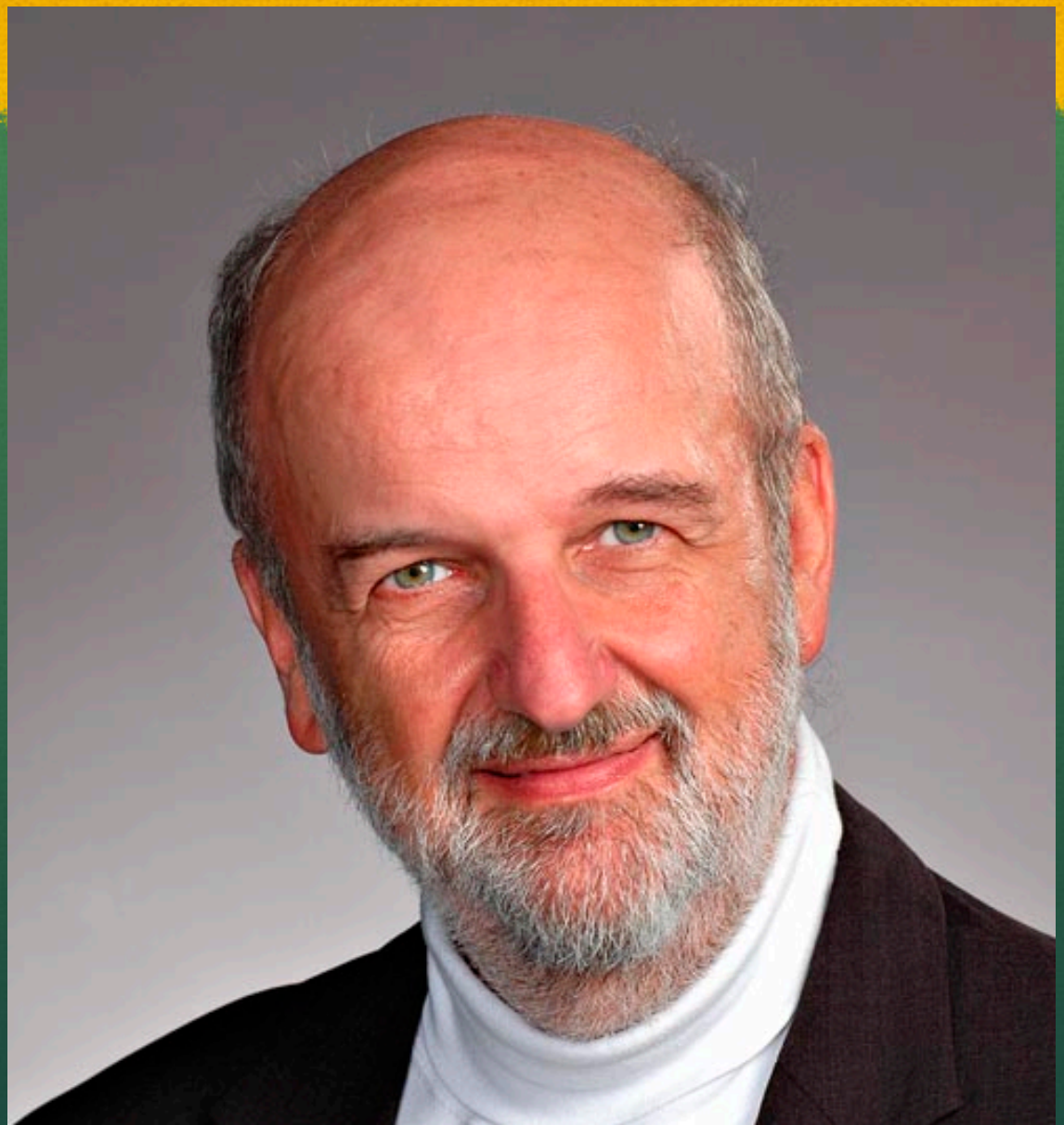
Disease Patterns in Text

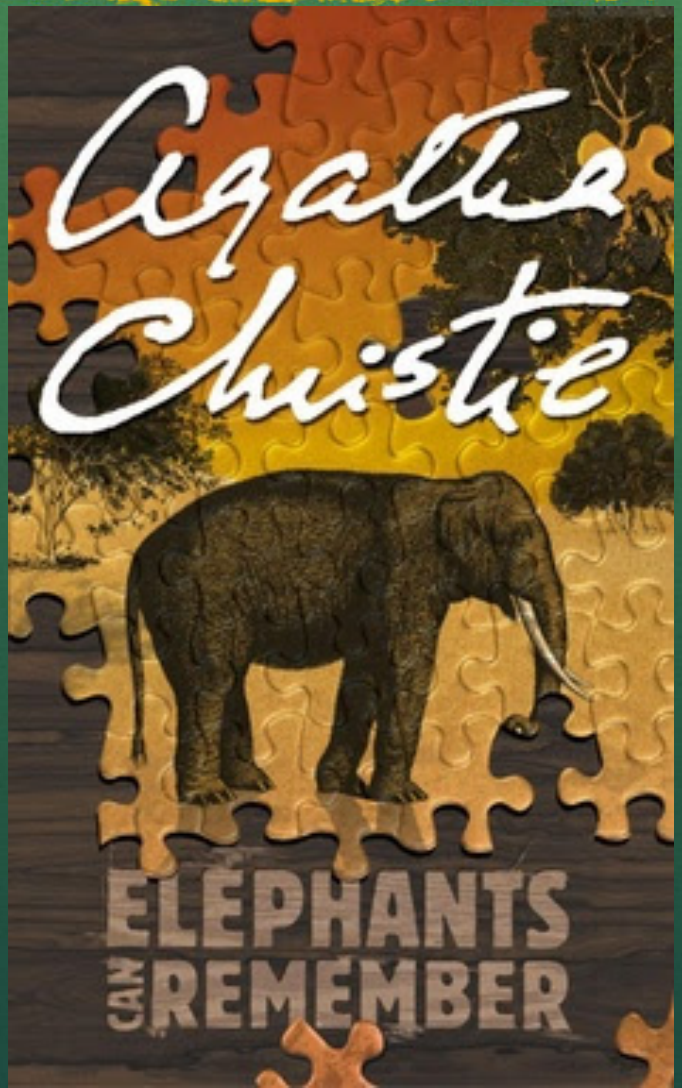
Using Alternative Biomarkers to make more accurate diagnoses

Dr. Ian Lancashire



The Question

Elephants Can Remember



Tools

- NLP: scikit-learn, gensim
- Exploration: IPython Notebook
- IMAP: Context.io
- Backend: Flask (Python microframework)
- Deployment: Heroku + PostgreSQL

Challenges

1. Not allowed to store data (only metadata)
2. Doctors must give permission (IRB)
3. Predicting Useful Features

Next Steps

- Stay Cognizant of Bonferroni's Principle
- http://nbviewer.ipython.org/github/gwulfs/Email-Analysis/blob/master/mom_vs_me.ipynb
- https://plot.ly/~gwulf/32

Bibliography

- ftp://ftp.cs.toronto.edu/pub/gh/Lancashire+Hirst-extabs-2009.pdf
- http://www.radiolab.org/story/91960-vanishing-words/
- ftp://ftp.cs.toronto.edu/dist/gh/Le-MSc-2010.pdf