

# The Right to Read is the Right to Mine - Text and Data Mining for Libraries and Your Users

Benjamin White

Head of Intellectual Property, British Library.

The image features a silhouette of an oil pumpjack against a gradient background of purple and blue. The pumpjack is positioned on the right side of the frame, with its long arm extending upwards and to the left. The background is a smooth gradient, transitioning from a darker purple at the top to a lighter blue at the bottom. The overall mood is industrial and futuristic.

DATA

is the new oil





Pre mid 1990s = pen, pencil and eyes

▪

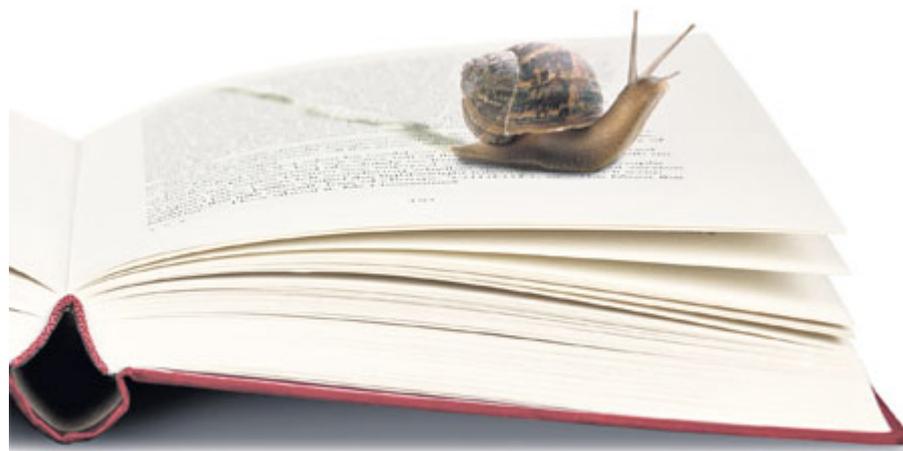
# Computers can now read



© Woodguy

[www.bl.uk](http://www.bl.uk)

And a lot faster than humans



# How to Do Research in 2013?

Post mid 1990s = pen, pencil, eyes **AND** computers.

Are off the shelf text and data mining tools from software providers, but researchers write their own programmes too.

# Text and Data Mining

- It is the reading of text and data to derive hypotheses / trends, links and extract facts.
- Also helps you structure and organise data and text.
- It has happened for years (unregulated by copyright law) using a pen and pencil.

bing

YAHOO!

KVASIC

ЯНДЕКС

ENIQA

Rambler

ilse

SEDUAM

volla.fr

virgilio

Bai du 百度

樂天

NAVER

Google

Google

YAHOO!

guruji

bing

# Many Tech Companies Text and Data Mine

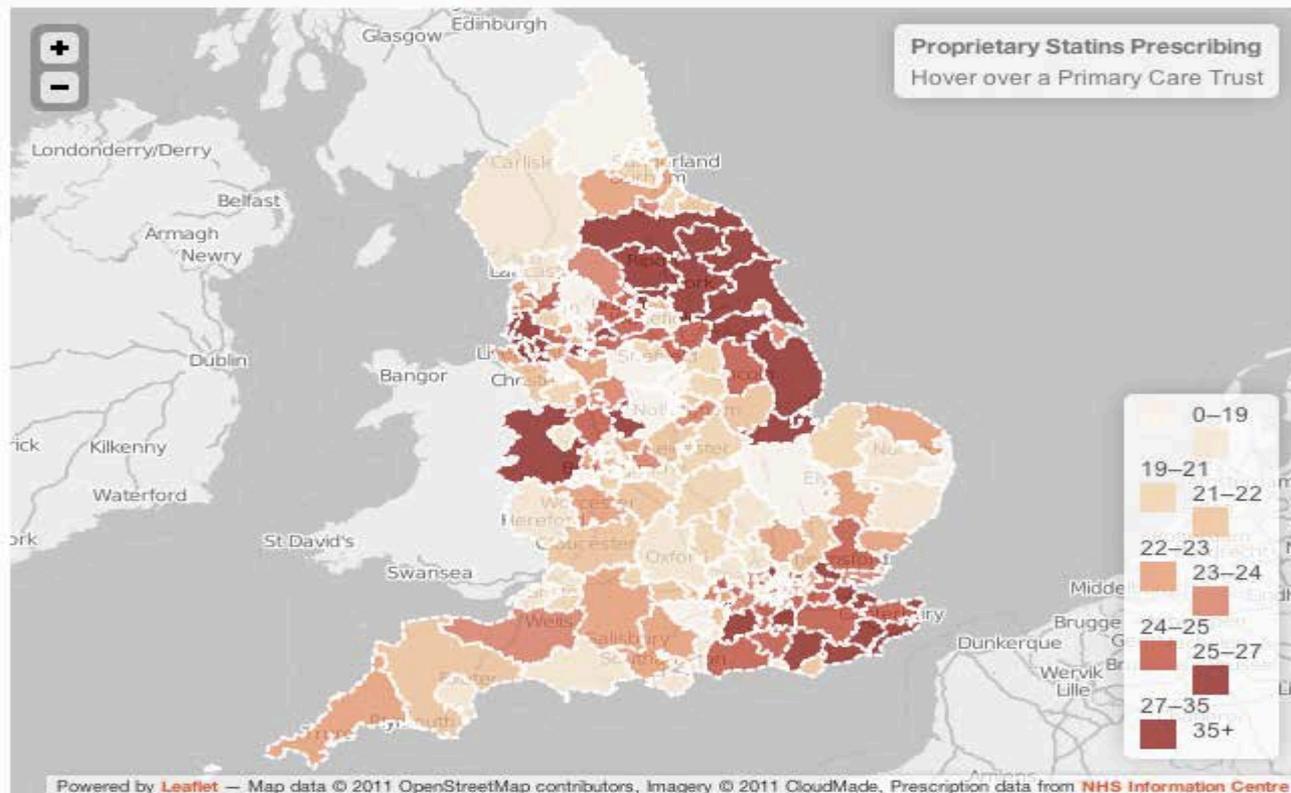
Video Time!

# Savings in the Health Sector

## NHS efficiency savings: the role of prescribing analytics

The NHS has been challenged to make £20 billion in “efficiency savings” by 2015 (1). £10 billion a year are spent by the NHS on essential drugs. Often, there’s a choice between a cheap “generic” medication, or an expensive “branded” one. Branded drugs can cost over ten times as much, for the same therapeutic benefit. “Prescribing Advisors” in the NHS, with the support of NICE, encourage doctors to use the most cost effective treatments. We have analysed exactly how much is spent on expensive “branded” medicines, for one class of drugs, namely statins, in England.

### Percentage of proprietary statin prescribing by CCG Sep 2011 - May 2012



Home

Aims & Objectives

NaCTeM Services

Text Mining Tools

Resources

Terms & Conditions

FAQ

News & Events

People

Projects

Publications

Community

External Collaboration

Vacancies

Teaching & Tutorials

Feedback

How to Find Us

Site Map

## Welcome to NaCTeM

The National Centre for Text Mining (NaCTeM) is the first publicly-funded text mining centre in the world. We provide text mining services in response to the requirements of the UK academic community. NaCTeM is operated by the University of Manchester.

On our website, you can find pointers to sources of information about text mining such as links to

- text mining services provided by NaCTeM
- software tools, both those developed by the NaCTeM team and by other text mining groups
- seminars, general events, conferences and workshops
- tutorials and demonstrations
- text mining publications

Let us know if you would like to include any of the above in our website.

### What text mining can do for you

Text mining offers a solution to the challenge of 'data deluge', information overload and information overlook. For more information, please see:

- [NaCTeM Brochure](#),
- [Text Mining Briefing Paper](#),
- [National Centre for Text Mining: an introduction to tools for researchers](#),
- [Vision for the Future](#),
- [Mining Biomedical Literature](#).
- [Event extraction for systems biology by text mining the literature](#)
- [Supporting the education evidence portal via text mining](#)

NaCTeM has developed text mining services and service exemplars for the UK academic community. Our services are underpinned by a

### Featured News

- [New paper and resources to support anatomical entity recognition at literature scale](#)
- [Keynote speech Pharma Documentation Ring special meeting in Bruges](#)
- [COLING 2014](#)
- [NaCTeM success at BioCreative IV](#)
- [Participation in Workshop on Text and Data Mining for Data Driven Innovation - Highlights available](#)
- [NaCTeM student selected to participate in Global Young Scientists Summit](#)

# New Medical Discoveries

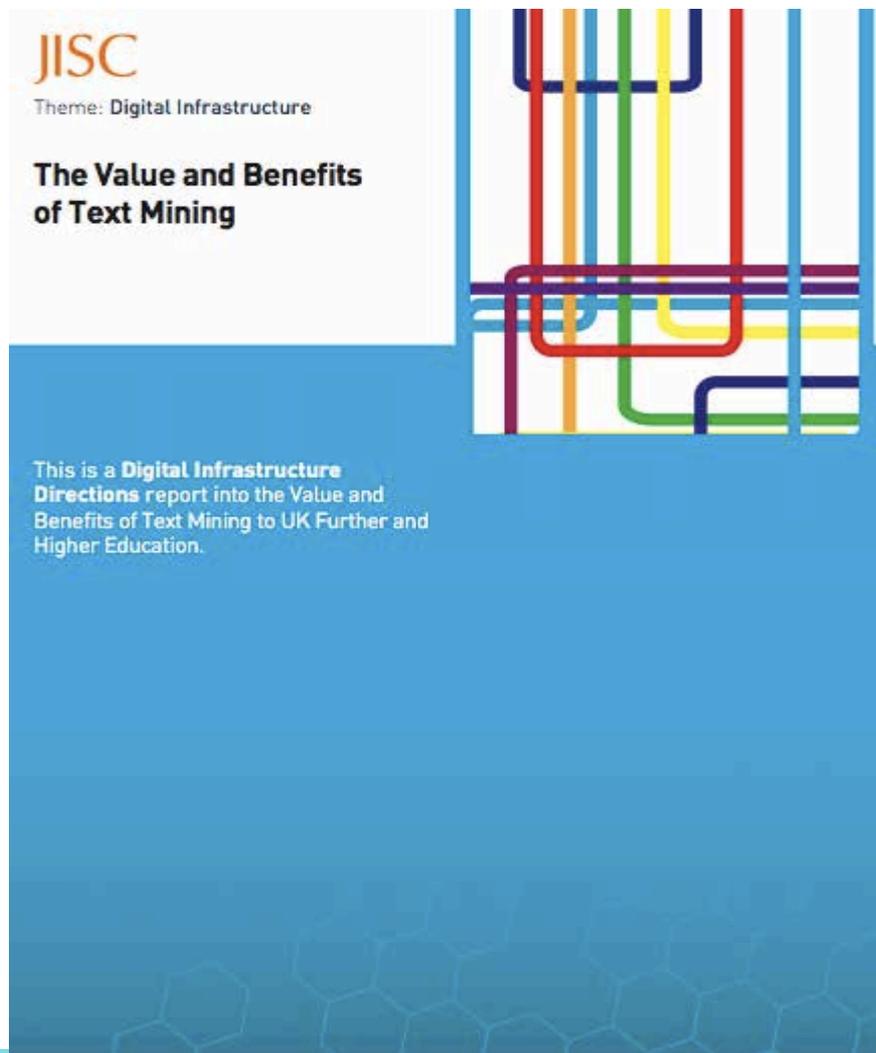
## Text mining suggests new uses for thalidomide

Marc Weeber and colleagues used automated text mining tools to infer that the drug thalidomide could treat several diseases it had not been associated with before. Thalidomide was taken off the market 40 years ago, but is still the subject of research because it seems to benefit leprosy patients via their immune systems. Weeber and Grietje Molema, an immunologist, used text mining tools to search the literature for papers on thalidomide and then pick out those containing concepts related to immunology. One concept, concerning thalidomide's ability to inhibit Interleukin-12 (IL-12), a chemical involved in the launch of an immune response, struck Molema as particularly interesting. A second automated search for diseases that improve when the action of IL-12 is blocked revealed several not previously linked with thalidomide, including chronic hepatitis, myasthenia gravis and a type of gastritis.

'Type in thalidomide and you get 2-3000 hits. Type in disease and you get 40,000 hits. With automated text mining tools we only had to read 100-200 abstracts and 20 or 30 full papers. We've created hypotheses for others to follow up,' says Weeber.

**Weeber et al. J Am Med Inform Assoc. 2003 10 252-259**

# Reduces Reading Times Exponentially



# Not Just Computer Scientists Either



© South Wiltshire Girls School

# The Right to Read is the Right to Mine?

- Facts and data not subject to copyright and database rights
- But computers have to copy in order to mine the data – so is it a licensable activity? (EU has an “internet browser” exception as browsers cache ...)
- European Union Commission stakeholder dialogue on TDM / “Licences for Europe” – Research / Library, Technology Sector and Open Access Publishers boycotted.



# The Right to Read is the Right to Mine?

- How would you license the internet?
- UKPMC – 75 publishers had articles with the word “malaria” in the title. BL’s experience of negotiating a new licence is 16 months on average.
- TDM goes across thousands / tens of thousands of articles which you **ALREADY** have legal access to. How can you **re**negotiate this with all publishers concerned?
- UK universities experiencing server access being suspended automatically when abnormal access is being detected.



Thank you



(unless indicated otherwise)