

# QUICK REFERENCE GUIDE TEXT AND DATAMINING

Text and data mining (TDM) is the automated searching of text and databases to discover (new) patterns, trends and connections. The techniques from TDM are increasingly being applied to scientific literature, such as journal articles or monographs. TDM can be applied within all stages of the research process.

TDM examples:

- **information retrieval:** a growing number of publishers explicitly allow text mining in their license terms. Although publishers may not contractually exclude TDM, they can impose conditions on it and on the ways in which text mining can be performed. These may differ per publisher, for example only allowing text mining with the tools they offer themselves.
- **information extraction:** identifying matters such as personal names, organisations or subjects within texts and the establishing of relationships.
- **datamining:** identifying correlations, regularities or other patterns within texts.

## COPYRIGHT AND TEXT- AND DATAMINING

Text and data mining techniques can be applied to resources that are in the public domain and to copyrighted material.

Within projects using TDM, local copies of texts are generally made. Making copies is normally reserved for the copyright holder. However, universities (and other non-commercial research institutions) are permitted to make copies for the purposes of TDM under sections 15n and 15o of the Copyright Act (effective from 7 June 2021).

TDM is possible for researchers for all works to which they have legitimate access. For example, material that is freely accessible via the Internet or through access to the organisation's catalogue.

Publishers may not contractually exclude TDM. They are not allowed to include in the license with the university that TDM is not allowed. Nor may publishers put up technical barriers that make TDM impossible. There is a growing number of publishers that explicitly allow text mining in their license conditions. Although publishers may not contractually exclude TDM, they can impose conditions on it and on the ways in which text mining can be performed. These may differ per publisher, for instance only allowing text mining with the tools they offer themselves.

## QUESTIONS?

Please contact one of the members of staff at the [Copyright Information Point \(AIP\)](#) of your institution.

SURF

[info@auteursrechten.nl](mailto:info@auteursrechten.nl) / [www.auteursrechten.nl](http://www.auteursrechten.nl)

Aan dit document kunnen geen rechten worden ontleend.

Beschikbaar onder de licentie Creative Commons Naamsvermelding 3.0 Nederland.

[www.creativecommons.org/licenses/by/3.0/nl](http://www.creativecommons.org/licenses/by/3.0/nl)

[Typ hier]

