CENL Statement on Big Data

The Council of European National Librarians represents the national librarians of the member states of the Council of Europe. National libraries have an important role to play in supporting national cultural heritage through advocacy on research issues, and in the digital world are committed to building an integrated digital national library of Europe for the benefit of European culture and learning.

The amount of data available to mankind has exploded – in 2009 alone it has been estimated that the amount of data created and copied was equivalent to two stacks of DVDs reaching from the earth to the moon.[1] Having the right technologies, skills and legal framework to create economic values from big data across the European economy are essential – it has been estimated that optimal use of big data available from public bodies in the EU alone can create €250 billion annual value to Europe's economy.[2]

People can no longer read and make sense of terabytes and exabytes of information and are increasingly using computers to analyse it for them - a technique known as data mining. This technique means we have the ability to use computer programmes to search for facts, linkages, trends and relationships which in turn will speed up scientific discovery and other innovation exponentially. For example scientists using this technique have without expensive laboratory research been able to discover the link between osteoporosis and particular genes, as well as innovative new potential uses of existing medicines.[3][4] Also libraries as well as commercial organisations can use the same technique allowing us to make significant savings in organising data - from the creation of online catalogues through to the structuring of data holdings that reduce search and processing time. It is certain also that one of the most useful means used to access and analyse mass digitised content in a digital library will be the technique of data mining.

Although the extraction of facts or individual words is not subject to copyright law, and a human being copying a word or a fact with a pen or pencil is perfectly free to do so, due to the fact a computer must make a copy of an entire work in order to perform the same activity, the process of data mining becomes subject to copyright law and also database rights. Few contracts also allow data mining. In short, the law currently allows a human to read digital content they have bought or have legal access to, and make associations between the facts contained therein, but it does not permit a computer programme, based at the same location as the human researcher, to undertake the same work.

---

[1] www.emc.com/leadership/programs/digital-universe.htm

[2] McKinsey Global Institute: Big Data:The next frontier for innovation, competition and productivity.

[3] Thomas C Wiegers, Allan Peter Davis, K Bretonnel Cohen, Lynette Hirschman and Carolyn J Mattingly Text mining and manual curation of chemical-gene-disease networks for the Comparative Toxicogenomics Database (CTD). BMC Bioinformatics 2009, 10:326doi:10.1186/1471-2105-10-326 / Weeber et al. Journal of the American Medical Informatics Association. 2003 10 252–259

In order to create new innovation, economic value, as well as make significant cost savings for society the practice of data mining, and the barriers that exist due to current European copyright and database laws must be addressed as a matter of urgency. Copyright law which was developed over 300 years ago was not designed with such uses in mind. In order to ensure that copyright supports innovation in the future as it has done in the past, we must seek to ensure it facilitates new technological innovation and computer based techniques to analyse data.

Recommendation

Given this computer-based technology offers enormous potential to speed up life-changing scientific discovery, create significant organisational savings for libraries and private companies, as well as create many other forms of economic value, existing limitations and exceptions should be extended to explicitly cover the technique of data mining. Without this the innovation, productivity and growth that can come from harnessing the full potential of big data will never be realised.