

IFLA Statement on Text and Data Mining

As the leading international professional association concerned with information and library services, IFLA represents associations and institutions worldwide that endeavour to provide equitable access to a diversity of information.

IFLA maintains that legal certainty for text and data mining (TDM) can only be achieved by (statutory) exceptions. As an organization committed to the principle of freedom of access to information, and the belief that information should be utilised without restriction in ways vital to the educational and cultural well-being of communities, IFLA believes TDM to be an essential tool to the advancement of learning, and new forms of creation.

Copyright and database laws can affect the ability of libraries to fulfil their mandates and deliver information services for the benefit of their patrons, and can impede the use of materials by library users in ways that would benefit communities – for scholarship, research, improvements in health and science, creativity and social inclusion.

Digital information opens new opportunities for research and innovation

We live in an era of “Big Data”. OECD figures show that more digital information was created between 2008 – 2011 than in all previous recorded history.¹ No human can read such vast volumes of information, which is why “computer based reading”, using tools such as text and data mining, is so important.

Text and data mining (TDM) includes various technologies for the computer-based organisation and analysis of text and data. It can be defined in regards to computer learning as:

“The computer-based process of deriving or organising information from text or data. It works by copying large quantities of material, extracting the data, and recombining it to identify patterns, trends and hypotheses or by providing the means to organise the information mined.”²

The goal of TDM is to discover new knowledge from already existing knowledge. It also helps to sort the vast amounts of information that organisations today rely on. 21st century data driven innovation and research relies on computers being able to analyse the vast amounts of information available digitally. In the internet environment, characterised by an abundance of information in a diversity of forms, text and data mining has become an essential tool for researchers and innovators.

Research organisations see TDM as an engine to improve the performance of science by speeding up new potential discoveries based upon existing literature without the need for further laboratory based research. TDM is a tool also increasingly being used by researchers and creators in the arts and humanities fields, to offer new interpretations of history, literature and art. Libraries are also increasingly undertaking TDM themselves, to improve information services and offer new insights into their collections. Government data sets are also increasingly being made available to researchers, archives and libraries undertaking TDM, as

¹ World Economic Forum (2012) ‘Global Information Technology Report: living in a hyper-connected world’ p 59 http://www3.weforum.org/docs/Global_IT_Report_2012.pdf

² Text Mining and Data Analytics in Call for Evidence Responses. UK Government <http://www.ipso.gov.uk/ipreview-doc-t.pdf>

they offer much potential economic value in an era of Big Data. Commercial innovators are also utilising TDM.

The legal situation

In common with any use of a computer, in order to analyse text or data a computer must make a copy. The data to be examined can be derived from many different sources – these include, inter alia, databases that may be subject to licensing agreements, or material on the open web.

While facts and data are not protected by intellectual property laws, the text, documents or databases that are mined may well be subject to copyright, related rights and/or database rights. The extraction and copying of content one already has legal access to, and its transformation into a machine readable format, can touch on the rights holder's exclusive reproduction right. In addition, technical protection measures attached to databases that prevent reproduction are subject to legal protection.

The technical act of copying involved in the process of TDM falls by accident, not intention, within the complexity of copyright laws – in fact analysis of facts and data has been the basis of learning for millennia. As TDM simply employs computers to “read” material and extract facts one already has the right as a human to read and extract facts from, it is difficult to see how the technical copying by a computer can be used to justify copyright and database laws regulating this activity.

“That these new uses happen to fall within the scope of copyright regulation is essentially a side effect of how copyright has been defined, rather than being directly relevant to what copyright is supposed to protect.”³

TDM is one of several new tools in the digital environment to which copyright norms devised 300 years ago do not readily apply.

Solution

Researchers must be able to share the results of text and data mining, as long as these results are not substitutable for the original copyright work - irrespective of copyright law, database law or contractual terms to the contrary. Without this right, legal uncertainty may prevent important research and data driven innovation putting researchers, institutions and innovators at risk.

IFLA does not support licensing as an appropriate solution for TDM. If a researcher or research institution, or another user accessing information through their library, has lawfully acquired digital content, including databases, the right to read this content should encompass the right to mine. Further, the sheer volume and diversity of information that can be utilised for text and data mining, which extends far beyond already licensed research data bases, and which are not viewed in silos, makes a licence-driven solution close to impossible.

³ Hargreaves Review of Intellectual Property and Growth (2011), UK Intellectual Property Office, <http://www.ipso.gov.uk/ipreview.htm>