

Influence of information technology companies in scientific publications

Bachelor thesis Information Science
Research proposal first draft

Joost Verkaik

April 12, 2017

1 Personal details

Author email joost.verkaik@student.uva.nl

Supervisor email maartenmarx@uva.nl

GitHub <https://github.com/joostverkaik/influence-tech-companies>

2 Introduction

Life without information technology is hard to think of. Humans find themselves increasingly intertwined with technology every day. This resonates into our daily lives, independent of ones age, education or profession. It is interesting to see how companies handle the increasing amount of technology in their surrounding. Companies like Google, Facebook, Twitter and others are impacting our lives more and more each day and these companies are conducting their own researches to invent new technologies. In this thesis I will attempt to show to what extent tech companies have influence in the scientific world, in particular on scientific publications.

3 Research question

Most studies that research influence of companies did not focus on scientific publications and because of the growing influence of tech companies on multiple aspects, the main research question of this thesis is: *What is the influence of large tech companies on subjects that are researched in information science publications?*

To answer that question, the main question is split in the following sub questions:

1. What are the trends in technology in the past decade from a common perspective?
2. What trends can be distinguished from scientific publications?

3. What are the affiliations of authors of scientific publications?
4. Is there a relation between research about certain technologies and companies that use those technologies in their products or services?

4 Related literature

The idea that will be researched in this thesis is quite new. Previous research mainly focuses on the influence of information technology itself on society, business, et cetera (Baloh & Trkman, 2003). Previous research has examined the role of public research on industrial R&D (Cohen, Nelson, & Walsh, 2002), stating that public research is very important for generating new ideas. Not a lot of publications can be found that turn the question about influence around to focus on the influence of a certain party or collection of parties on the area of information technology in scientific research. In 2006, an article was published in Research Policy about the influence of R&D and innovation strategies between companies and universities. From a business perspective, the alliances were examined and comparisons were made between companies and universities on the one hand and companies and other external partners on the other hand (Bercovitz & Feldman, 2007). While this article focused on the relation itself, this thesis will focus on the greater picture of influence of companies on research.

In an article written by Critchley & Nicol, support for research is assessed depending on the source of funding. The impact of commercialization is noticeable, with private funding resulting in lower support of the research (Critchley & Nicol, 2011). Research with funding from, or research conducted by, large technology companies could count on less support from the public.

5 Methodology

To find the answer to the main research question, the following methodology will be applied.

- Determine the best collection of scientific articles to scrape
- Collect technology trends of the past decade from annual trends reports from Deloitte
- Build a Python scraper to gather data (title, authors + affiliations, abstract, keywords/topics) about scientific articles from sources like ACM
- Clean up the data in Python so that it is usable
- Search the dataset for popular trends, e.g. no. of articles about tweets
- Analyze all affiliations of the authors and check for example how many occurrences of internal research departments of big tech companies can be found

6 Risk assessment

As this research will be conducted on a dataset that is yet to be retrieved and constructed, there could be some obstacles in obtaining the data. Scraping the data could take some time because the library that will be used does not offer a downloadable data file and sending a lot of web requests could lead to being blocked by the party that offers the data. However, with Python libraries such as Scrapy, certain measures can be taken to minimize the chance of "getting caught" (like waiting between requests).

If the method above does not work, a less recent, readily available dataset would probably suffice as well.

7 Project plan

Week	Task
14	Determine subject for thesis
15	Search literature, hand in research proposal
16	Search literature, start with theory section
17	Theory section
18	Finish theory section, explore scraping possibilities
19	Write scraping scripts for gathering data
20	Scrape data
21	Analyze data
22	Analyze data
23	Process results into paper
24	Conclusion, discussion
25	Finish thesis, polish unfinished parts
26	Wrap-up thesis

Table 1: Project plan

References

- Baloh, P., & Trkman, P. (2003). Influence of Internet and Information Technology on Work and Human Resource Management. In *In-site*. Retrieved from <http://www.proceedings.informingscience.org/IS2003Proceedings/docs/071Baloh.pdf>
- Bercovitz, J. E. L., & Feldman, M. P. (2007). Fishing upstream: Firm innovation strategy and university research alliances. *Research Policy*, 36, 930–948. doi: 10.1016/j.respol.2007.03.002
- Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2002). Links and Impacts: The Influence of Public Research on Industrial R&D. *Management Science*, 48(1), 1–23. doi: 10.1287/mnsc.48.1.1.14273
- Critchley, C. R., & Nicol, D. (2011). Understanding the impact of commercialization on public support for scientific research: is it about the funding source or the organization conducting the research. *Public understanding of science*, 20(3), 347–366. doi: 10.1177/0963662509346910