RESEARCH

A sample article title

Jane E Doe[1*†] and John RS Smith[1,2]

Abstract

First part title: Text for this section.

Second part title: Text for this section.

Keywords: sample; article; author

Content

Text and results for this section, as per the individual journal's instructions for authors.

Background

Incessant influx of new data on chemical entities and their properties steadily increases our capabilities for information retrieval in biomedical domain. However, ever-growing rates at which new information is being produced make it infeasible for any specialist to analyze manually even if confined within a specific subject. On the other hand the prevalent formats for such publication (paper, theses or a patent) is not structured in a way that is suitable for automatic processing[1]. Various databases containing diverse chemical information, such as PubChem[2], ChemSpider[3], CAS Registry[4], etc., partially address this problem by accumulating information from different sources and storing it in an organized, easy-to-process manner. But the maintenance and curation of this kind of databases still requires substantial manual labour from the experts in the field. Moreover, a standardized entry to a database usually omits many of the nuances transmitted by a natural languge. For this reason, in recent years special interest is given to the instruments for automatic information extraction from text[5].

Named entities recognition (NER) is an essential step in the process of extracting chemical information from text. The problem of NER is not uniqeue to chemical entities, but there are several features characteristic to the given subject. A lack of universal naming convention, the existance of multiple synonyms for a single entity (e.g. EXAMPLE), high sensitivity to spelling errors and a hard-to-parse syntax with multiple whitespaces, punctuation marks and numbers inside single entity name are some of the most obviouse examples[6, 7]. Significance of text mining and information extraction in biomedical domain is outlined by such initiatives as BioCreative challenge[8] and chemical named entities recognition (CNER) in particular by CHEMDNER task[9].

Historically, first approaches used for CNER were dictionary-based and rule-based. Dictionary-based algorithms rely on string matching to a domain-specific collection of terms. They are highly dependent on the quality of this term collection, which can get outdated in time and hence requires constant expansion and curation[10, 11]. Rule-based algorithms make use of manually created and updated rules based on morphological and contextual information. These rules allow users to describe potentially endless set of names if they follow an orderly convention such as IUPAC, but do not port as well on trivial names and like dictionaries require maintenance by the experts[12].

*Correspondence: jane.e.doe@cambridge.co.uk
[1]Department of Zoology, Cambridge, Waterloo Road, London, UK
Full list of author information is available at the end of the article
†Equal contributor

Competing interests
The authors declare that they have no competing interests.

Author's contributions
Text for this section ...

Acknowledgements
Text for this section ...

Author details
[1]Department of Zoology, Cambridge, Waterloo Road, London, UK.
[2]Marine Ecology Department, Institute of Marine Sciences Kiel, Düsternbrooker Weg 20, 24105 Kiel, Germany.

References

1. Eltyeb, S., Salim, N.: Chemical named entities recognition: a review onapproaches and applications. Journal of Cheminformatics **6**, 17 (2014)
2. The PubChem Project. https://pubchem.ncbi.nlm.nih.gov/
3. ChemSpider Database. http://www.chemspider.com/
4. CAS Registry. https://www.cas.org/content/chemical-substances

5. Swain, M.C., Cole, J.M.: Chemdataextractor: A toolkit for automated extraction of chemical information from the scientific literature. Journal of Chemical Information and Modeling **56**(10), 1894–1904 (2016). doi:10.1021/acs.jcim.6b00207. PMID: 27669338. http://dx.doi.org/10.1021/acs.jcim.6b00207
6. Brecher, J.: Name=struct: A practical approach to the sorry state of real-life chemical nomenclature. J. Chem. Inf. Comput. Sci. **39**, 943–950 (1999)
7. Hettne, K.M., et al.: Automatic vs. manual curation of a multi-source chemical dictionary: the impact on text mining. J. Chem. Inf. **2:3** (2010)
8. Hirschman, L., Yeh, A., Blaschke, C., Valencia, A.: Overview of biocreative: critical assessment of information extraction for biology. BMC Bioinformatics **6**(1), 1 (2005). doi:10.1186/1471-2105-6-S1-S1
9. Krallinger, M., Leitner, F., Rabal, O., Vazquez, M., Oyarzabal, J., Valencia, A.: Chemdner: The drugs and chemical names extraction challenge. Journal of Cheminformatics **7**(1), 1 (2015). doi:10.1186/1758-2946-7-S1-S1
10. Hettne, K.M., Stierum, R.H., Schuemie, M.J., Hendriksen, P.J.M., Schijvenaars, B.J.A., Mulligen, E.M.v., Kleinjans, J., Kors, J.A.: A dictionary to identify small molecules and drugs in free text. Bioinformatics **25**(22), 2983–2991 (2009). doi:10.1093/bioinformatics/btp535. /oup/backfile/content_public/journal/bioinformatics/25/22/10.1093_bioinformatics_btp535/2/btp535.pdf
11. Rebholz-Schuhmann, D., Kirsch, H., Arregui, M., Gaudan, S., Riethoven, M., Stoehr, P.: Ebimed—text crunching to gather facts for proteins from medline. Bioinformatics **23**(2), 237–244 (2007). doi:10.1093/bioinformatics/btl302. /oup/backfile/content_public/journal/bioinformatics/23/2/10.1093/bioinformatics/btl302/2/btl302.pdf
12. Budi, I., Bressan, S.: A biological named entity recognizer. Pac Symp Biocomput, 427 (2003)

Figures

Figure 1 Sample figure title. A short description of the figure content should go here.

Figure 2 Sample figure title. Figure legend text.

Tables

Table 1 Sample table title. This is where the description of the table should go.

	B1	B2	B3
A1	0.1	0.2	0.3
A2	...	..	.
A3	..	.	.

Additional Files

Additional file 1 — Sample additional file title
Additional file descriptions text (including details of how to view the file, if it is in a non-standard format or the file extension). This might refer to a multi-page table or a figure.

Additional file 2 — Sample additional file title
Additional file descriptions text.