Knowledge Graphs as Educational Tools in Biomedical Education

Ana Carolina Pereira
carolinadpereira18@gmail.com
Lisbon School of Engineering, IPL
Lisbon, Portugal

Matilde Pato
matilde.pato@isel.pt
Lisbon School of Engineering, IPL
IBEB & LASIGE, Faculty of Sciences of the University of Lisbon
NOVA LINCS, NOVA School of Science and Technology
Monte da Caparica, Portugal

Nuno Datia
nuno.datia@isel.pt
Lisbon School of Engineering, IPL
NOVA LINCS, NOVA School of Science and Technology
Monte da Caparica, Portugal

Abstract

Knowledge graphs (KGs) offer powerful visualization capabilities that can transform how complex biomedical concepts are analyzed and potentially taught. This paper explores the educational potential of our biomedical KG methodology – originally developed for drug research – as a tool for biomedical education. By integrating data from DailyMed and FDA databases through Named Entity Recognition, we create intuitive visualizations of drug-disease relationships, approval timelines, and administration patterns that reveal pharmaceutical knowledge in ways traditional formats cannot. Our KG framework encompasses 561 drugs and 176 diseases connected through more than 1,000 relationships, providing rich visual contexts to understanding complex biomedical interactions. We discuss how the intuitive representation of pharmaceutical relationships aligns with how people naturally process information, suggesting promising applications for enhancing understanding of biomedical concepts in educational settings. This work highlights how computer science innovations can potentially catalyse educational change in health sciences, making complex domain knowledge more accessible through visual, interactive experiences. The relevant datasets and code are available at https://github.com/matpato/MedJsonify.git.

CCS Concepts

• **Information systems → Hierarchical data models**; **XPath**; **Ontologies**; **Extraction, transformation and loading**; **Data cleaning**; • **Applied computing → Bioinformatics**.

Keywords

Knowledge Graphs, Biomedical Data Visualization, Drug-Disease Relationships, Visual Analytics, Graph Databases, Neo4j, Named Entity Recognition, Pharmaceutical Knowledge Representation

ACM Reference Format:

Ana Carolina Pereira, Matilde Pato, and Nuno Datia. 2018. Knowledge Graphs as Educational Tools in Biomedical Education. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX).* ACM, New York, NY, USA, 2 pages. https://doi.org/XXXXXXX.XXXXXXX

Conference acronym 'XX, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Medical and pharmacy education faces challenges in helping students comprehend complex interconnections between drugs, diseases, and biological mechanisms. Traditional textbook approaches often struggle to convey multidimensional pharmaceutical knowledge, leading to fragmented understanding. Knowledge graphs (KGs) offer a solution by visually representing relationships between entities in an intuitive and explorable format [1]. This paper extends our previous work on biomedical KGs for drug research [2] – originally developed for research purposes, to address educational challenges in health sciences. We demonstrate how graph-based visualization techniques can transform abstract pharmaceutical knowledge into accessible visual patterns that enhance student learning. We focus on three applications: (1) visualizing drug-disease relationships, (2) illustrating historical drug approval patterns, and (3) representing administration routes for practical training.

2 Methodology

Our KG framework builds upon our established methodology for biomedical data integration as depicted in Figure 1. We integrate data from DailyMed, Orange and Purple Books databases, with preprocessing to ensure data quality ❶. Named Entity Recognition (NER) establishes unique identifiers using a hierarchical approach: DrugBank [3], ChEBI [4], or ingredient-based IDs for drugs; Disease Ontology [5] and Orphanet [6] for conditions. We implemented data wrangling to standardize heterogeneous sources into JSON format, enabling consistent processing across datasets ❷. NER was applied to assign unique identifiers, following a hierarchical approach: DrugBank, ChEBI, and active ingredient-based IDs. Disease entities were mapped using both DO and Orphanet to maximize coverage of rare conditions ❸. The resulting KG was implemented in Neo4j, facilitating efficient traversal and sophisticated querying of complex biomedical relationships ❹. We developed this framework in Neo4j, enabling interactive visualization through its browser interface. The graph represents drugs as primary nodes connected via semantic relationships (*has_indication*, *has_contraindication*, *has_administration*, *is_approved*) to disease nodes, administration routes, and approval years.

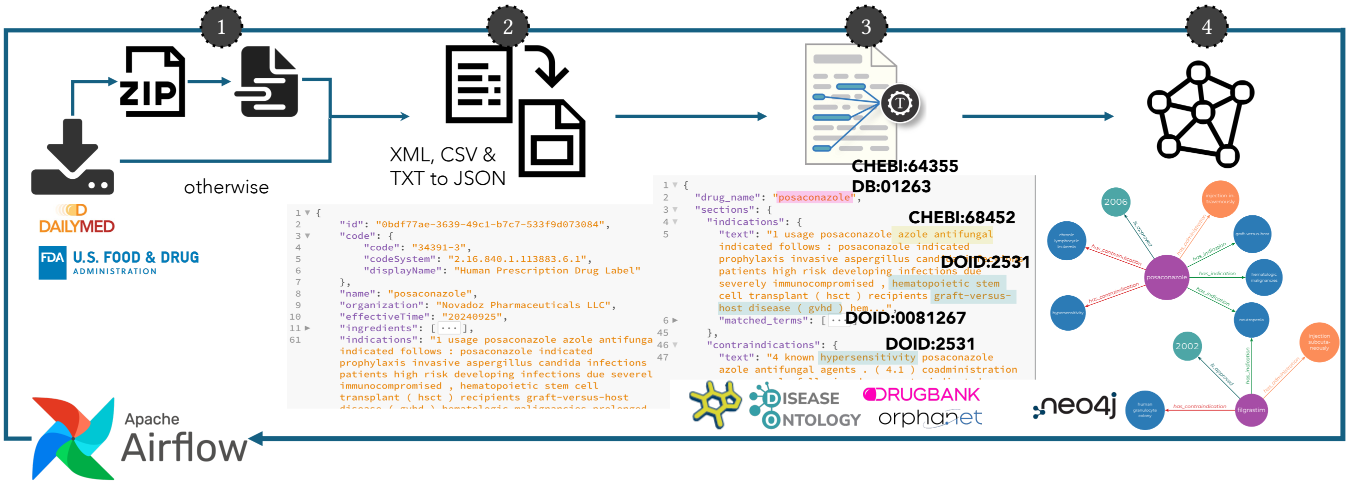

Figure 1: Overview of the workflow. The knowledge graph integrates drugs, indications, and contraindications from DailyMed, Orange and Purple Books databases. Apache Airflow orchestrates monthly updates, while NER identifies drugs via ChEBI and DrugBank, and diseases via Disease Ontology and Orphanet.

3 Visualization and Analytical Capabilities

The resulting KG enables multiple analytical dimensions that would be difficult to perceive through traditional tabular analysis: (1) **Drug-Disease Relationship Mapping**: Our KG contains 737 nodes (561 drugs and 176 diseases) connected through 1,037 relationships (539 therapeutic indications and 498 contraindications). The visualization reveals therapeutic clusters and regulatory patterns that highlight important pharmaceutical connections. (2) **Temporal Analysis of Drug Approvals**: The bipartite graph structure comprising 947 nodes (894 drugs and 53 approval years) connected by 1,047 relationships reveals the evolution of regulatory patterns over time, with notable acceleration in recent years. (3) **Administration Route Analysis**: The graph-based representation of 329 drug nodes and 20 administration route nodes connected by 344 relationships reveals hierarchical patterns in drug delivery methods, highlighting both common approaches and specialized routes.

4 Potential for Educational Impact

While our current work focuses on research applications, the visual and interactive nature of our KG system demonstrates significant potential for transforming how pharmaceutical knowledge could be presented in educational contexts. The intuitive representation of complex relationships aligns with how people naturally process information, suggesting promising applications for enhancing understanding of biomedical concepts.

5 Conclusion and Future Works

This work demonstrates how computer science innovations – specifically KG technology – can transform biomedical data exploration. By converting abstract pharmaceutical knowledge into interactive visual representations, we enable more intuitive analysis that better reflects the interconnected nature of biomedical domains. Computer science serves as a catalyst for research advancement by providing tools that align with how researchers naturally conceptualize complex information, making KGs a powerful technology for exploring intricate biomedical relationships. We plan to extend this research in several directions: (1) Incorporating additional biomedical data sources including clinical trial results and pharmacovigilance data; (2) Implementing advanced graph algorithms for identifying novel drug interaction patterns; and (3) Developing more intuitive interactive visualization interfaces to support exploratory research.

Acknowledgments

This work is supported by UID/04516/NOVA Laboratory for Computer Science and Informatics (NOVA LINCS), UID/BIO/00645/IBEB, UIDB/00408/LASIGE and UIDP/00408/LASIGE with the financial support of FCT.IP.

References

[1] David N. Nicholson and Casey S. Greene. 2020. Constructing knowledge graphs and their biomedical applications. *Computational and Structural Biotechnology Journal* 18 (2020), 1414–1428. https://doi.org/10.1016/j.csbj.2020.05.017

[2] Ana Carolina Pereira, Matilde Pato, and Nuno Datia. 2025. Mapping Drug Interactions and Therapeutic Clusters through Visual Graph Analytics. (2025), Accepted in 29th International Conference Information Visualisation, Darmstadt, Germany

[3] Craig Knox, Mike Wilson, Christen M Klinger, Mark Franklin, Eponine Oler, Alex Wilson, Allison Pon, Jordan Cox, Na Eun Chin, Seth A Strawbridge, et al. 2023. DrugBank 6.0: the DrugBank Knowledgebase for 2024. *Nucleic Acids Research* 52, D1 (11 2023), D1265–D1275. https://doi.org/10.1093/nar/gkad976

[4] Kirill Degtyarenko, Paula de Matos, Marcus Ennis, Janna Hastings, Martin Zbinden, Alan McNaught, Rafael Alcántara, Michael Darsow, Mickaël Guedj, and Michael Ashburner. 2007. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Research* 36, suppl_1 (10 2007), D344–D350. https://doi.org/10.1093/nar/gkm791

[5] Lynn M Schriml, James B Munro, Mike Schor, Dustin Olley, Carrie McCracken, Victor Felix, J Allen Baron, Rebecca Jackson, Susan M Bello, Cynthia Bearer, Richard Lichenstein, Katharine Bisordi, Nicole Campion Dialo, Michelle Giglio, and Carol Greene. 2021. The Human Disease Ontology 2022 update. *Nucleic Acids Research* 50, D1 (11 2021), D1255–D1261. https://doi.org/10.1093/nar/gkab1063

[6] SS Weinreich, R Mangon, JJ Sikkens, M E en Teeuw, and MC Cornel. 2008. Orphanet: a European database for rare diseases. *Nederlands tijdschrift voor geneeskunde* 152, 9 (March 2008), 518—519.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009