Query-based Text Summarization using Generative Adversarial Networks

On behalf of the de Melo Lab for the NSF

Abhinav Madahar

2018 April 25

Rutgers University

Outline

Introduction

Needs

Background

Organization Information

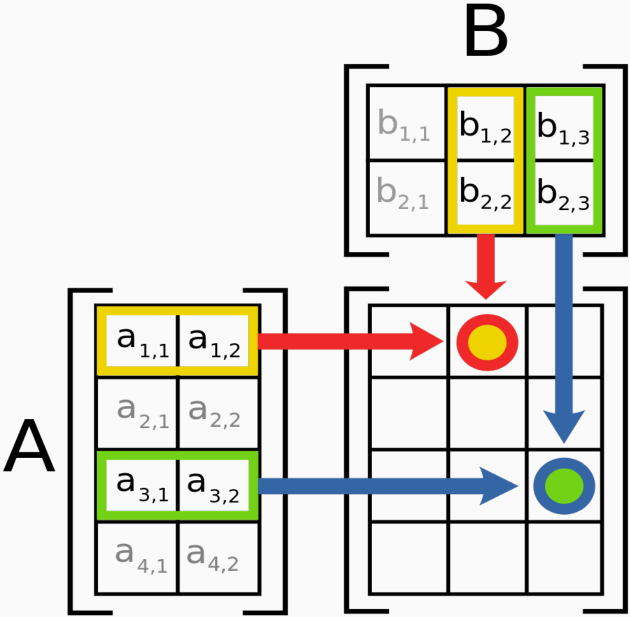
Our Project

Budget

Impact

Introduction

What are artificial intelligence and machine learning?



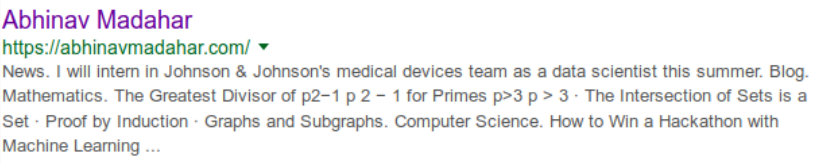
Motivation

> *Systematic reviews are a cornerstone of evidence-based care and a necessary foundation for care recommendations to be labeled clinical practice guidelines. However, they become* **outdated relatively quickly** *and require substantial resources to maintain relevance. One particularly time-consuming task is* **updating the search to identify relevant articles published since the last search.** *[. . .] Machine learning shows promise for* **decreasing the effort** *involved in updating searches for systematic reviews.*

(Shekelle, Shetty, Newberry, Maglione, and Motala, 2017)

Goals

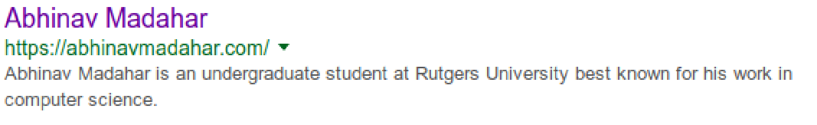
Objectives



Abhinav Madahar

https://abhinavmadahar.com/ ▾

News. I will intern in Johnson & Johnson's medical devices team as a data scientist this summer. Blog. Mathematics. The Greatest Divisor of p2−1 p 2 − 1 for Primes p>3 p > 3 · The Intersection of Sets is a Set · Proof by Induction · Graphs and Subgraphs. Computer Science. How to Win a Hackathon with Machine Learning ...



Abhinav Madahar

https://abhinavmadahar.com/ ▾

Abhinav Madahar is an undergraduate student at Rutgers University best known for his work in computer science.

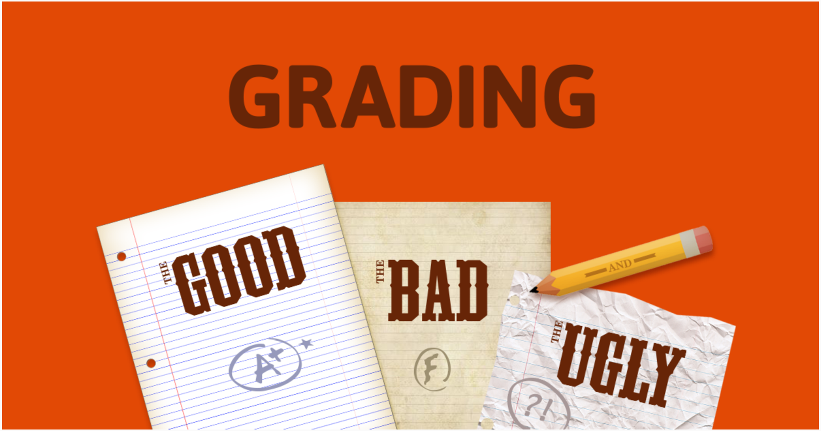
Intellectual Merit

- New techniques for summarizing a search result
- New techniques to train generative adversarial networks

Broader Impact Activities

- Educating students on modern machine learning via Rutgers Masters' of Computer Science and online courses.
- Publication of example implementation code to help software engineers apply our research.

Evaluation



Organization Information

清華大學
TSINGHUA UNIVERSITY
1911
自強不息
厚德載物

RUTGERS THE STATE UNIVERSITY OF NEW JERSEY
1766
SOL IUSTITIAE ET OCCIDENTEM ILLUSTRA

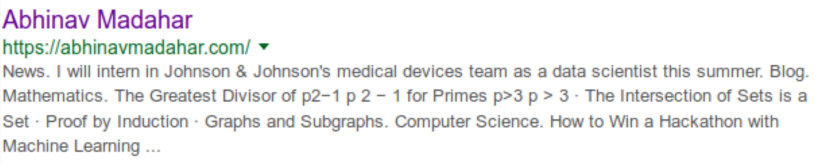
Needs

The Growth of the Internet

> *In todays era, when the size of information and data is increasing exponentially, there is an upcoming need to create a concise version of the information available.*

(Bhartiya and Singh, 2014),
(Mishra et al., 2014),
(Farzindar and Lapalme, 2004)

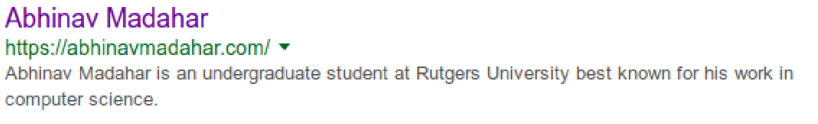
The Existing Solution to Finding Information Online



Abhinav Madahar

https://abhinavmadahar.com/ ▾

News. I will intern in Johnson & Johnson's medical devices team as a data scientist this summer. Blog. Mathematics. The Greatest Divisor of p2−1 p 2 − 1 for Primes p>3 p > 3 · The Intersection of Sets is a Set · Proof by Induction · Graphs and Subgraphs. Computer Science. How to Win a Hackathon with Machine Learning ...



Abhinav Madahar

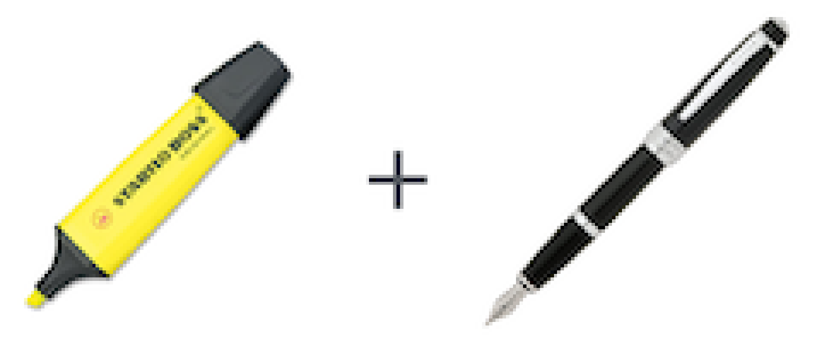
https://abhinavmadahar.com/ ▾

Abhinav Madahar is an undergraduate student at Rutgers University best known for his work in computer science.

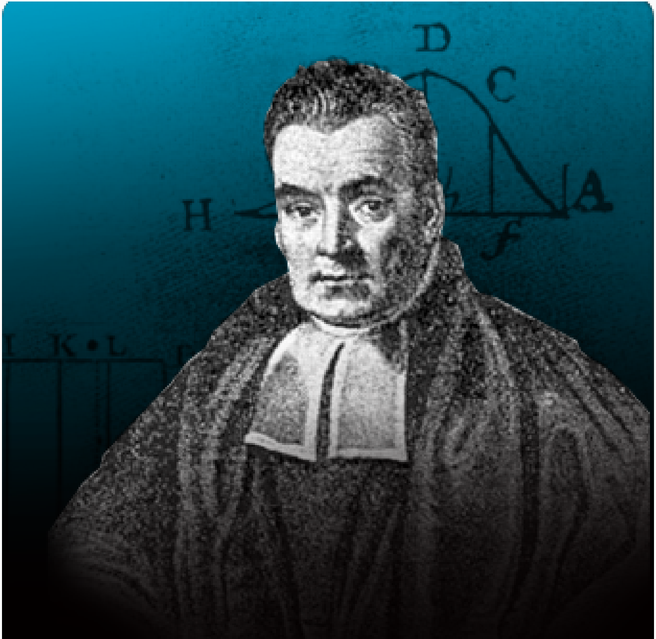
(Hasselqvist, Helmertz, and Kågebäck, 2017),
(Torres-Moreno and Torres-Moreno, 2014)

Background

Extractive vs Abstractive



Bayesian Model



Word Duplication

Predicted Summary: *the large to euthanasia is a natural death* **life life** *use*

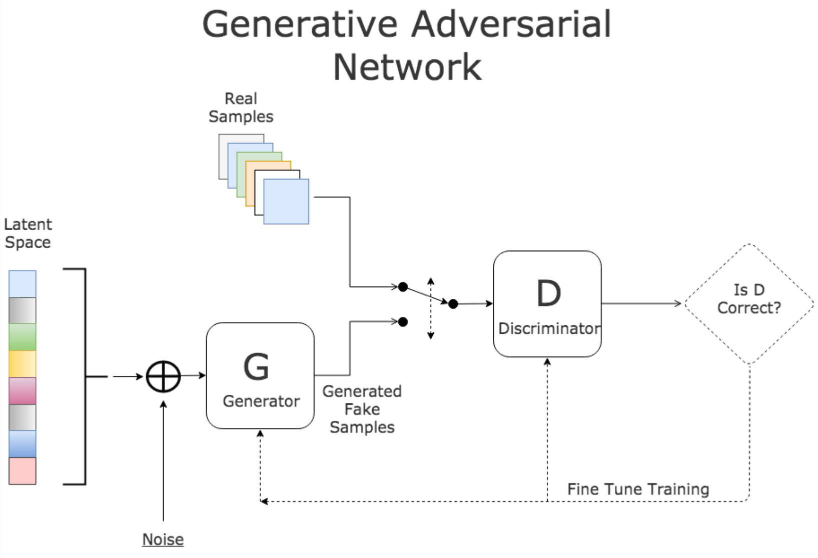
(Nema, Khapra, Laha, and Ravindran, 2017)

Question/Answer Model

How does this document relate to that query?

(Hasselqvist et al., 2017)

Generative Adversarial Network



Organization Information

Members

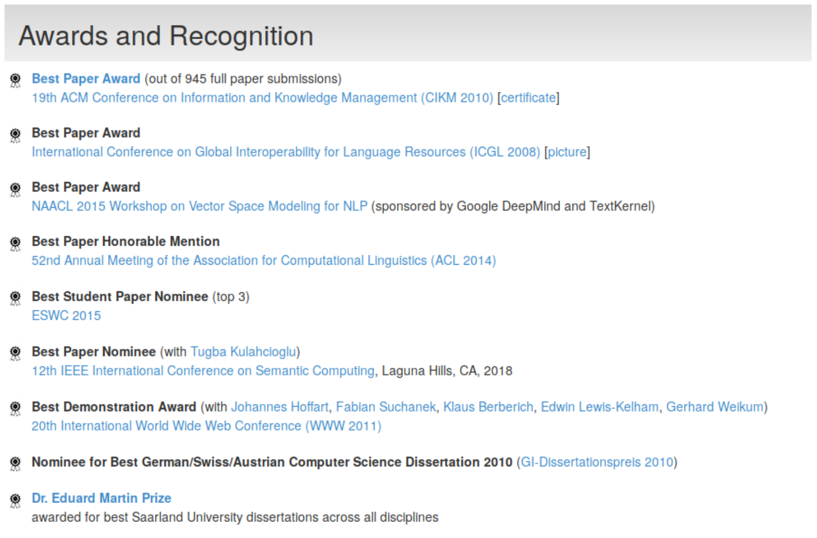
- 11 PhD students
- A few masters and undergraduate students

Location

清華大學
TSINGHUA UNIVERSITY
1911
自強不息
厚德載物

RUTGERS THE STATE UNIVERSITY OF NEW JERSEY
1766
SOL IUSTITIAE ET OCCIDENTEM ILLUSTRA

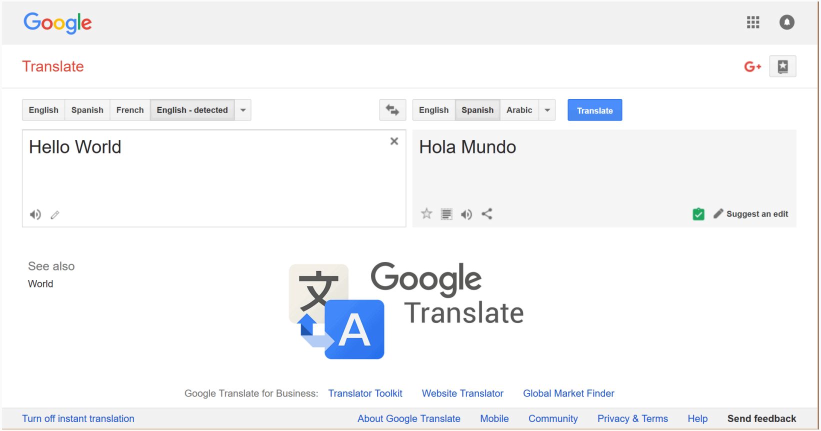
Awards



Awards and Recognition

- **Best Paper Award** (out of 945 full paper submissions)
  19th ACM Conference on Information and Knowledge Management (CIKM 2010) [certificate]
- **Best Paper Award**
  International Conference on Global Interoperability for Language Resources (ICGL 2008) [picture]
- **Best Paper Award**
  NAACL 2015 Workshop on Vector Space Modeling for NLP (sponsored by Google DeepMind and TextKernel)
- **Best Paper Honorable Mention**
  52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014)
- **Best Student Paper Nominee** (top 3)
  ESWC 2015
- **Best Paper Nominee** (with Tugba Kulahcioglu)
  12th IEEE International Conference on Semantic Computing, Laguna Hills, CA, 2018
- **Best Demonstration Award** (with Johannes Hoffart, Fabian Suchanek, Klaus Berberich, Edwin Lewis-Kelham, Gerhard Weikum)
  20th International World Wide Web Conference (WWW 2011)
- **Nominee for Best German/Swiss/Austrian Computer Science Dissertation 2010** (GI-Dissertationspreis 2010)
- **Dr. Eduard Martin Prize**
  awarded for best Saarland University dissertations across all disciplines

Previous Industry Work



Google

Translate

English | Spanish | French | English - detected

English | Spanish | Arabic

Translate

Hello World

Hola Mundo

Suggest an edit

See also

World

Google Translate

Google Translate for Business: Translator Toolkit Website Translator Global Market Finder

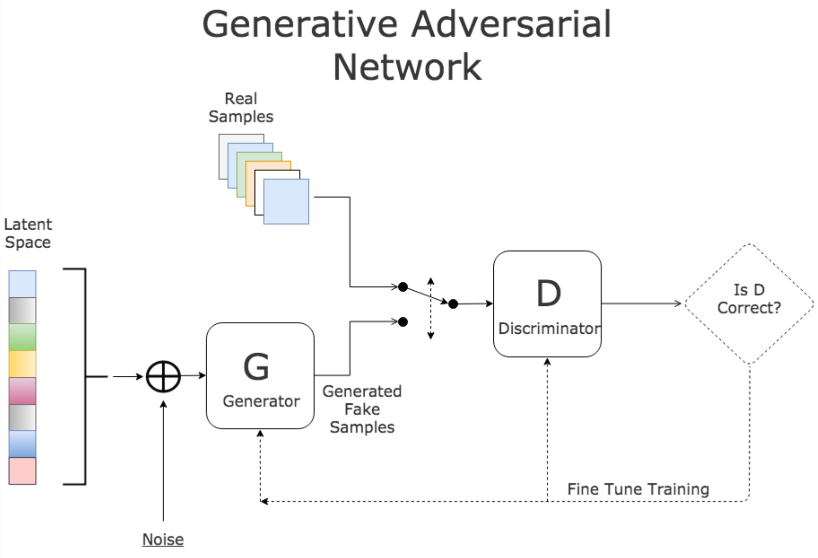
Turn off instant translation About Google Translate Mobile Community Privacy & Terms Help Send feedback

Previous Grants

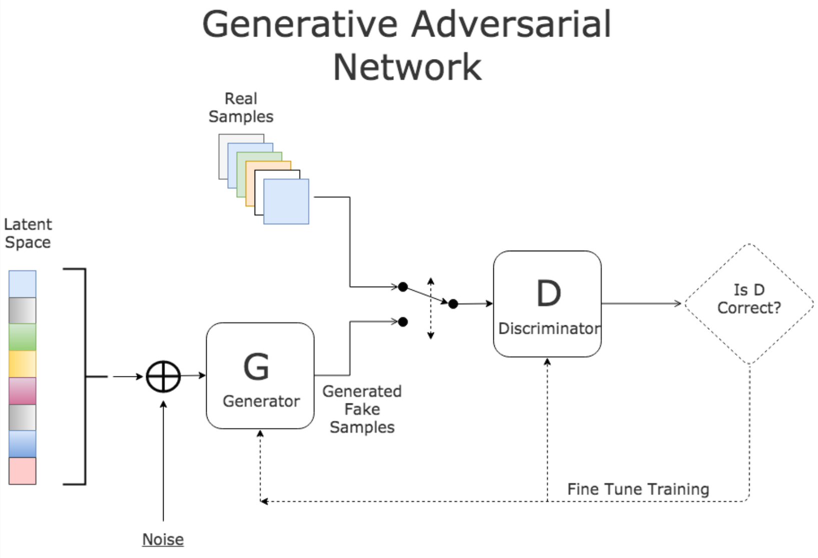
Our Project

1. Collect and Compile Data

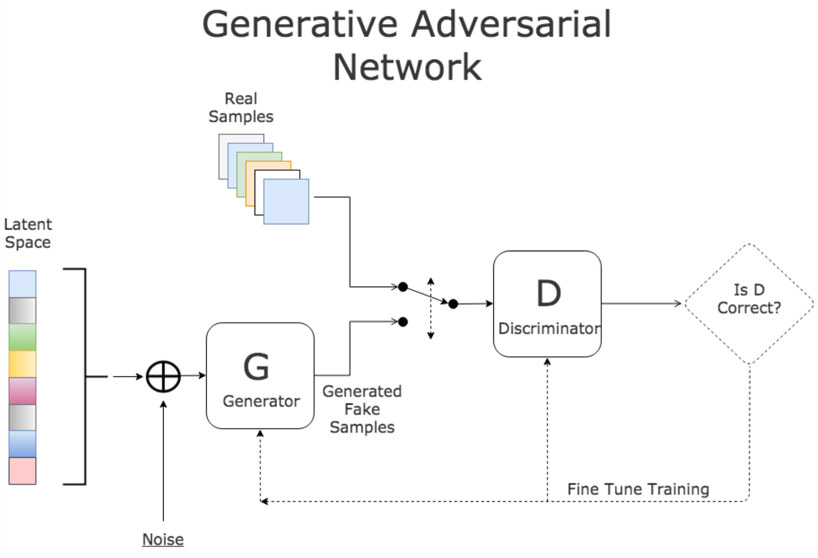
2. Develop a Generator Model



3. Develop a Discriminator Model



4. Go back to step 1 and Implement Recent Developments



Budget

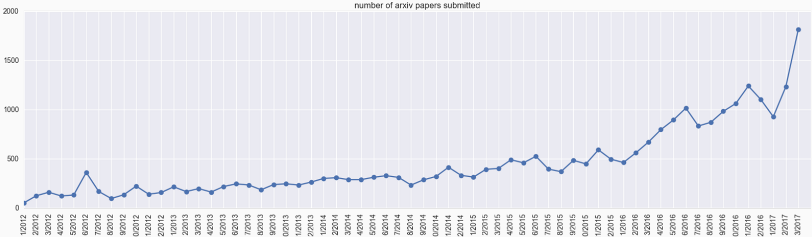
Overall cost

$90 000	Gerard de Melo’s Salary for 2 semesters (spread over 2 years)
$120 000	PhD students (2 students for 4 years)
$110 000	Postdoc (1 postdoc for 2 years)
$30 000	Conference publication (3 conferences)
$50 000	Laboratory tools (hardware, software, etc.)

Total: $400 000

Impact

Intellectual Merit



- New generative models
- New discriminative models
- Training GANs (e.g. loss, optimizers, etc.)
- More datasets

Broader Impact Activities

- Easier application into industry with sample code
- Rutgers courses
- Online course

References

Bhartiya, D. & Singh, A. (2014). A Semantic Approach to Summarization. *arXiv:1406.1203 [cs]*. arXiv: 1406.1203. Retrieved March 6, 2018, from http://arxiv.org/abs/1406.1203

Farzindar, A. & Lapalme, G. (2004). Letsum, an automatic legal text summarizing system. *Legal knowledge and information systems, JURIX*, 11–18.

Hasselqvist, J., Helmertz, N., & Kågebäck, M. (2017). Query-Based Abstractive Summarization Using Neural Networks. *arXiv:1712.06100 [cs]*. arXiv: 1712.06100. Retrieved January 30, 2018, from http://arxiv.org/abs/1712.06100

Mishra, R., Bian, J., Fiszman, M., Weir, C. R., Jonnalagadda, S., Mostafa, J., & Fiol, G. D. (2014). Text Summarization in the Biomedical Domain: A Systematic Review of Recent Research. *Journal of biomedical informatics*, *0*, 457–467. doi:10.1016/j.jbi.2014.06.009

Nema, P., Khapra, M., Laha, A., & Ravindran, B. (2017). Diversity driven Attention Model for Query-based Abstractive Summarization. *arXiv:1704.08300 [cs]*. arXiv: 1704.08300. Retrieved February 24, 2018, from http://arxiv.org/abs/1704.08300

Shekelle, P. G., Shetty, K., Newberry, S., Maglione, M., & Motala, A. (2017). Machine Learning Versus Standard Techniques for Updating Searches for Systematic Reviews: A Diagnostic Accuracy Study. *Annals of Internal Medicine*, *167*(3), 213. doi:10.7326/L17-0124

Torres-Moreno, J.-M. & Torres-Moreno, J.-M. (2014). Why Summarize Texts? In *Automatic Text Summarization* (pp. 1–21). John Wiley & Sons, Inc. doi:10.1002/9781119004752.ch1