

GutBrainKB: Exploring the Gut–Brain Interaction through a Reliable Biomedical KB

Ornella Irrera^[0000–0003–2284–5699], Marco Martinelli^[0009–0001–1596–8642],
Samuel Piron^[0009–0006–5558–4295], and Gianmaria Silvello^[0000–0003–4970–4554]

University of Padova, Italy
{name.surname}@unipd.it

Abstract. GutBrainKB is a web platform based on a structured KB to aid clinicians and researchers in studying gut–brain interactions. It features a 79K-triple KB from expert-annotated literature in the Gut-BrainIE dataset (BioASQ Lab @ CLEF 2025), offering a reliable resource for exploring the gut-brain axis. Users can explore through natural language and faceted search, investigating entity relationships, accessing scientific literature, and visualizing connections with interactive graphs. GutBrainKB combines a semantically validated KB with an accessible interface, providing an efficient tool for analyzing gut–brain interactions and their impact on neurological and psychiatric disorders.

1 Introduction

In recent years, interest in the gut–brain axis has grown steadily, as demonstrated by numerous studies linking gut microbiota to neurological disorders such as Parkinson’s and Alzheimer’s disease, multiple sclerosis and various mental health conditions [1–3]. This expanding body of literature creates the need for systematic extraction and integration of knowledge from publications to render insights machine-actionable and to reveal connections among medical concepts that might otherwise remain hidden. In this context, Information Extraction (IE) serves as the foundation for processing large volumes of biomedical text, constructing structured Knowledge Bases (KBs), and enabling downstream applications in drug discovery, disease modeling, and precision medicine. The core tasks within IE encompass Named Entity Recognition (NER) for identifying and classifying biomedical entities, Named Entity Linking (NEL) for normalizing and disambiguating those entities, and Relation Extraction (RE) for detecting and categorizing semantic relationships among them. The effectiveness of systems performing these tasks largely relies on the availability of high-quality human-annotated data, which serves as the gold standard for training, evaluation, and benchmarking. This open issue was addressed, specifically within the gut-brain interplay domain, by the GutBrainIE task, part of the BioASQ Lab at CLEF2025 [8, 11], which introduces a dataset featuring over 350 PubMed abstracts manually annotated by biomedical experts for entity mentions and relations.

To fully leverage the knowledge embedded in this collection, the extraction, interpretation, and integration of its content are crucial. This is typically

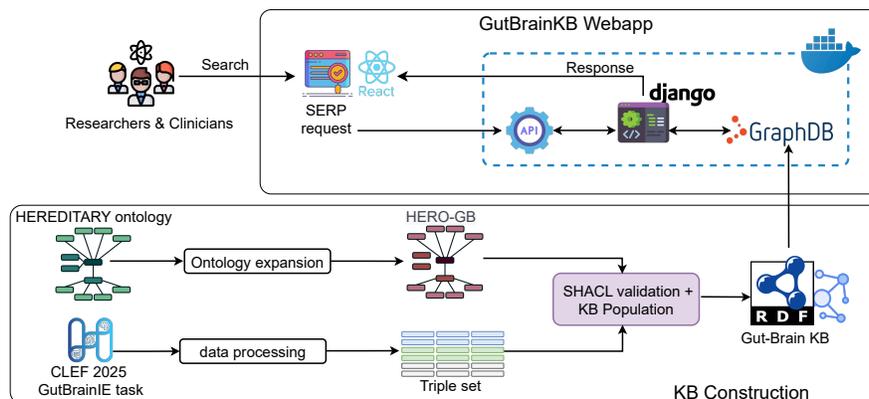


Fig. 1. Overview of the GutBrainKB system.

achieved through KBs, which model entities as nodes and relations as edges, producing a machine-readable format that supports easy access and advanced querying. Relevant examples in the literature which target gut-brain interactions are: the Pre-/Probiotics Knowledge Graph (PPKG) [6], the Food4HealthKG [4], the Microbiota Knowledge Graph for Mental Disorders (MiKG4MD) [7], and Microbiota Knowledge Graph (MiKG) [12]. Typically, such KBs are represented using RDF triples and queried via SPARQL [13], which provides powerful capabilities for data retrieval and reasoning. However, formulating SPARQL queries is challenging and requires technical expertise [5, 14].

GutBrainKB is a web platform based on a 79K-triple KB from GutBrainIE, enabling exploration of gut-brain relationships with scientific evidence. Users can query via natural language or faceted search and view interactive graphs of entity connections, aiding researchers and clinicians in understanding gut-brain interactions and tracing claims to source publications.

2 GutBrainKB

Figure 1 shows the development of the GutBrainKB system, composed of two main phases: KB construction and GutBrainKB WebApp implementation.

KB construction. To construct the KB, we relied on two resources: (i) the GutBrainIE collection [8, 11], and (ii) the HEREDITARY Gut-Brain Ontology (HERO-GB in Figure 1), which also captures the structure of the annotated corpus, modeling sentences that contain biomedical concepts.¹

We used the expert-curated annotations from the GutBrainIE collection comprising 359 abstracts, 9.9K entity mentions, and 4K relationships. From these annotated abstracts, entity mentions were linked to controlled biomedical vocabularies, thus normalizing and disambiguating textual variants. Then, we ex-

¹ <https://hereditary.dei.unipd.it/ontology/gutbrain/>

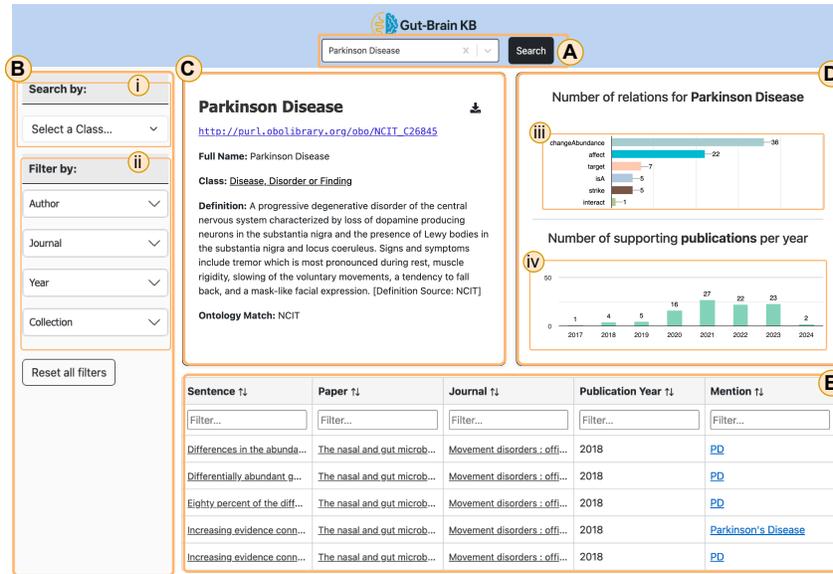


Fig. 2. GutBrainKB User Interface.

tracted RDF triples capturing both the semantic relationships among entities identified through RE annotations and the textual context in which each entity appears (e.g., sentence- and document-level evidences), resulting in a final KB of 79K triples. To ensure consistency with the underlying ontology, the resulting RDF graph has been validated with SHACL. The code and the data used for KB generation are available online.^{2 3}

Importantly, the KB construction pipeline is fully reusable and scalable: new documents from future GutBrainIE expansions can be processed using the same workflow to generate additional RDF triples, which can be seamlessly integrated into the KB. As a result, the KB can grow over time while preserving schema consistency.

GutBrainKB WebApp. The GutBrainKB WebApp adopts a three-layer architecture. At the top layer, the User Interface (UI) is built with React.js, providing users with an interactive front-end for querying and interacting with the KB; a back-end layer, implemented using the Python web framework Django, is responsible for business logic, REST APIs, and service orchestration; the data layer is a GraphDB instance used to store and manage RDF triples of the generated KB. The system is fully dockerized with each component running in a separate container to simplify deployment and maintainability.

² <https://github.com/GutBrainKB/GutBrainKB-WebApp/>

³ <https://zenodo.org/records/16845409>

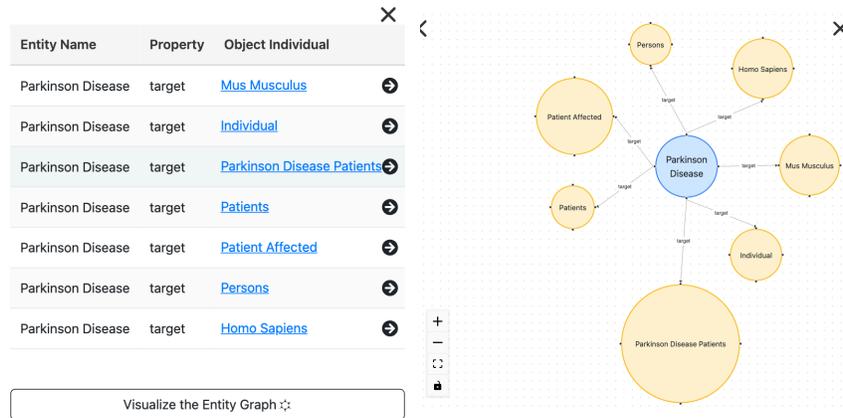


Fig. 3. The panel on the left reports the list of relationships for the subject *Parkinson disease* and predicate *target*. On the right, the related Entity Graph.

The GutBrainKB platform allows users to search for factual information on gut–brain relations supported by scientific literature. Users can access concept definitions and explore relationships with their supporting textual and bibliographic evidences. Specifically, when a query such as “Alzheimer’s disease” is submitted, the backend first sanitizes the input and then executes a predefined SPARQL query template, injecting the term into a case-insensitive filter to retrieve all matching entity mentions. For each matched entity, the system returns associated paper metadata (e.g., PMID, title, authors, journal, year), supporting evidence (sentences and mention texts), linked ontology class or individual information, and the relations in which the entity participates. The results are then deduplicated at the paper–sentence–mention level and packaged into a compact JSON format for the UI. The reliability of the results returned to the user directly is related to the quality of the GutBrainIE annotations underlying the KB. Entity linking achieved an overall accuracy of 0.915 ± 0.0473 [10]. The quality of relation annotations is further ensured by relying on the two highest-quality folds of GutBrainIE, curated by expert annotators with domain knowledge and prior annotation experience [9].

In Figure 2, we present the GutBrainKB UI showing the **Parkinson Disease** concept. (A) provides an autocomplete search bar, which lets users enter queries for specific concepts and their relationships. (B) features a faceted search panel, allowing users to filter concepts by ontological class or named individual and to narrow publication results by author, journal, collection, or year. (C) displays detailed information about the selected concept, such as its ontology reference, name, class, definition, and vocabulary. Users may also download a summary of the concept, including its relationships and supporting sentences, in either JSON or TTL format. (D) summarizes the relationships associated with the selected

concept and shows how many publications mention it. A horizontal histogram visualizes relationship distribution by predicate, and users can click any bar to see all relationships for the chosen predicate in a modal window. Each listed object links to its own description page, enabling further exploration within the KB. A graph view is also available, visualizing the selected concept and its related entities. In Figure 3, the left panel shows the list of relationships with *Parkinson disease* as subject and *target* as predicate, while the right presents an overview of the generated graph. Another histogram shows the number of related publications per year and allows users to view yearly publication details in a modal window. (E) presents a table that summarizes the sentences where the concept appears. Users can interact with each table cell to access additional information, such as the full text of the article and journal details.

The GutBrainKB WebApp is available at <https://hereditary.dei.unipd.it/app/gutbrainkb/>.

3 Final remarks

In this work, we introduced GutBrainKB, a web application for exploring validated gut–brain facts within a biomedical KB. We built the KB from the expert-annotated GutBrainIE collection and ensured semantic consistency by validating against HERO-GB. GutBrainKB combines curated data, semantic modeling, and interactive visualizations to help clinicians and researchers query and interpret complex gut–brain relationships.

Future directions include the integration of large language models to assist in translating users’ natural language queries into SPARQL queries over the KB.

Acknowledgments

The work was supported by the HEREDITARY project, as part of the EU Horizon Europe program under Grant Agreement 101137074.

Discolsure of interests

The authors have no competing interests to declare that are relevant to the content of this article.

Bibliography

- [1] Appleton, J.: The gut-brain axis: influence of microbiota on mood and mental health. *Integrative Medicine: A Clinician's Journal* **17**(4), 28 (2018)
- [2] Carabotti, M., Scirocco, A., Maselli, M.A., Severi, C.: The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. *Annals of gastroenterology: quarterly publication of the Hellenic Society of Gastroenterology* **28**(2), 203 (2015)
- [3] Cryan, J.F., O'Riordan, K.J., Sandhu, K., Peterson, V., Dinan, T.G.: The gut microbiome in neurological disorders. *The Lancet Neurology* **19**(2), 179–194 (2020)
- [4] Fu, C., Huang, Z., van Harmelen, F., He, T., Jiang, X.: Food4healthkg: Knowledge graphs for food recommendations based on gut microbiota and mental health. *Artificial Intelligence in Medicine* **145**, 102677 (2023). <https://doi.org/https://doi.org/10.1016/j.artmed.2023.102677>, <https://doi.org/10.1016/j.artmed.2023.102677>
- [5] Giachelle, F., Marchesin, S., Silvello, G., Alonso, O.: Searching for reliable facts over a medical knowledge base. In: Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval. pp. 3205–3209 (2023). <https://doi.org/https://doi.org/10.1145/3539618.3591822>, <https://doi.org/10.1145/3539618.3591822>
- [6] Liu, T., Lan, G., Feenstra, K.A., Huang, Z., Heringa, J.: Towards a knowledge graph for pre-/probiotics and microbiota–gut–brain axis diseases. *Scientific Reports* **12**(1), 18977 (2022). <https://doi.org/https://doi.org/10.1038/s41598-022-21735-x>, <https://doi.org/10.1038/s41598-022-21735-x>
- [7] Liu, T., Pan, X., Wang, X., Feenstra, K.A., Heringa, J., Huang, Z.: Predicting the relationships between gut microbiota and mental disorders with knowledge graphs. *Health information science and systems* **9**(1), 3 (2020). <https://doi.org/https://doi.org/10.1007/s13755-020-00128-2>, <https://doi.org/10.1007/s13755-020-00128-2>
- [8] Martinelli, M., Silvello, G., Bonato, V., Di Nunzio, G.M., Ferro, N., Irrera, O., Marchesin, S., Menotti, L., Vezzani, F.: Overview of GutBrainIE@CLEF 2025: Gut-Brain Interplay Information Extraction. In: Faggioli, G., Ferro, N., Rosso, P., Spina, D. (eds.) CLEF 2025 Working Notes (2025)
- [9] Martinelli, M., Silvello, G., Bonato, V., Di Nunzio, G., Ferro, N., Irrera, O., Marchesin, S., Menotti, L., Vezzani, F., et al.: Overview of gutbrainie@ clef 2025: Gut-brain interplay information extraction. In: Faggioli, G., Ferro, N., Rosso, P., Spina, D. (eds.) CLEF 2025 Working Notes. vol. 4038, pp. 65–98 (2025)
- [10] Martinelli, M., Marchesin, S., Silvello, G.: Efficient and reliable estimation of named entity linking quality: A case study on gutbrainie (2026), <https://arxiv.org/abs/2601.06624>

- [11] Nentidis, A., Katsimpras, G., Krithara, A., Krallinger, M., Rodríguez-Ortega, M., Rodríguez-López, E., Loukachevitch, N., Sakhovskiy, A., Tutubalina, E., Dimitriadis, D., Tsoumakas, G., Giannakoulas, G., Bekiaridou, A., Samaras, A., Di Nunzio, G.M., Ferro, N., Marchesin, S., Martinelli, M., Silvello, G., Paliouras, G.: Overview of BioASQ 2025: The thirteenth BioASQ challenge on large-scale biomedical semantic indexing and question answering. *Lecture Notes in Computer Science*, vol. TBA, p. TBA. Springer (2025)
- [12] Pudavar, A.E., Baksi, K.D., Pokhrel, V., Kuntal, B.K.: Microbiome knowledge graph as a tool to understand bacteria-host associations. *Archives of Microbiology* **207**(9), 222 (2025). <https://doi.org/10.1007/s00203-025-04413-0>, <https://doi.org/10.1007/s00203-025-04413-0>
- [13] Weikum, G., Dong, X.L., Razniewski, S., Suchanek, F.M.: Machine knowledge: Creation and curation of comprehensive knowledge bases. *Found. Trends Databases* **10**(2-4), 108–490 (2021). <https://doi.org/10.1561/19000000064>, <https://doi.org/10.1561/19000000064>
- [14] Wu, W.: Proactive natural language search engine: tapping into structured data on the web. In: *Proceedings of the 16th International Conference on Extending Database Technology*. pp. 143–148 (2013). <https://doi.org/https://doi.org/10.1145/2452376.2452394>, <https://doi.org/10.1145/2452376.2452394>