

Databases and ontologies

Ontoclick: a web browser extension to facilitate biomedical knowledge curation

Anthony Xu¹, Aravind Venkateswaran¹, Lianguizi Zhou¹ and Andreas Zankl ^{2,3,4,*}

¹School of Computer Science and Engineering, UNSW, Sydney, NSW 2052, Australia, ²Department of Clinical Genetics, The Children's Hospital at Westmead, Sydney, NSW 2145, Australia, ³Faculty of Medicine and Health, The University of Sydney, Sydney, NSW 2006, Australia and ⁴Bone Biology Division and Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Sydney, NSW 2010, Australia

*To whom correspondence should be addressed.

Associate Editor: Janet Kelso

Received on November 20, 2020; revised on June 24, 2021; editorial decision on July 2, 2021; accepted on July 14, 2021

Abstract

Motivation: Knowledge curation from the biomedical literature is very valuable but can be a repetitive and laborious process. The paucity of user-friendly tools is one of the reasons for the lack of widespread adoption of good biomedical knowledge curation practices.

Results: Here, we present Ontoclick, a web browser extension that streamlines the process of annotating a text span with a relevant ontology term. We hope this tool will make biocuration more accessible to a wider audience of biomedical researchers.

Availability and implementation: Ontoclick is freely available under the GPL-3.0 license on the Chrome Web Store and on the Mozilla Add-Ons for Firefox Store. Source code and documentation are available at: <https://github.com/azankl/Ontoclick>.

Contact: andreas.zankl@sydney.edu.au

1 Introduction

The Human Phenotype Ontology (HPO) includes over 10 000 terms to describe the clinical features of patients with rare diseases and has proved its utility in areas that require computational processing of phenotype data, such as clinical genomics (Köhler *et al.*, 2017). However, annotating a biomedical text such as a case report in the medical literature with HPO terms remains a laborious process. Typically, a curator would identify a text span that represents a clinical phenotype (e.g. 'abnormally large head') and then search for this term on the HPO website. As it's not always clear what the correct HPO term is, this often requires some trial and error. For example, searching for 'abnormally large head' returns no results, while searching for 'large head' leads to the term 'HP:0000256 macrocephaly'. Studying the term's definition and list of synonyms provided by the HPO website reveals that this is the correct term (macrocephaly is medical jargon for an abnormally large head). Depending on the curation workflow, the curator might then copy the term identifier (HP:0000256), the term label ('macrocephaly') or both into a spreadsheet or other application. The curator might also have to manually copy the associated text span from the original text ('abnormally large head') and a source text identifier (e.g. the

PubMed ID) into this spreadsheet or other application to complete the annotation.

2 The Ontoclick browser extension

To facilitate this process, we have developed Ontoclick, an extension for the Chrome and Firefox web browser. Ontoclick lets the user highlight a piece of text and then tries to find a matching HPO term for it. Matching terms are returned to the user for review (Fig. 1).

Ontoclick can display a term's definition and synonyms and provides multiple options for copying the chosen term's identifier and/or label, with or without the highlighted text span. The copied information can be stored on the system clipboard, ready to be pasted into a spreadsheet or other application. Alternatively, multiple annotations can be stored in an internal list and exported in CSV format. Each row in the CSV file represents an annotation with separate columns for the highlighted text span, term ID and term label. The first row of the file contains the URL of the page on which the annotations were created. If the annotations were created on an abstract on the NCBI PubMed website, the first row will contain the PubMed ID of the annotated abstract.

Li Hu^{1,2}, Xike Wang³, Tingting
 , Shujuan Yan², Xiaoling Fu³,

Affiliations + expand

PMID: 32141654 PMCID: PMC

Free PMC article

Abstract

Background: 3-M syndrome is a rare genetic disorder characterized by growth retardation, large head changes, which is associated with CCDC8. Only a few 3-M syndromes

Methods: Children with unexplained intelligence in two Chinese families (trio-WES) and pathogenicity prediction conservative analysis of the mutant wild-type (WT) and mutant CUL7

Results: We identified a homozygous p.Thr1633Met) in CUL7 gene in

The screenshot shows the Ontoclick interface. At the top, there is a search bar containing the text 'large head'. Below the search bar, there is a dropdown menu showing 'Human Phenotype Ontology (HPO)'. Underneath, there is a section for 'NCBO Bioportal Search' which contains a table of results. The table has three columns: 'Identifier', 'Label', and 'Actions'. The first row is highlighted and shows the identifier 'HP:0000256' with a link icon, the label 'Macrocephaly' with a link icon, and three action icons (copy, edit, delete). Below this row, there is a 'Definition' section: 'Occipitofrontal (head) circumference greater than 97th centile compared to appropriate, age matched, sex-matched normal standards. Alternatively, a apparently increased size of the cranium.' followed by a 'Synonyms' section: 'Big head, Increased size of head, Big calvaria, Big skull, Large cranium, Large calvaria, Large skull, Big cranium, Megacephaly, Increased size of cranium, Macrocrania, Large head, Large head circumference, Increased size of skull'. Below the definition and synonyms, there are four more rows in the table, each with a '+' sign in the first column, an identifier, a label, and action icons.

Identifier	Label	Actions
- HP:0000256	Macrocephaly	[Copy] [Edit] [Delete]
+ HP:0001176	Large hands	[Copy] [Edit] [Delete]
+ HP:0002003	Large forehead	[Copy] [Edit] [Delete]
+ HP:0030865	Large elbow	[Copy] [Edit] [Delete]
+ HP:0100729	Large face	[Copy] [Edit] [Delete]

Fig. 1. The Ontoclick popup showing HPO terms matching the highlighted text span

In addition to the HPO, Ontoclick also allows searching for matching terms in other ontologies. We currently support searches against the Gene Ontology (Ashburner et al., 2000; Gene Ontology Consortium, 2021), Mondo (Mungall et al., 2017), Disease Ontology (Schriml et al., 2019), Orphanet Rare Disease Ontology (<http://www.orphadata.org/>). We plan to add more ontologies in the future. Requests to add specific ontologies can be submitted through the Ontoclick Github Issue Tracker.

Ontoclick uses the NCBO BioPortal Search API (Musen et al., 2012) by default, but searches can also be performed with the NCBO BioPortal Annotator API, the EBI Ontology Lookup Service Search API (Jupp et al., 2015) and the HPO Search API (Köhler et al., 2021) (the latter two only support searches against the Human Phenotype Ontology).

Detailed installation and usage instructions are available at: <https://github.com/azankl/Ontoclick>.

3 Conclusion

We believe Ontoclick is a quick and user-friendly way to annotate biomedical texts on the web with ontology terms. While the current version is mostly geared at the HPO, the tool can be easily adapted to other ontologies. We hope that streamlined curation tools like Ontoclick will lead to more widespread adoption of biomedical curation in biomedical research.

Financial Support: none declared.

Conflict of Interest: none declared.

Acknowledgements

Many thanks to Graham Towse for help on early versions of Ontoclick.

References

- Ashburner, M. et al. (2000) Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.*, 25, 25–29.
- Gene Ontology Consortium. (2021) The gene ontology resource: enriching a GOld mine. *Nucleic Acids Res.*, 49, D325–D334.
- Jupp, S. et al. (2015) A new ontology lookup service at EMBL-EBI. In: *Proceedings of the 8th Semantic Web Applications and Tools for Life Sciences (SWAT4LS) International Conference, Cambridge UK*.
- Köhler, S. et al. (2017) The human phenotype ontology in 2017. *Nucleic Acids Res.*, 45, D865–D876.
- Köhler, S. et al. (2021) The human phenotype ontology in 2021. *Nucleic Acids Res.*, 49, D1207–D1217.
- Mungall, C.J. et al. (2017) The monarch initiative: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res.*, 45, D712–D722.
- Musen, M.A. et al.; NCBO Team. (2012) The national center for biomedical ontology. *J. Am. Med. Inform. Assoc.*, 19, 190–195.
- Schriml, L.M. et al. (2019) Human disease ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Res.*, 47, D955–D962.