Abstract

Copyright Limitations and Exceptions for 'Big Data' Analysis and 'Data Mining'

Kim, Byungil*, Shin, Hyunchul**, Ahn, Chang-Won***

Text and data mining (TDM) is the use of automated analytical techniques to analyse text and data for patterns, trends and other useful information. 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. TDM is considered to be a tool to facilitate research at all levels and bring to light new aspects of and insights into existing data by analysing, combining, isolating or synthesizing information with the aid of an automated tool.

TDM almost always involves copying, but not all copying amounts to copyright infringement. However, many researchers want to keep a durable copy of the articles that underlie their TDM analysis as a reference to validate their results. In other cases, the durable outputs of TDM may have some copyrighted content, such as images. Researchers using TDM in their research risk infringing copyright unless they had specific permission from the copyright owner. This paper gives a brief overview of the copyright issues under which TDM practices fall and legal solutions for a copyright exception for TDM.

Keywords

Big Data, Text Mining, Data Mining, Copyright Limitations and Exceptions, Copyright Infringement

* Hanyang University
** Osaka University
*** ETRI