DIGITALEUROPE

DIGITALEUROPE Position Paper
The issue of text and data mining within the modernisation of the EU copyright framework

Brussels, 1 February 2016

KEY MESSAGES

1. Text and data mining has an important potential for Europe's growth and competitiveness.

Text and Data Mining (TDM) is an evolving and burgeoning area of innovation. As such, it is unpredictable and likely to continue to evolve. Threats of copyright assertion will inevitably lead to slower and more difficult take up in Europe, and a competitive gap.

2. As technology companies, we believe the right to read, i.e. when a person has lawful access to content, includes the right to mine.

Once one has obtained lawful access to copyright protected content, the mining of this content is completely outside the realm of copyright. Therefore, a specific permission for TDM is not actually needed from a copyright perspective.

3. The EU's approach should not state or imply that text and data mining infringe copyright rules.

In particular, should a distinction be made between research and other uses of TDM, it should remain clear that TDM is not an infringement of copyright.

4. The EU should rather ensure that a clear legal framework encourages the use of computing techniques such as text and data mining in Europe.

We call on the Commission to ensure that the review of EU copyright rules will not impede data-driven innovation based on text and data mining.

THE SITUATION TODAY

The issue of **text and data mining (TDM)** is currently being considered within the frame of the EU copyright review. This issue was also discussed in the "Licenses for Europe" stakeholder dialogue and in a study on the legal framework of text and data mining", by De Wolf & partners in March 2014.

DIGITALEUROPE members are **concerned about the potential impact of the inclusion of text and data mining in the upcoming EU copyright review**. Notably, if not carefully considered, this can result to limitations to access and analysis of commercial and non-commercial copyrighted content, implications for big data analysis, creation of precedent relating to public data in relation to which copyright rules may not be clear today.

In more detail, DIGITALEUROPE members are **concerned by increasing attempts to assert copyright over text and data mining**. In the recent discussions in 'Licences for Europe, a number of stakeholders argued that TDM required licences above and beyond the right to access the content. Furthermore, although the starting point was limited to TDM for research and scientific purposes, similar considerations were discussed on TDM in other contexts. A legislative solution that implies that TDM infringes copyright can have a negative impact on a range of activities: innovation in online businesses; research across a wide swath of fields; and a host of other legitimate activities that do not harm the marketability of original works.

For more information please contact:
Damir Filipovic, Director – Digital Enterprise and Consumer Policy
+32 2 609 53 25 or damir.filipovic@digitaleurope.org

DIGITALEUROPE
Rue de la Science, 14 - 1040 Brussels [Belgium]
T. +32 (0) 2 609 53 10 F. +32 (0) 2 431 04 89
www.digitaleurope.org | info@digitaleurope.org | @DIGITALEUROPE
Transparency register member for the Commission: 64270747023-20

2

BACKGROUND INFORMATION

A technical illustration of text and data mining

Text mining varies in complexity and sophistication but generally requires a semantic analysis of the text is carried out using text mining software.

In our example, the semantic analysis merely counts the number of occurrences of individual terms in an individual document. Take for example the following two very short documents:

Doc 1: You can take that money to the bank.

Doc 2: You can take the walk along the river bank.

The semantic analysis is actually just a 'term by document matrix' or 'bag of words', looking like this:

Term	Document 1	Document2	Document…..
You	1	1	
Can	1	1	
Take	1	1	
That	1	0	
Money	1	0	
To	1	0	
The	1	2	
bank (1)	1	0	
Walk	0	1	
Along	0	1	
bank (2)	0	1	

These numbers can then be further analysed to give information about documents within the collection.

While mining is carried out at scale and with more sophistication, this is the basis process that unlocks the opportunities referred to above.

Therefore, when considering this technical analysis in the light of Article 2 on the Reproduction right and Article 5 on Exceptions and limitations of the Copyright Directive (Directive 2001/29/EC), it can be argued that from a legal perspective, the mining of such content is outside the realm of copyright.

DIGITALEUROPE
Rue de la Science, 14 - 1040 Brussels [Belgium]
T. +32 (0) 2 609 53 10 F. +32 (0) 2 431 04 89
www.digitaleurope.org | info@digitaleurope.org | @DIGITALEUROPE
Transparency register member for the Commission: 64270747023-20

3

Copyright & text and data mining

Copyright in its most basic and traditional sense protects an author from having his or her work copied by others. The "work" of an author is his or her creative content, or stated another way, the expression of the author's idea. Facts or news are not protected by copyright per se. Copyright protection only covers 'expression' and not ideas, procedures, methods of operation or mathematical concepts[1]. In addition, under EU law, facts are incapable of being 'original' in the EU sense[2], as result, using reported facts in research is not an infringement of copyright under the Infopaq test. In the UK Ian Hargreaves considers in Digital Opportunity: A review of Intellectual Property and Growth' (2011) that "Copyright is not intended to prevent use of facts for research". Furthermore, text and data mining is a clear instance of non-expressive or transformative use as it does not take away value from the original nor is it a substitute for the original work. It creates new value.

In all cases where e.g. text is being read on a computer, there are technically reproductions. Such content is temporarily reproduced on computers, in the cache or similar memories. Copyright would generally prohibit reading any copyrighted text on a computer unless there is some exception to copyright. When a rights holder (author or publisher) places an author's work on the internet or otherwise makes it accessible for someone with a browser to read, it should not be considered a violation of copyright to make a temporary copy of that text as it is only by making such copy that the text can be read.

Once the technically necessary copy is made, the further semantic analysis carried out by text mining does not implicate copyright protection; the "work" of the author is not copied in the traditional sense where the meaning behind the text is carried forward in a copy. Semantic analysis carried out in text mining does not "copy" any "expression" of an author. It simply analyses the individual words which are not, in and of themselves, subject to copyright protection and Article 2 of the Copyright Directive. This leads to the conclusion that once the work is made available to be read on a computer, the mining of that text is completely outside the realm of copyright.

Therefore, despite some suggestions to the contrary, TDM is not currently, and cannot be, subject to copyright protection. Text and data mining potentially encompasses a broad range of technologies across-sectors. These technologies essentially serve to analyse information through new computing techniques. In many ways, what a researcher could achieve through reading and analysing a text and a set of data - an activity that has never been, and should never be, an infringement of copyright - can be achieved on a larger scale through computer analysis/ In so far as these technologies consist in extracting new information or facts from existing materials, they are not in general subject to copyright protection.

---

1 See for example. TRIPs Article (2), WIPO Guide to the Berne Convention: Article 2(1)
2 As described in CJEU case C-5/08 Infopaq

ABOUT DIGITALEUROPE

DIGITALEUROPE represents the digital technology industry in Europe. Our members include some of the world's largest IT, telecoms and consumer electronics companies and national associations from every part of Europe. DIGITALEUROPE wants European businesses and citizens to benefit fully from digital technologies and for Europe to grow, attract and sustain the world's best digital technology companies.

DIGITALEUROPE ensures industry participation in the development and implementation of EU policies. DIGITALEUROPE's members include 61 corporate members and 37 national trade associations from across Europe. Our website provides further information on our recent news and activities: http://www.digitaleurope.org

DIGITALEUROPE MEMBERSHIP

Corporate Members

AMD, Airbus, Apple, BlackBerry, Bose, Brother, CA Technologies, Canon, Cassidian, Cisco, Dell, Epson, Ericsson, Fujitsu, Google, Hitachi, Hewlett Packard Enterprise, HP Inc., Huawei, IBM, Ingram Micro, Intel, iQor, JVC Kenwood Group, Konica Minolta, Kyocera, Lenovo, Lexmark, LG Electronics, Loewe, Microsoft, Mitsubishi Electric Europe, Motorola Solutions, NEC, Nokia, Nvidia Ltd., Océ, Oki, Oracle, Panasonic Europe, Philips, Pioneer, Qualcomm, Ricoh Europe PLC, Samsung, SAP, SAS, Schneider Electric IT Corporation, Sharp Electronics, Siemens, Sony, Swatch Group, Technicolor, Texas Instruments, Toshiba, TP Vision, VMware, Western Digital, Xerox, Zebra Technologies, ZTE Corporation.

National Trade Associations

Austria: IOÖ
Belarus: INFOPARK
Belgium: AGORIA
Bulgaria: BAIT
Cyprus: CITEA
Denmark: DI Digital, IT-BRANCHEN
Estonia: ITL
Finland: FFTI
France: AFDEL, AFNUM, Force Numérique

Germany: BITKOM, ZVEI
Greece: SEPE
Hungary: IVSZ
Ireland: ICT IRELAND
Italy: ANITEC
Lithuania: INFOBALT
Netherlands: Nederland ICT, FIAR
Poland: KIGEIT, PIIT, ZIPSEE
Portugal: AGEFE
Romania: ANIS, APDETIC

Slovakia: ITAS
Slovenia: GZS
Spain: AMETIC
Sweden: Foreningen Teknikföretagen i Sverige, IT&Telekomföretagen
Switzerland: SWICO
Turkey: Digital Turkey Platform, ECID
Ukraine: IT UKRAINE
United Kingdom: techUK

DIGITALEUROPE
Rue de la Science, 14 - 1040 Brussels [Belgium]
T. +32 (0) 2 609 53 10 F. +32 (0) 2 431 04 89
www.digitaleurope.org | info@digitaleurope.org | @DIGITALEUROPE
Transparency register member for the Commission: 64270747023-20

5