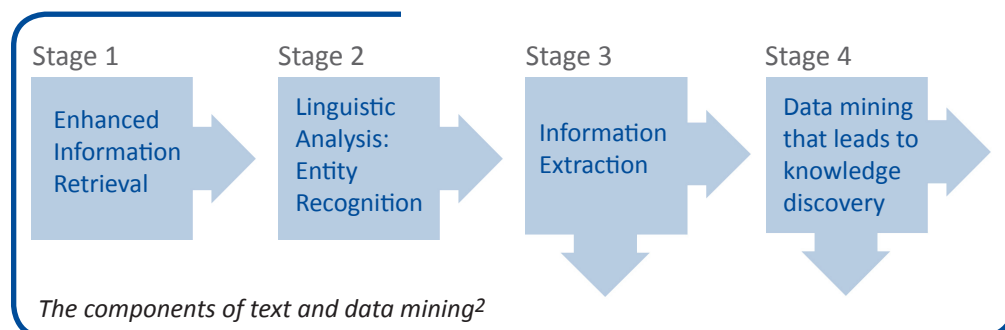# Text and Data Mining: Its importance and the need for change in Europe.

We believe that the right to read is the right to mine. Libraries and their users should be empowered to contribute to an innovative and competitive Europe.

What Is Text And Data Mining?

"Text and data mining (TDM) is the process of deriving information from machine-read material. It works by copying large quantities of material, extracting the data, and recombining it to identify patterns."[1]

There are four stages to the TDM process. First, potentially relevant documents are identified. These documents are then turned into a machine-readable format so that structured data can be extracted. The useful information is extracted (stage 3) and then mined (stage 4) to discover new knowledge, test hypotheses, and identify new relationships.

Stage 1	Stage 2	Stage 3	Stage 4
Enhanced Information Retrieval	Linguistic Analysis: Entity Recognition	Information Extraction	Data mining that leads to knowledge discovery

The components of text and data mining[2]

The bottom line is that text and data mining is concerned with the extraction of facts and data from vast quantities of material for research purposes, and not with substituting the original material.

Why Is It Important?

TDM will increase the progress of science exponentially. It has the potential to facilitate the discovery of cures for diseases such as cancer and Parkinson's. It has already been used to discover how existing drugs can be used to treat other conditions. It will also act as a foundation for innovation and new industry.

For libraries, who provide access to a growing amount of scientific content, it means that the researchers we support will be able to fully realise the value of the content we hold. This will, in turn, ensure a more rigorous approach to research, including more through reviews of the literature.

[1] Text Mining and Data Analytics in Call for Evidence Responses. UK Government http://www.ipo.gov.uk/ipreview-doc-t.pdf

[2] Value and Benefits of Text Mining. Dr Diane McDonald. Joint Information Systems Committee. 2012. http://www.jisc.ac.uk/publications/reports/2012/value-and-benefits-of-text-mining.aspx

Here are five key reasons that show why the current situation of negotiating special licenses for TDM is untenable:

1. Not all of the content that researchers wish to mine can be licensed. The internet is one example of this.

2. Libraries have already invested considerable time and effort in negotiating access to paid-for content. We believe that the right to read this content is also the right to mine it.

3. The negotiation of licences for TDM on a case-by-case basis is not scalable or sustainable over the long term.

4. It damages Europe's competitive advantage. Researchers in other countries such as the United States, Japan, Israel, Taiwan and South Korea already benefit from legal limitations and exceptions for such activities.

5. It goes against the open-science agenda, which advocates for research to be conducted in such a way that others can contribute, collaborate and add to the research effort.

Why Do Libraries Want A Copyright Exception?

Research libraries are at the centre of the data deluge. They are investing heavily in preserving born-digital content, digitising cultural heritage and in facilitating data sharing. Libraries are campaigning for freedom to apply TDM to this data because they know that doing so will maximise the return on investment of public money. A copyright exception for TDM will enable libraries and their users to contribute greatly to an innovative and competitive Europe.

This vision stands in stark contrast to the current situation. At the moment, researchers are hugely discouraged by a lack of clarity around copyright and ownership of derived works, as well as the complexity of negotiating licences. A copyright exception would allow them to use text and data mining methods to speed the pace of their research and to make new discoveries by analysing massive amounts of data with increasingly faster and more powerful technologies.

How Can You Help?

LIBER is actively advocating for a more flexible copyright system. So far, over 60 other influential organisations and individuals, representing researchers, science organisations and industry, have supported our call. You can join them. To learn more, please visit our website.

→ www.libereurope.eu/news/tdm

Look but don't touch...

Day and night, countless pieces of information flow through internet data centres such as this one owned by Google. For Europe's researchers, however, the lack of clarity surrounding copyright means they cannot use TDM to examine this knowledge and make new discoveries. The debate surrounding TDM was recently featured in the national Dutch newspaper *NRC*, and in the prominent science journal *Nature*. You can read the *Nature* article online to learn how TDM has already been used in projects to map the human brain and discover treatments for diseases. Read it here: www.nature.com/news/text-mining-spat-heats-up-1.12636

Image courtesy of Google.

www.libereurope.eu