Digital Scholarship and Leiden University Libraries as a Partner in Knowledge

Isabel Brouwer
Leiden University Libraries

22 September 2014

Universiteit Leiden
The Netherlands

Strategic plan Leiden University Libraries 2011-2015

Mission:

- We are a Partner in Knowledge for researchers, teachers and students,
- by making available, accessible and managing
- scholarly information for education and research …

New areas of expertise:

- Data management
- Data curation
- Rights management
- Text- and data mining
- GIS

Leiden University Libraries Project Text- & data mining

- TDM digital content will become more important in Science and Humanities
- How do/will Leiden researchers use TDM on our collections?
- How can we – LeidenUniversity Libraries – support them?

Project team

- Librarians
  - Leiden University Libraries
  - Walaeus Library (LUMC)
- Researchers
  - Faculty of Humanities (LUCL, LUCAS)
  - Biosemantics group (LUMC)

Approach

Focus on two trends:

1) Humanities: increasing use of TDM on digital text corpora in the Humanities

2) Science: growing importance of TDM digital content for Literature Based Discovery

Phases

I. Exploration: desk research, symposia, workshops

II. Interviews
Pilots
Available sources
Selection tools for text analysis
Website

III. Evaluation, recommendations for library services

Preliminary conclusions (1)

Scale of research with TDM of digital collections:

- Science
  - Internationally TDM of STM content highly valued
  - In Leiden mainly done in Biomedical Sciences

European attention

Text and Data Mining: Its importance and the need for change in Europe.

Mapping Text and Data Mining in Academic and Research Communities in Europe

special briefing Issue 16/2014

By Sergey Filippov

Sergey Filippov is associate director of the Lisbon Council. A Dutch national, he previously served as assistant professor of innovation management at Delft University of Technology and holds a PhD in economics and policy studies of technical change from UNU-MERIT, a joint research institute of the United Nations University and University of Maastricht.

This report aims to map the scale of the use of text and data mining practices in the academic and research community in several European countries, and to benchmark the state of play in Europe against leading Asian and American countries.[1] The study first provides an overview of academic publications and patents pertaining to text and data mining. This quantitative data is then supplemented by in-depth interviews with leading researchers and text and data mining experts from a number of European countries and the United States. The paper is intended to serve as impartial input into current policy debates on reforming European copyright legislation to make it fit for the challenges and opportunities of the digital age.

Among the study's key findings, which will be analysed and developed in greater detail in the following pages, are the following:

1. Text and data mining is already an important tool for making sense of and finding value in data. The world of data is growing exponentially, offering new insights, better analytics and deeper understanding in a wealth of areas (including human health, analysis of traffic and migratory patterns, climate and environmental systems and more).[2]
2. There has been a strong increase in recent years in the number of publications and patents referring to text and data mining around the world. This growth is driven mostly by US and Asia (China in particular). US nationals are responsible for almost half of all publications and patents in the text and data mining field.

This interactive special briefing seeks to make knowledge more accessible through online circulation and interactive features, such as hotlinks to articles cited in the footnotes and a web-friendly format.

The opinions expressed in this special briefing are those of the author alone and do not necessarily reflect the views of the Lisbon Council or any of its associates.

1 The author would like to thank Christian Reimsbach-Kounatze of the Organisation for Economic Co-operation and Development (OECD), Gaetan De Rassenfosse of the University of Melbourne, and Ian Hargreaves and Paul Hofheinz of the Lisbon Council for comments on an early draft. Special thanks as well to Maja Bogataj Jančič, Antal van den Bosch, Henrik Boström, Alex Coad, Rishab Ghosh, José Guimón, Jun Hou, Can Huang, Pim Huijnen, Ashwin Ittoo, Dinar Kale, Paul Keller, Jan Knoerich, Izaskun Lacunza, Georg Licht, Boris Lokshin, Sarianna M. Lundan, Alex Mohr, Ismael Rafols, Nico Rasters, Francesco Rentocchini, Alexander Settles, Mariagrazia Squicciarini, Dariusz T. Stepniak, Jie Tang, Vlad Vaiman, Tim Vink, Fardad Zand and Erika Zoeller Veras. As always, any errors of fact or judgment are the author's sole responsibility.

2 Paul Hofheinz and Michael Mandel, Bridging the Data Gap: *How Digital Innovation Can Drive Growth and Create Jobs* (Brussels & Washington, DC: Lisbon Council and PPI, 2014).

Preliminary conclusions (2)

- Humanities
  Use of TDM in research & education:

  - Computational linguistics (LUCL)
  - History & Area Studies (Institute for History, LIAS, KITLV)
  - Book & Digital Media Studies (LUCAS)

Preliminary conclusions (3)

- Humanities collections issues:
  - Ocr quality
  - Intellectual Property Right
- Lot of interest but ‘how?’
- Not always that popular…

Interviewees: views &recommendations (1)

- There certainly is a role
- Collections:
  - enriched text, tools
  - advise on collections
  - advise on making digital text corpora

Interviewees: views &recommendations (2)

- Consultancy & information:
  - IPR, Licenses in relation toTDM
  - Ocr software & quality
  - TDM software, tools
  - How to make a researchbase?
  - Data storage & data management
  - ICT infrastructure (Leiden, CLARIN, etc.)
  - Organizations with other expertise like DANS
  - What research projects?

Interviewees: views &recommendations (3)

- Library as a physical & virtual place for collaboration in DH → visibility

- Organisation of meetings and workshops in collaboration with Faculty