Copyright Exceptions for Text and Data Mining in China: Inspiration from Transformative Use

Dr. (Jerry) Jie HUA
Associate Professor, Shanghai International College of Intellectual Property, Tongji University (Shanghai, China)

Abstract: Text and data mining (TDM) refers to the creation of new, non-obvious information such as patterns, trends or relationships from a collection of textual documents including books, images, webpages, emails, or reports. TDM helps to dig out the hidden gold from textual information and leaps from old-fashioned retrieval to information and knowledge discovery. Combined with artificial intelligence and deep learning technologies, TDM is widely applied to and thrives in linguistic applications, scientific research, legal research, financial analysis, health care, education and biomedicines. TDM plays an important role in China’s policies of promoting data sharing and opening. Chinese companies which have massive data and the ability to process massive data are expected to launch TDM projects.

TDM often works by copying substantial quantities of materials, extracting the data and recombining it to identify patterns into the final output. TDM thus will arouse copyright infringement problem if materials used for TDM are original works still under copyright term of protection. As large number of materials will be used in TDM, it is unrealistic to seek permission from every single copyright owner before copying the materials. Even if mass permission can be pursued with the aid of copyright collective management organizations, costs for conducting TDM thus will be very high, which will hinder many non-commercial entities or small and medium enterprises to engage in TDM and open data wave. Exceptions that allow use of copyright works without prior authorization from, and payment of remuneration to, the copyright owners, are key to resolve the issue.

This article will analyze copyright exceptions for TDM based on the third amendment of the Copyright Law and judicial practice in China. Based on the discussion on failure of the third amendment of the Copyright Law to cover TDM exception, this article will observe the TDM exceptions in the European Union and the United States, two typical jurisdictions that incorporate TDM exception by different models; analyze copyright exceptions in legislations and judicial practice in China, especially the transformative use recently developed by Chinese practitioners; and ultimately suggest recommendations for China to embody both certainty and flexibility in copyright exceptions generally and for TDM especially.