Visual Exploration of Ontologies Supported by Language Models and Interactive Lenses

Benedikt Kantz
benedikt.kantz@tugraz.at
Graz University of Technology
Graz, Austria

Peter Waldert
peter.waldert@tugraz.at
Graz University of Technology
Graz, Austria

Stefan Lengauer
s.lengauer@tugraz.at
Graz University of Technology
Graz, Austria

Tobias Schreck
tobias.schreck@tugraz.at
Graz University of Technology
Graz, Austria

Abstract

Ontologies contain semantic relationships and dependencies for, either, ontological studies of a subject or as a blueprint for connected data within Knowledge Graphs (KGs). The ontologies themselves, however, can become difficult to comprehend as specialized class diagrams or complete views. We present an improved interactive visualization for ontology exploration using circle packed hierarchical views within our Ontology and Semantic Exploration Toolkit (OnSET) as the base layer, with interactive visual lenses. The circle packed visualization is enriched in two novel ways incorporating Language Models (LMs): first, by linking sparse datasets to the ontology using semantic matching. The second enrichment is performed by employing topic modelling on the ontology and its connection to find groups of topics that cluster the ontology in a hierarchical manner.

CCS Concepts

• **Information systems** → **Ontologies**; • **Human-centered computing** → *Information visualization*; Visualization theory, concepts and paradigms.

ACM Reference Format:
Benedikt Kantz, Peter Waldert, Stefan Lengauer, and Tobias Schreck. 2025. Visual Exploration of Ontologies Supported by Language Models and Interactive Lenses. In *Proceedings of the 18th International Symposium on Visual Information Communication and Interaction (VINCI 2025), December 01–03, 2025, Linz, Austria.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3769534.3769538

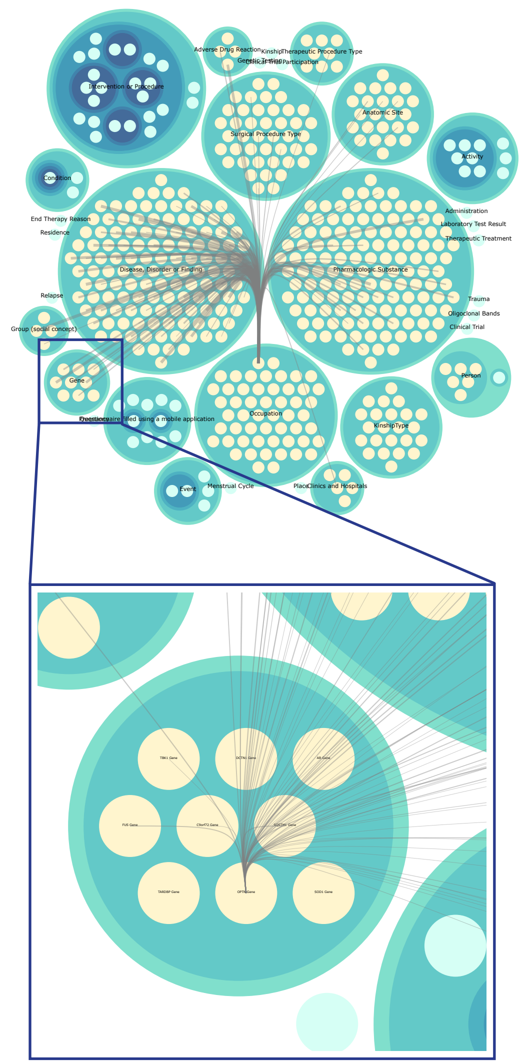


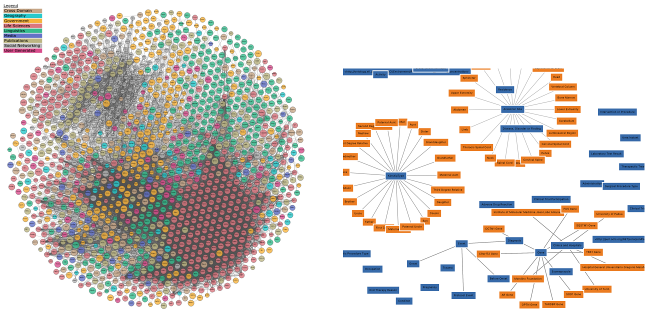
Figure 1: Different zoom levels of the linked view in OnSET enable the discovery of related attributes in a sparse dataset.

1 Introduction

Ontologies are widely used across many domains [1], requiring an understanding of them – especially for the unsupervised acquisition of ontologies. The OnSET enables a **two-step exploration of arbitrary ontologies**. It offers two avenues to explore the ontologies, first by offering a topic overview for arbitrary ontologies using a LM. The second, related visualization approach **connects**

VINCI 2025, Linz, Austria

ACM ISBN 979-8-4007-1845-8/25/12
https://doi.org/10.1145/3769534.3769538



(a) Tightly packed ontology [7], emphasizing the need for detail on demand and zooming.

(b) Overflowing Graph, illustrating the need for more a more effective and global view.

Figure 2: Different Existing Ontology Visualizations

high-dimensional sparse datasets to the ontology based on the column descriptions and a greedy matching strategy.

These two approaches enable users with little to no experience in ontologies to tackle the following tasks:

(1) Gather an overview of arbitrary ontologies for users not familiar with the usual notation of ontologies using exploration-driven discovery,
(2) the relation (co-occurrence) of specific instances, their properties, and classes given a sparse dataset,
(3) and the discovery of topical concepts hidden within rich ontologies.

While this tool targets non-experts as the user base, we also intend the tool to be used as an initial overview of ontologies and as a point of entry for more experienced user to discover the depth of the ontology they are analyzing. This exploration and analysis are enabled through the novel combination of LMs with visual ontology analysis, allowing for fuzzy matching of components and topic modeling in our case. While OnSET is capable of supporting any ontology and sparse dataset combination, we use the Brainteaser Ontology (BTO) [11] and the GutbrainIE ontology [3] for testing and evaluating the tool. The ontology relates attributes of patients with various diseases and diagnoses, as well as clinical procedures.

The sparse dataset serving as the basis for the linked visualization is the related BRAINTEASER Amyotrophic lateral sclerosis (ALS) and Multiple sclerosis (MS) dataset [10], which can also be replaced with another sparse dataset related to an arbitrary loaded ontology. This relation to the data is achieved using text-based matching of the dataset metadata and a textual representation of the ontology.

This **connection of ontology (visualization) with LMs** represents the core novelty and contribution of our work, showing how these **rigid systems** can be explored using the **fuzzy tools** of state-of-the-art natural language processing.

2 Related Works

Principal visualization and interaction techniques. Ontologies and graphs are usually visualized using different graph placement strategies like manual placement, tight packing, or predetermined shapes like circles [9], as shown in Figure 2. *Hierarchical Visualizations* [21] enable an exploratory and hierarchical view of the ontologies. This approach can, compared to traditional network visualizations, operate intuitively on multiple scales. The technique can therefore provide users with a comprehensive overview of the data and structure, while providing details on demand. The circle packing, although not as efficient in utilizing space as other packing strategies, still allows for additional relational visualization on top, as the circle can be seen as the backdrop for the links among the nodes represented as circles. This packing algorithm can also be used for three-dimensional visualization to emphasize the depth of relations, at the cost of reducing the possibilities for labeling the nodes. Scheibel et al. [18] further compare the distinctions of hierarchical visualization algorithms, particularly the choice of layout algorithms that utilize latent similarity measures, effectively creating a topology. *Nmap* [8] and *IsoMatch* [12] are two such examples for encoding a spatial layout; derived from feature vectors or bag-of-word approaches from Natural Language Processing (NLP). These systems, however, do not inherently support deep hierarchical visualizations, which are required for ontology visualizations. *Software Landscapes* [2] tackles these requirements for hierarchical visualization in an adjacent field, software system visualization. They employ a landscape view with connected and enclosed spheres, connected through trees. Another approach to alter the view is *Edge Bundling* [14], which provides a continuous interpolation of the topic view in a two-dimensional space between the shown high-level view and a possible low-level view.

Ontology visualization approaches Dudáš, et al. [9] identify four core use cases for visualizing ontologies: *editing, inspecting, learning*, and *sharing*. *Editing* and *inspecting* focus on the creation of ontologies, where users, typically, experts, visually build and debug ontologies. These systems provide a more comprehensive view of the ontologies, featuring detailed diagrams of their structure and exact constraints. Our use case focuses on the *learning* and *sharing aspects*, catering to non-expert users exploring ontologies, drawing inspiration from the concepts of exploratory search systems [22]. Existing systems already experiment with similar views and connections [4, 6] where they connect parts of the ontology using linkages. Our system, in comparison, uses related data to create these links from sparse data and the descriptions within the ontology in natural language. These descriptions also inform the topic modelling in natural language to further the ontology analysis and exploration. We also focus on a global view, while providing details on demand.

3 Methodology

Our visualization approach requires preprocessing of the ontology and related sparse data using LMs to discover the relations and topics present in the embedding space within the textual representation of the ontology.

3.1 Preprocessing

Our preprocessing pipeline follows a ETL-pipeline schema [16], a schema intended to streamline data processing, shown in Figure 3. The pipeline first extracts data from the CSV datasets by merging the tables, given a collection of them, and extracts classes and relations from the ontology. Next, we transform the columns of the dataset and the textual representation of the ontology elements

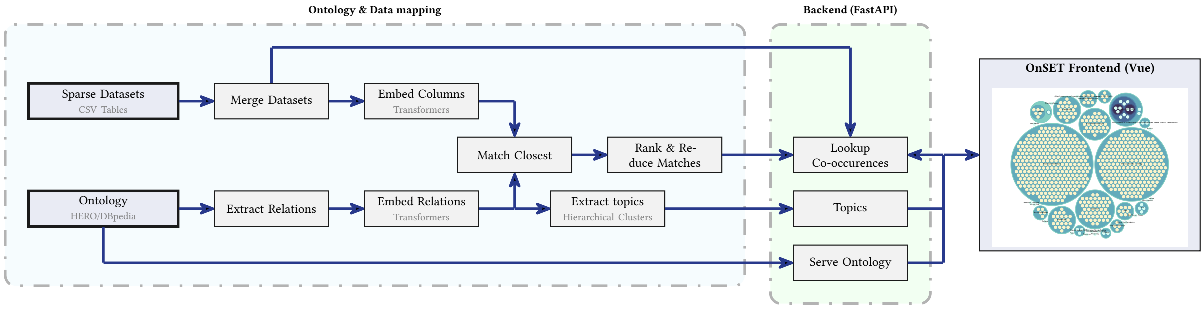


Figure 3: The OnSET semantic discovery architecture. We employ an Extraction, Transform and Load (ETL)-like system to first gather the data from the CSVs and Ontology, transform it using an embedding model and finally load it using topic modelling and matching techniques.

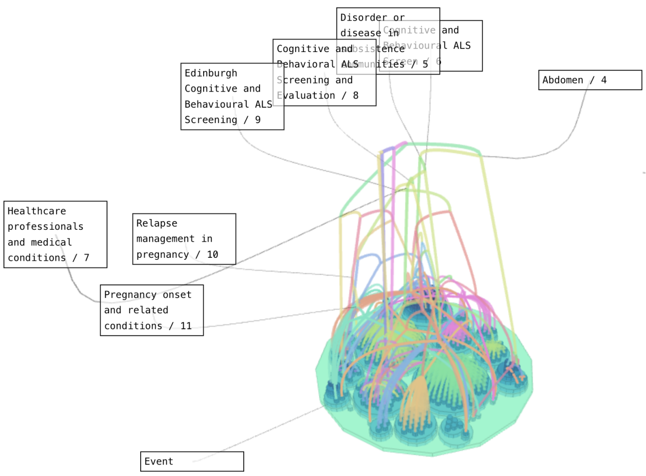


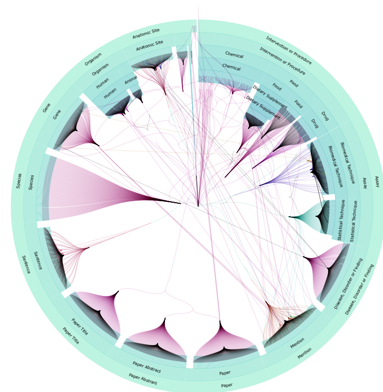
Figure 4: Topic view of BTO within OnSET with parent topic labels of the class "Event".

into a suitable embedding space using a LM, more specifically, the `sentence-transformers` packages developed for similarity-search applications [17]. These are then matched to their closest pair in cosine distance using a greedy, iterative algorithm that searches for the best distance between the ontology and table columns in a pairwise manner and reduces these candidates. The embeddings generated from the text representation are also used to build a hierarchical topic model of the ontology using BERTopic [13], which provides an additional level of linkage over the ontology.

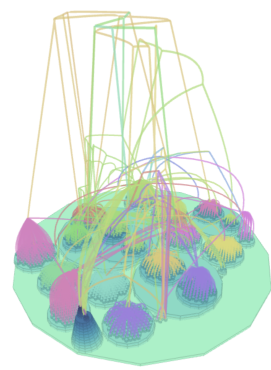
3.2 Linked Sparse Connections

Our visualization technique takes arbitrary ontologies and maps the classes, subclasses and named instances into an overview using a hierarchical representation of the ontology and the circle-packing algorithm [21]. This mapping enables the overview given in Figure 1. The user can, based on their interest in specific areas, click on any level within the circle packing to zoom and pan, revealing more details about the individual subclasses and instances of a class.

The visualization, furthermore, uses related sparse data sources and maps the data to the ontologies, providing co-occurrences and relations across the ontology graph. These co-occurrences are then shown on demand by clicking on a specific node. These links inform the user at both the overview level and in the detail views, as shown in the zoomed view in Figure 1. The overview level provides a top-level estimation of co-occurrences across the entire ontology. This



(a) Topic tree view with the classes as a sunburst diagram and the hierarchical topics as edges between them.



(b) Topic tree view within OnSET in three dimensions.

Figure 5: Topic tree view of the GutbrainIE ontology [3] within OnSET with two different visualizations.

enables the user to inform themselves about related concepts within the data. The detailed link view shown fosters this learning within the selected levels. This is helpful if the relations occur locally and might be missed at the global scale.

3.3 Topic View

Our system provides a linked view, shown in Figure 4, to incorporate the topics modeled in the preprocessing steps using a three-dimensional view. The decision to add an additional dimension to the visualization stems from the need to visualize a greater number of overlapping and intersecting connections within the ontology, compared to the simple linked connections presented above. The links are derived from the hierarchical topic model [13], based on the classes and named instances present in the ontology. This view incorporates the topics modeled within a three-dimensional view, which can be advantageous for network visualization and terrain metaphors [15].

Our visualization approach utilizes the HSV spectrum to convey the relationship between the topics. The assignment of colors is based on splitting the color (hue) part of each tree branch by the number of nodes present in a tree, splitting the color. The resulting link colors indicate, to some extent, how close these topics are in relation to each other, illustrated through Figures 4 and 5b. We also place labels indicating nodes and parent topics radially outside the tree visualization when hovering over them. The placement of labels within the circle packing and tree visualization would occlude relevant parts of the visualization, already noted in prior works highlighting the difficulties of three-dimensional visualization and label placement [20, 21].

We provide a comparison to our topic trees with an inverted version of the Sunburst diagram [19] with the linked edges within the circle in Figure 5a.

4 Implementation

The described visual analytics systems are achieved using a three-step system architecture illustrated in Figure 3. This system is based upon an ETL pipeline, where we first extract the datasets and ontologies and match them to each other as described in Section 3.1. These reduced matches are then loaded into a database and served, along with the ontology itself, on our *FastAPI*[1] backend and visualized using our *Vue.js*[2] frontend with the help of *D3.js* [5] and *three.js*[3]. The code is available at https://github.com/hereditary-eu/OnSET.

5 Use Case

Our approach enables workflows for users who have acquired an ontology for a new domain of interest and would like to explore the hierarchical relationships and covered topics. Ontologies can be loaded into OnSET and explored, i.e., examining concepts of interest and verifying whether they align with those in the ontology. If the researcher additionally gathers related tabular data, the tool can load it and match it to the ontology. This connection of schematic data and tabular data allows the researcher to search for relations within their data.

[1]https://fastapi.tiangolo.com/
[2]https://vuejs.org/
[3]https://threejs.org/

In the case of the BTO, the user loads the ontology and the related sparse ALS and MS dataset. After the user clicks on a specific occupation, as shown in Figure 1, links to different other instances are shown – there seems to be a high link count (indicated by width) to a particular group, a few specific diseases, disorders, or findings, and adverse drug reactions. The user could then hover over the linked circles to discover which ones are affected and check, by clicking on them, how they might be linked to other instances, indicating correlations or higher-order dependencies. If one link is very localized, for example, in the cropped view for the Gene, the user can investigate these small-scale localized links by clicking on them. The user can also view the topic map if there is no associated dataset, or if the exploration is targeted towards the ontology itself. The user can, using the topic view, first get a grasp of the distribution of topics within the circle view. Hovering over individual circles, representing classes, as shown in Figure 4, reveals the class and parent topic labels, thereby organizing the knowledge. Comparing this view to the two-dimensional view of the sunburst diagram in Figure 5a reveals why a third dimension is helpful for this double-hierarchical view, as the user cannot accurately pinpoint the individual instances, and the structure of topics is not as easily distinguishable.

6 Discussion

The proposed system **combines NLP in the form of LMs with ontology visualization**, arriving at novel exploration strategies and views which should foster the **understanding of non-experts through visual lens systems and hierarchical approaches**, enabling a coarse-to-fine view. This intersection enhances the visualization and reveals latent information hidden within ontologies, thereby enriching the familiarization process for non-experts in the field of building and conceptualizing ontologies. Interactive views, especially lenses, facilitate this introductory setting, allowing users to choose how to focus and explore different aspects of the ontology at their own pace, with a global view at its core.

There are, however, opportunities evolving from related works to enrich our visualization and choice of packing algorithms.

First, the approach to lay out the basis of our visualization, circle packing, could be enriched using spatially aware algorithms like *Nmap* [8] and *IsoMatch* [12]. These could reduce the spatial disarray found in some linked views and reduce the entanglement of the topic maps. Our visualization system, furthermore, only considers the class hierarchies, foregoing the richness that some ontologies offer in the properties linking classes within the ontology [9]. *Edge Bundles* [14] could, furthermore, provide a continuous interpolation of the topic view in the two-dimensional view between the shown high-level and a possible low-level view as a further point of comparison. These improvements and comparisons could, furthermore, be studied in comparative evaluations to compare the legibility of the topic map and visual links, potentially reducing the visual load placed upon the users.

Acknowledgments

This work is partially supported by the HEREDITARY Project, as part of the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101137074.

References

[1] Muhammad Nabeel Asim, Muhammad Wasim, Muhammad Usman Ghani Khan, Waqar Mahmood, and Hafiza Mahnoor Abbasi. 2018. A survey of ontology learning techniques and applications. *Database* 2018 (10 2018), bay101. doi:10.1093/database/bay101

[2] Michael Balzer, Andreas Noack, Oliver Deussen, and Claus Lewerentz. 2004. Software Landscapes: Visualizing the Structure of Large Software Systems. In *Eurographics / IEEE VGTC Symposium on Visualization*, Oliver Deussen, Charles Hansen, Daniel Keim, and Dietmar Saupe (Eds.). The Eurographics Association. doi:/10.2312/VisSym/VisSym04/261-266

[3] Vanessa Bonato, Giorgio Maria Di Nunzio, Nicola Ferro, Ornella Irrera, Stefano Marchesin, Marco Martinelli, Laura Menotti, Gianmaria Silvello, and Federica Vezzani. 2025. GutBrainIE @ CLEF 2025. https://hereditary.dei.unipd.it/challenges/gutbrainie/2025/ [Online; accessed 2025-06-09].

[4] A. Bosca and D. Bonino. 2006. OntoSphere3D: A Multidimensional Visualization Tool for Ontologies. In *17th International Workshop on Database and Expert Systems Applications (DEXA'06)*. 339–343. doi:10.1109/DEXA.2006.106

[5] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D³ Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2301–2309. doi:10.1109/TVCG.2011.185

[6] Isabel Cristina Siqueira da Silva, Giuseppe Santucci, Carla del Sasso Freitas, et al. 2012. Ontology Visualization: One Size Does Not Fit All.. In *EuroVA@ EuroVis*.

[7] Javier de la Cueva and Eva Méndez. 2022. *Open Science and Intellectual Property Rights. How Can They Better Interact? State of the Art and Reflections. Report of Study. European Commission.* European Commission, Brussels.

[8] Felipe S. L. G. Duarte, Fabio Sikansi, Francisco M. Fatore, Samuel G. Fadel, and Fernando V. Paulovich. 2014. Nmap: A Novel Neighborhood Preservation Space-filling Algorithm. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (Dec. 2014), 2063–2071. doi:10.1109/tvcg.2014.2346276

[9] Marek Dudáš, Steffen Lohmann, Vojtěch Svátek, and Dmitry Pavlov. 2018. Ontology visualization methods and tools: a survey of the state of the art. *The Knowledge Engineering Review* 33 (2018), e10. doi:10.1017/S0269888918000073

[10] Guglielmo Faggioli, Stefano Marchesin, Laura Menotti, Helena Aidos, Roberto Bergamaschi, Giovanni Birolo, Pietro Bosoni, Paola Cavalla, Adriano Chiò, Arianna Dagliati, Mamede de Carvalho, Giorgio Maria Di Nunzio, Piero Fariselli, Jose Manuel García Dominguez, Marta Gromicho, Alessandro Guazzo, Enrico Longato, Sara C. Madeira, Umberto Manera, Gianmaria Silvello, Eleonora Tavazzi, Erica Tavazzi, Isotta Trescato, Martina Vettoretti, Barbara Di Camillo, and Nicola Ferro. 2024. BRAINTEASER ALS and MS Datasets. doi:10.5281/ZENODO.12789962

[11] Guglielmo Faggioli, Stefano Marchesin, Laura Menotti, Giorgio Maria Di Nunzio, Gianmaria Silvello, and Nicola Ferro. 2024. The BrainTeaser Ontology for ALS and MS Clinical Data. doi:10.5281/ZENODO.12789731

[12] O. Fried, S. DiVerdi, M. Halber, E. Sizikova, and A. Finkelstein. 2015. IsoMatch: Creating Informative Grid Layouts. *Computer Graphics Forum* 34, 2 (2015), 155–166. doi:10.1111/cgf.12549 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.12549

[13] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794* (2022).

[14] D. Holten. 2006. Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data. *IEEE Transactions on Visualization and Computer Graphics* 12, 5 (Sept. 2006), 741–748. doi:10.1109/tvcg.2006.147

[15] Kim Marriott, Jian Chen, Marcel Hlawatsch, Takayuki Itoh, Miguel A. Nacenta, Guido Reina, and Wolfgang Stuerzlinger. 2018. *Immersive Analytics: Time to Reconsider the Value of 3D for Information Visualisation.* Springer International Publishing, Cham, 25–55. doi:10.1007/978-3-030-01388-2_2

[16] Aiswarya Raj, Jan Bosch, Helena Holmström Olsson, and Tian J. Wang. 2020. Modelling Data Pipelines. In *2020 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. 13–20. doi:10.1109/SEAA51224.2020.00014

[17] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. https://arxiv.org/abs/1908.10084

[18] Willy Scheibel, Daniel Limberger, and Jürgen Döllner. 2020. Survey of treemap layout algorithms. In *Proceedings of the 13th International Symposium on Visual Information Communication and Interaction* (Eindhoven, Netherlands) *(VINCI '20)*. Association for Computing Machinery, New York, NY, USA, Article 1, 9 pages. doi:10.1145/3430036.3430041

[19] John Stasko, Richard Catrambone, Mark Guzdial, and Kevin McDonald. 2000. An evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human-Computer Studies* 53, 5 (Nov. 2000), 663–694. doi:10.1006/ijhc.2000.0420

[20] Thierry Stein and Xavier Décoret. 2008. Dynamic label placement for improved interactive exploration. In *Proceedings of the 6th International Symposium on Non-Photorealistic Animation and Rendering* (Annecy, France) *(NPAR '08)*. Association for Computing Machinery, New York, NY, USA, 15–21. doi:10.1145/1377980.1377986

[21] Weixin Wang, Hui Wang, Guozhong Dai, and Hongan Wang. 2006. Visualization of large hierarchical data by circle packing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI '06)*. Association for Computing Machinery, New York, NY, USA, 517–520. doi:10.1145/1124772.1124851

[22] Ryen W. White and Resa A. Roth. 2009. *Exploratory Search: Beyond the Query—Response Paradigm.* Springer International Publishing. doi:10.1007/978-3-031-02260-9

Received 13.6. 2025