, even in areas where there is a much higher proportion of research to engineering, the "Research Team" we have established is not as formally separate from engineering activities as those in other organizations, and for example runs large production systems, too. Overall, we undertake research work when we feel its substantially higher risk is warranted by a chance of more significant potential impact. Additionally, research also has the potential to impact the world through Google's products and services, and through the academic research community. We recognize that the wide dissemination of fundamental results often benefits us by garnering valuable feedback, educating future hires, providing collaborations, and seeding additional work.

In no way do we feel that our model precludes long term research: we just try hard to "factorize" it into shorter-term, measurable components. This provides benefits to us in terms of team motivation (based upon evidence of concrete progress in reasonable time periods) and the potential for commercial benefit (in advance of the complete fulfillment of all objectives). Even if we cannot fully factorize work, we have sometimes undertaken longer term efforts. For example, we have started multi-year, large systems efforts (e.g., Google Translate, Chrome, Google Health) that have important research components. These projects were characterized by the need for complex systems and research (e.g., web-scale identification of parallel corpora for Translate [6] and various complex security features in Chrome [7] and Health). At the same time, we have recently shown that even in longer term, publicly launched efforts, we are unafraid to refocus our work (e.g., Health), if it seems we are not achieving success.

Clearly, this approach benefits from the mainly evolutionary nature of CS research, where great results are usually the composition of many discrete steps. If the discrete steps required large leaps in vastly different directions, we admit that our primarily hill-climbing-based approach might fail. Thus, we have structured the Google environment as one where new ideas can be rapidly verified by small teams through large-scale experiments on real data, rather than just debated. The small-team approach benefits from the services model, which enables a few engineers to create new systems and put them in front of users. This in turn enables us to conduct experiments at a scale that is generally unprecedented for research and development projects. One consequence is that many projects can directly affect billions of users. This naturally influences how researchers choose to spend their time, balancing the opportunity to have impact through Google's services, with the opportunity to have impact in the academic community. Google encourages both kinds of impact, and some of the most successful projects achieve both.

We thus define our hybrid research model as one that (i) aims to generate scientific and engineering advances in fields of import to Google, that (ii) does so in a way that tends to factorize longer projects (perhaps with very challenging goals) into discrete, achievable steps (each of which may be of commercial value), where (iii) we maximally leverage our cloud computing models and large user base to support in vivo research, where (iv) we allow for the maximal amount of organizational flexibility so that we can support both projects that require some room to grow unfettered by current constraints, as well as projects that require close integration with existing products, and where (v) we emphasize knowledge dissemination using a flexible collection of different approaches.

2.3 Example Research Patterns

1. An advanced project in a product-focused team that, by virtue of its creativity and newness, changes the state of the art and thereby produces new research results.
The first and most prevalent pattern exemplifies how blurry the line between research and development work can be. Operating at large scale, engineering teams are often faced with novel challenges which, when overcome, constitute research results. Organizationally, research is done in situ by the product team to achieve its goals. The most successful high-profile examples of this pattern are systems infrastructure projects such as MapReduce [8], Google File System [9] and BigTable [10].

2. A project in the research group that results in new products or services.
The second pattern is research followed by the operation of the production service based on that research. Both Google Translate and Voice Search [11] are examples of this pattern, where the cloud computing infrastructure enabled small research teams to build systems that could be deployed. This pattern applies best when continuing research can further improve and extend the resulting products.

3. A project in the research group that creates new concepts and technologies, which are then applied to existing products or services.
The third pattern is a traditional research and development model. Google's success with this model of research benefits from the services model and from the emphasis on data-driven evaluation. For instance, some new audio and video fingerprinting techniques [12], which researchers were able to demonstrate not only on small test cases, but on real data at production scale, were then productized by YouTube engineers.

4. A joint research project between an engineering team and the research group which is then used by that engineering team.
The fourth pattern is a collaborative integration of research and development teams. Many of our products require novel algorithmic solutions to support high performance, thus posing a blend of research and engineering challenges. An example for this pattern is the work done by our Market Algorithms group in collaboration with teams working on our advertisement systems. Together, they design, modify and analyze the core algorithms and economic mechanisms used for ad selection and optimization.

5. A research project in an engineering team that is transitioned to the research group (and eventually becomes (2), (3) or (4) above).
The fifth pattern, transitioning a project from an engineering team to the research team is an important mechanism for giving a project more time or resources, when the work is important more broadly than for a specific engineering team. An example of this pattern is work on YouTube recommendations, which started in various engineering groups, but then moved to a research team, where the work continued using a different, and perhaps deeper, algorithmic basis.

2.4 Successes

In the same way that it is difficult to define what exactly constitutes "research," it can be difficult to measure its "success." In our opinion, a research project is successful if it has academic or commercial impact, or ideally, both. Commercial impact at Google is perhaps easier to measure, and the company has benefitted from numerous advances in systems, speech recognition, language translation, machine learning, market algorithms, computer vision, and more.

By academic impact we refer to impact on the academic community, other companies or industries, and the field of Computer Science in general. Of course, this type of impact has most traditionally come from publications,

and Google continues to publish research results at increasing rates (from 13 papers published in 2003, to 130 in 2006, to 279 in 2011). Some of our papers are highly regarded and have received extensive references [8, 9, 10]. But we feel that publications are by no means the only mechanism for knowledge dissemination: Googlers have led the creation of over 1000 open source projects, contributed to various standards (e.g. as editor of HTML5), and produced hundreds of public APIs for accessing our services. In some cases, we have used these different channels in symbiotic ways, following up an initial publication describing the high-level ideas (e.g. MapReduce, GFS, BigTable) with open source implementations of particular aspects (e.g. Protocol Buffers). In other cases, projects have started as open source initiatives from day one: Android and Chromium are probably the two most well-known examples of open source projects and demonstrate the effectiveness of this approach.

3   Discussion

Technology companies invest in research for a number of reasons, including: (i) importance to the companys products and services, (ii) prestige and contributions to the public good, and (iii) reducing the risk of getting blindsided by new technology developments.

Research at Google is built on the premise that connecting research with development provides teams with powerful, production-quality infrastructure and a large user base, resulting not only in innovative research, but also in valuable new commercial capabilities. By coupling research and development, our goal is to minimize or even eliminate the traditional technology transfer process, which has proven challenging at other companies. Most of our projects involve people working with a given technology from the research stage through to the product stage. This close collaboration and integration furthermore ensures the reality of the problems being investigated: research is conducted on real systems and with real users. Our flexible organization also provides diverse opportunities for our employees and has positive implications on our innovation culture and hiring ability.

Of course, this close integration also brings some risks with it. Being so close to the users and to the day-to-day activities of product teams, it is easy to get drawn in and miss new developments. To mitigate this risk, we engage with the academic community through various initiatives such as our visiting faculty program, our intern program or our faculty research awards program. We also encourage publication of research results, though we sometimes get criticized for not publishing enough. One reason for this is that researchers at Google have multiple avenues for having impact, publishing papers not being the only one. As a result, Googlers publish fewer papers, but the ones that they publish can be more impactful, because they describe experience with well-tested and implemented systems, not just proposed ideas. Another potential pitfall of the hybrid research model is that it is probably more conducive to incremental research. We therefore do support paradigmatic changes as well, as exemplified by our autonomous vehicles project, Google Chauffeur, among others.

4   Conclusions

Many of the world's Computer Science research questions are of great relevance to Google's business, our technical leaders, and our user community. We have chosen to organize Computer Science research differently at Google by maximally connecting research and development. This yields not only innovative research results and new technologies, but also valuable new capabilities for the company. Our hybrid approach to research enables us to conduct experiments at a scale that is generally unprecedented for research projects, resulting in stronger research results that can have a wider academic and commercial impact. We also provide flexible opportunities across the R&D spectrum for our team members. While our hybrid research model exploits a number of things that are particular to Google, we hypothesize that it may also serve as an interesting model for other technology companies.

Acknowledgments

References

[1] Donald E. Stokes. *Pasteur's Quadrant - Basic Science and Technological Innovation*. Brookings Institution Press, 1997.

[2] Robert Buderi. *Engines of Tomorrow: How The Worlds Best Companies Are Using Their Research Labs To Win The Future*. Simon & Schuster, 2000.

[3] Mark Dodgson, David Gann, and Ammon Salter. *The Management of Technological Innovation: Strategy and Practice*. Oxford University Press, 2008.

[4] Richard Leifer, Gina OConnor, and Mark Rice. Implementing radical innovation in mature firms: The role of hubs. *The Academy of Management Executive*, 15, 2001.

[5] Ellen Enkel, Oliver Gassmann, and Henry Chesbrough. Open r&d and open innovation: Exploring the phenomenon. *R&D Management*, 39, 2009.

[6] Jakob Uszkoreit, Jay Ponte, Ashok Popat, and Moshe Dubiner. Large scale parallel document mining for machine translation. In *Proc. of COLING*, 2010.

[7] Charles Reis, Adam Barth, and Carlos Pizano. Browser security: Lessons from google chrome. *ACM Queue*, 7, 2009.

[8] Jeffrey Dean and Sanjay Ghemawat. Mapreduce: Simplified data processing on large clusters. In *Proc. of OSDI*, 2004.

[9] Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung. Google file system. In *Proc. of ACM SIGOPS*, 2003.

[10] Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach, Mike Burrows, Tushar Chandra, Andrew Fikes, and Robert E. Gruber. Bigtable: A distributed storage system for structured data. In *Proc. of OSDI*, 2006.

[11] Johan Schalkwyk, Doug Beeferman, Francoise Beaufays, Bill Byrne, Ciprian Chelba, Mike Cohen, Maryam Garrett, and Brian Strope. "your word is my command": Google search by voice: A case study. In Amy Neustein, editor, *Advances in Speech Recognition*. Springer, 2010.

[12] Shumeet Baluja and Michele Covell. Waveprint: Efficient wavelet-based audio fingerprinting. *Pattern Recognition*, 11, 2008.

Additional references can be found at http://research.google.com/pubs/papers.html