

Text Mining for Chemical Compounds

1. The ambiguity of non-systematic chemical identifiers within chemical databases is much lower than the ambiguity of these identifiers across databases. (this thesis)
2. Integrating existing databases by using systematic identifiers as indexes for integration will introduce inconsistencies to the integrated database. (this thesis)
3. Well-defined and documented chemistry standardisation rules applied to all compounds can greatly decrease the number of errors within chemical databases and expedite integration of such databases. (this thesis)
4. The performance of a text-mining tool should be judged in view of the inter annotator agreement on the corpus that was used to develop the tool. (this thesis)
5. A dictionary-based approach to extract chemical entities from patents will result in low recall and should be supplemented by rule-based or statistical approaches. (this thesis)
6. Quality is not a coincidence. It is always the result of intelligent effort.
7. Thanks to big data, machines can now be programmed to do the next thing right. But only humans can do the next right thing. (Dov Seidman)
8. Without big data analytics, companies are blind and deaf, wandering out onto the web like deer on a freeway. (Geoffrey Moore)
9. The science of today is the technology of tomorrow. (Edward Teller)
10. The price of light is less than the cost of darkness. (Arthur Nielsen)
11. The wind is always against you when you are cycling.

Saber A. Akhondi

Rotterdam, 2 October 2018