

Researchers review the state-of-the-art text mining technologies for chemistry

June 22 2017

In a recent *Chemical Reviews* article, Spanish researchers have published the first exhaustive revision of the state-of-the-art methodologies underlying chemical search engines, named entity recognition and text mining systems.

The rapidly growing field of big data applications in biomedical research, together with the use of machine learning and artificial intelligence technologies for text data mining, has resulted in promising tools. The authors write, "This review is organised to serve as a practical guide to researchers entering in this field but also to help them to envision the next steps in this emerging data science field."

"Through the release of Gold Standard datasets and the organisation of several community challenge benchmark events, the Biological Text Mining Unit has played a critical role in the development and evaluation of current [chemical](#) text mining systems, as highlighted in this article," explains Martin Krallinger, head of the unit and co-first author of the review.

A huge amount of unstructured data

A considerable fraction of biomedically relevant data is only available in the form of unstructured data. This type of data includes rapidly growing scientific literature, medicinal chemistry patents, electronic health records and clinical trial documents. In fact, every year, over 20,000 new

compounds are published in medicinal and biological chemistry journals.

Being able to transform unstructured biomedical research data into structured databases that can be more efficiently processed by machines or queried by humans is critical for a range of heterogeneous applications. These include the identification of new drug targets and chemical probes to validate/discard those new potential targets, re-purposing of approved drugs, the identification of adverse drug events or retrieval of systems biology associated with chemical-disease or chemical-gene networks.

As a therapeutic strategy to treat medical needs, chemical compounds constitute a key entity type of critical relevance for [biomedical research](#). "The construction of large chemical knowledge bases, integrating chemical information with biological and clinical data, is crucial to identify and validate new therapeutic targets for unmet medical needs as well as to speed up the drug discovery process," says Julen Oyarzabal, director of Translational Sciences at CIMA and co-leader of this report.

More information: Martin Krallinger et al, Information Retrieval and Text Mining Technologies for Chemistry, *Chemical Reviews* (2017).
[DOI: 10.1021/acs.chemrev.6b00851](https://doi.org/10.1021/acs.chemrev.6b00851)

Provided by Centro Nacional de Investigaciones Oncológicas (CNIO)

Citation: Researchers review the state-of-the-art text mining technologies for chemistry (2017, June 22) retrieved 19 April 2024 from <https://phys.org/news/2017-06-state-of-the-art-text-technologies-chemistry.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.