

KnowTeX: Visualizing Mathematical Dependencies

Elif Uskuplu

Lawrence S. Moss

Valeria de Paiva

January 20, 2026

Abstract

Mathematical knowledge exists in many forms, ranging from informal textbooks and lecture notes to large formal proof libraries, yet moving between these representations remains difficult. Informal texts hide dependencies, while formal systems expose every detail in ways that are not always human-readable. Dependency graphs offer a middle ground by making visible the structure of results, definitions, and proofs. We present KnowTeX, a standalone, user-friendly tool that extends the ideas of Lean’s Blueprints, enabling the visualization of conceptual dependencies directly from L^AT_EX sources. Using a simple `\uses` command, KnowTeX extracts relationships among statements and generates previewable graphs in DOT and TikZ formats. Applied to mathematical texts, such graphs clarify core results, support education and formalization, and provide a resource for aligning informal and formal mathematical representations. We argue that dependency graphs should become a standard feature of mathematical writing, benefiting both human readers and automated systems.

1 Introduction

Mathematical writing is inherently complex. Definitions, theorems, and proofs form dense networks of interdependent ideas, where results often build on earlier ones in subtle and intertwined ways. In textbooks and lecture notes, these connections are spread across multiple chapters and sections, making it difficult to perceive the overall structure of the knowledge being presented. The richness of mathematical discourse, while essential for depth and rigor, also creates challenges for navigation, comprehension, and reuse of information.

In contrast, formal proof assistants such as Lean ([MKA⁺15]), Agda ([Nor09]), or Rocq¹ ([BC10]) make every dependency explicit. When a mathematical theory is formalized, each proof must declare precisely which results it uses, since omitting even a single dependency would render the proof invalid. These systems, therefore, expose the finest level of detail in mathematical relationships, capturing the complete web of definitions, lemmas, and theorems in code. However, this level of detail—while powerful for verification—can be overwhelming for humans, as it demands close familiarity with the entire formal environment.

Between the implicit structure of informal texts and the exhaustive precision of formal systems lies a useful middle ground: dependency graphs. A dependency graph represents mathematical statements as nodes and their conceptual or logical relationships as edges, providing a readable overview of how results depend on one another. Such visualizations make the structure of a text more transparent without requiring formalization, offering both researchers and educators a practical way to explore, analyze, and align mathematical knowledge.

Some tools have already explored ways to generate dependency graphs from mathematical documents. One notable example is Patrick Massot’s ‘Lean Blueprint’²) system, which extends the concept of dependency between components of a proof by linking informal mathematical text with other corresponding formal proofs in the Lean theorem prover. Another similar tool is plasTeXdepgraph³, a simplified version of the Lean blueprint, a plugin for plasTeX. This simply removes the connection to the Lean prover, still linking informal mathematical text in plasTeX to interactive dependency graphs. Together, these tools demonstrate the value of dependency graphs for understanding mathematical structure, though they remain closely tied to specific technical ecosystems.

¹Formerly known as Coq.

²Code in <https://github.com/PatrickMassot/leanblueprint>

³See <https://github.com/PatrickMassot/plastexdepgraph>

Another related approach is the recently-introduced ‘Trouver’ system⁴, which applies machine learning to automatically infer dependency structures from L^AT_EX documents. Unlike tools that rely on explicit markup commands, Trouver allows users to simply upload a L^AT_EX file, and it generates a dependency graph using statistical and linguistic models. This approach removes the manual effort required to annotate dependencies, offering a convenient and automated workflow. However, its effectiveness depends strongly on the clarity of the source text: when relationships are implicit or linguistically complex, the model may miss important links or produce spurious ones, resulting in overly dense or incomplete graphs. Trouver thus represents a promising yet imperfect data-driven alternative, practical for quick visualization but less reliable for precise structural analysis. There is also a recent commercial service ‘ScienceStack’⁵ that provides HTML rendering as well as dependency graphs for the user’s L^AT_EX papers, for a price and using closed source software.

In this work, we introduce KnowTeX, a lightweight and user-friendly tool for generating dependency graphs directly from L^AT_EX sources. Conceptually, KnowTeX follows the same principles as plasTeXdepgraph and the tool Lean Blueprint, using explicit markup to indicate relationships among statements. However, unlike systems integrated into broader ecosystems or relying on machine learning, KnowTeX focuses purely on the structure expressed within the text itself. Authors can specify connections by inserting simple L^AT_EX commands such as `\uses{...}`, from which the program automatically builds a dependency graph. The tool operates independently—no external framework, theorem prover, or AI model is required—and outputs graphs in standard DOT and TikZ formats. By emphasizing transparency and simplicity, KnowTeX aims to make dependency visualization accessible to anyone working with mathematical documents, bridging the gap between readability and formal precision.

Our goal here is not only to introduce KnowTeX as an independent tool, but also to highlight the broader role that dependency graphs could play in mathematical communication. We focus on how such graphs can serve as a standard resource for organizing, teaching, and formalizing mathematical knowledge. Through comparisons with existing systems such as Lean Blueprint, plasTeXdepgraph, and the machine learning based Trouver, we discuss the strengths and limitations of each approach and show how KnowTeX extends this line of work in a simpler, more accessible direction. Ultimately, we argue that dependency graphs deserve to become a standard component of mathematical writing, benefiting both human understanding and automated reasoning.

2 Dependency graphs

A natural starting point for understanding the value of dependency graphs lies in their simplest application: supporting human learning. Dependency graphs serve a practical role in navigating and learning mathematical material. Beyond illustrating conceptual structure, they help readers plan efficient study paths and identify the minimal background required for a given result. In our case study based on Tom Leinster’s ‘Basic Category Theory’ [Lei14], a graph mapping definitions, lemmas, and propositions makes it clear which earlier concepts support key theorems such as the Yoneda Lemma or the Adjoint Functor Theorem. Rather than reading linearly from the beginning, a learner can focus on the relevant portion of the conceptual network, reducing cognitive load and saving time.

Dependency graphs also play a crucial role in the technical side of mathematical formalization. When formalizing a theory within a proof assistant, efficiency often depends on identifying exactly which concepts and results are required for a particular proof. A dependency graph provides a clear map of these relationships, allowing researchers to include only the necessary components while omitting unrelated material. Although fully formalizing every result in a textbook can be an ideal long-term goal, in practice it is more efficient to formalize selectively – focusing on the minimal set of definitions and lemmas needed for a given theorem. This selective approach reduces both computational load and human effort, as shorter, more focused formalizations are easier to maintain and verify. In this way, dependency graphs help optimize the design and scalability of formal libraries, supporting faster and more resource-efficient proof development.

Dependency graphs can also contribute to the construction of machine learning datasets and structured mathematical corpora. The need for such datasets and math corpora has been discussed by several people, e.g., de Paiva and Fong [FdP21], Collard et al [CDPS24], Welleck et al [WLB⁺21].

⁴See <https://github.com/hyunjongkimmath/trouver>.

⁵See <http://www.sciencestack.ai>.

There are initiatives to compile a large-scale, machine-readable collection of articles in research mathematics, starting with category theory abstracts from the journal Theory and Applications of Categories⁶, enabling the study of mathematical language and concept extraction for large language models [dPGKM23].

One of the major challenges in such work is ensuring that all relevant concepts and their inter-relations are accurately captured. Missing definitions, unlinked results, or redundant statements can limit the dataset’s usefulness for both linguistic and logical analysis. By providing an explicit representation of how mathematical statements depend on one another, dependency graphs can help detect these issues and guide curation. They offer a principled way to align informal mathematical writing with formal structures, improving both dataset quality and interpretability for AI-driven research in mathematics. In summary, dependency graphs serve as versatile tools that reveal the structure of mathematical knowledge at multiple levels, from supporting individual learning to guiding formalization and lexicon/corpus construction. Their ability to make conceptual relationships explicit allows both humans and machines to navigate mathematics more effectively.

To explore one concrete way of generating such graphs, the next section introduces KnowTeX, our standalone tool designed to create dependency graphs directly from L^AT_EX documents in a simple and transparent manner.

3 KnowTeX

KnowTeX⁷ is a standalone Python application that generates dependency graphs directly from L^AT_EX sources. The program takes a single L^AT_EX file as input and produces two main outputs: a DOT file and a TikZ-based L^AT_EX file representing the same graph. The DOT file is a standard Graphviz format that can be rendered or made interactive using external visualization tools, such as Graphviz⁸ or viz.js⁹, making it suitable for integration into HTML pages or online repositories. The TikZ¹⁰ file provides a static version of the graph that can be included directly in L^AT_EX documents or exported as a PDF. Before producing these outputs, KnowTeX offers a preview feature within the application interface, allowing users to inspect the generated graph visually.

The design of KnowTeX emphasizes both flexibility and control for the user. After loading a L^AT_EX source file, the user specifies which types of mathematical environments should be scanned by the program.

Commonly recognized environments include ‘Definition’, ‘Theorem’, ‘Lemma’, ‘Proposition’, ‘Corollary’, ‘Construction’, ‘Example’, and ‘Remark’. These can be selected individually or collectively, depending on the desired level of detail. To accommodate variations in authoring styles, KnowTeX also supports multiple naming conventions through regular expressions, for example, recognizing shortened forms such as “thm” or “def” as valid environments. When a document contains multiple chapters, the user can choose to analyze the entire text or focus on a specific chapter, allowing for localized graph generation and more targeted exploration. Once the selections are made, the program automatically scans the specified environments, preparing the data for dependency extraction and graph construction as described below.

To construct the dependency graph, KnowTeX relies on a lightweight command-based linking system embedded within the L^AT_EX source. We illustrate the procedure for introducing dependencies in KnowTeX with a simple example. The author writes standard mathematical environments in L^AT_EX assigns labels in the usual way, and records dependencies by inserting the \uses command.

```
\begin{definition}\label{def:ring}
A ring is a set with two operations satisfying ...
\end{definition}

\begin{lemma}\label{lem:ring-unit}
\uses{def:ring}
In a ring, if $1=0$ then every element is zero.

```

⁶<https://github.com/ToposInstitute/tac-corpus>

⁷See <https://github.com/ElifUskuplu/KnowTeX>.

⁸<https://graphviz.org/>

⁹<https://viz.js.org/>

¹⁰<https://tikz.dev/>

```

\end{lemma}

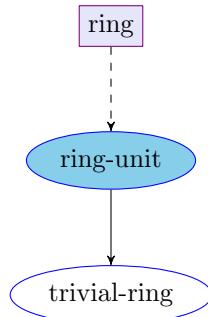
\begin{proof}
Trivial from the axioms.
\end{proof}

\begin{corollary}\label{cor:trivial-ring}
\uses{def:ring}
If a ring satisfies  $1 = 0$ , then it is the trivial ring  $\{0\}$ .
\end{corollary}

\begin{proof}
\uses{lem:ring-unit}
By Lemma~\ref{lem:ring-unit}, if  $1 = 0$ , then every element equals  $0$ .
Hence, the ring contains only one element,  $0$ , and is therefore the trivial ring.
\end{proof}

```

From this annotated source, KnowTeX extracts labeled mathematical objects and their declared dependencies. When the `\uses` command appears inside the statement of a definition, lemma, or corollary, it records a conceptual dependency; in the generated graph, this relation is visualized with a dashed edge. When it appears inside a proof environment, it records a direct logical dependency arising from the proof itself, represented by a solid edge. Running KnowTeX on the example above produces the following TikZ graph:



The resulting graph contains a node for the definition of a ring, a node for the lemma, and a node for the corollary. The dashed edge from the definition node to the lemma node represents a conceptual dependency declared in the lemma statement. The solid edge from the lemma node to the corollary node represents a logical dependency declared in the proof of the corollary. KnowTeX computes the transitive closure of the dependency relation when constructing the graph. For this reason, although the corollary conceptually depends on the definition of a ring, no additional dashed edge from the definition node to the corollary node is drawn. This dependency is already implied by the path through the lemma, and omitting the redundant edge keeps the visualization readable while preserving the full dependency information.

In addition to the `\uses` command, KnowTeX supports an optional `\proves` command, used when the link between a proof and its corresponding statement cannot be inferred automatically, for example, when intervening remarks or text interrupt the expected structure. In such cases, `\proves` explicitly declares which labeled theorem or lemma is being proved, ensuring that all dependencies are correctly represented in the resulting graph.

In Figure 1, the graph is extracted from the first chapter of *Basic Category Theory* by Leinster. In this graph, boxes represent definitions, white diamonds constructions, blue diamonds propositions, and ellipses lemmas. Additional chapters and examples involving more environments are provided in the GitHub repository of the program.

The most influential system for representing mathematical dependencies is Lean’s ‘Blueprint’, which provides structured links between informal text and formal proof code. Building on this idea, KnowTeX extends the same command logic to the broader L^AT_EX ecosystem while remaining independent of any specific proof assistant. It can thus be viewed as an agnostic blueprint tool that could complement

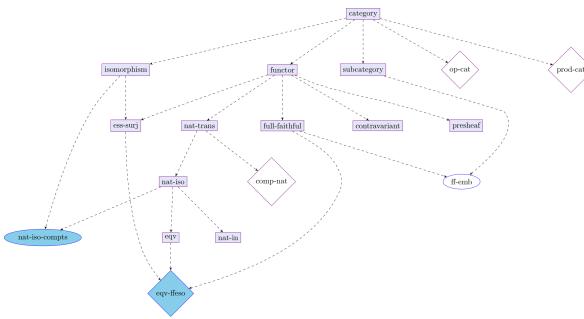


Figure 1: The dependency graph of Chapter 1 of Leinster’s ‘Basic Category Theory’

frameworks such as Lean, Rocq, Isabelle, or Agda. The command system in KnowTeX follows the same underlying structure of Massot’s plasTeXdepgraph plugin, which the Lean Blueprint relies on for graph generation. This deliberate compatibility ensures that \LaTeX documents prepared for either system can be processed directly by KnowTeX without modification, yielding consistent dependency relations.

We verified this compatibility across several case studies, confirming that KnowTeX reproduces the same dependency mappings as these earlier tools while offering a more streamlined, user-oriented workflow. PlasTeX¹¹ is a Python package that processes \LaTeX documents into an XML-DOM-like object, which can then be used to generate various types of output, including HTML and Braille. The plasTeXdepgraph plugin compiles \LaTeX files into an HTML format that integrates the dependency graph directly into the rendered document, allowing users to click on nodes and navigate to the corresponding results. This interactivity is effective when the \LaTeX source is written for plasTeX, but compiling legacy or complex documents, those with custom macros or nonstandard environments, can be challenging. In our experiments with Tom Leinster’s Basic Category Theory ([Lei14]), successful compilation required extensive editing and removal of non-essential macros. The Lean Blueprint system extends this HTML-based foundation by connecting dependency graphs with formal proof development, using visual cues to indicate proof status and links to both \LaTeX and Lean sources. These features are powerful but Lean-centered and presuppose an existing formalization.

In contrast, KnowTeX isolates and simplifies this functionality. It focuses solely on dependency visualization, requiring no external ecosystem, HTML conversion, or formal proof integration. Like plasTeX, it allows color and shape customization, but through a lightweight, standalone interface that makes dependency analysis accessible to authors and researchers seeking to understand or present the structure of mathematical texts without entering a formal workflow.

The Trouver¹² system takes a machine-learning approach to extracting dependency information from \LaTeX sources. Unlike KnowTeX, which relies on explicit commands, Trouver infers relationships automatically from textual and structural patterns. This automation makes it attractive for large-scale or legacy corpora. But it also makes it less predictable: inferred links may miss subtle logical dependencies or introduce spurious ones. KnowTeX, by contrast, prioritizes explicit author control and reproducibility over automation, providing a complementary rather than competing model.

These comparisons illustrate how KnowTeX occupies a distinct position within the existing ecosystem: it preserves compatibility with established tools while remaining lightweight, independent, and focused solely on the visualization of the logical structure. We should also remark that while our experiments and code are geared towards mathematical research text, most of the so-called hard sciences could, perhaps, benefit from a tool like KnowTeX. \LaTeX is used by physicists, chemists, biologists, etc, and the logical dependency trees that KnowTeX makes explicit should be beneficial to these other sciences too. We turn to the broader implications of this approach and outline possible directions for future development.

¹¹See <https://plastex.github.io/plastex/>.

¹²<https://github.com/hyunjongkimmath/trouver>

4 Future work

Beyond the current version, several directions remain open for extending both KnowTeX and the broader use of dependency graphs in mathematical writing:

Expanding case studies: Future work will involve applying KnowTeX to a wider variety of mathematical texts and educational materials. Demonstrating its usefulness across different writing styles and domains will help establish the practical value of dependency graphs and support their adoption as a writing standard.

AI-assisted command generation: Inspired by systems such as Trouver, one possible enhancement would be to automate the insertion of `\uses` and `\proves` commands through AI-based text analysis. Such a feature could reduce authors' manual workload while preserving accuracy and alignment with the original text structure in addition to other dependency graph tools.

Greater interactivity and customization: The visualization could be made more dynamic by allowing users to define every graphical parameter from color schemes and node shapes to edge styles directly within the interface. Additionally, the preview window could evolve into an interactive display, enabling users to click nodes that link back to corresponding definitions or theorems within the compiled PDF.

Integration with large-scale corpora: In large mathematical corpora, dependency graphs can serve as structured metadata, recording which results depend on which definitions. Such metadata enable corpus-level querying, visualization, and linking between informal and formal representations. Integrating KnowTeX within corpus-building pipelines would make it possible to study the structure of mathematical theories at scale, supporting both formalization and machine learning research.

These extensions point toward a future in which dependency graphs are not only a visualization aid but an integral part of how mathematical knowledge is written, shared, and explored.

5 Conclusion

In summary, our motivation for developing KnowTeX as an independent, L^AT_EX-oriented tool stems from a desire to make dependency graphs a standard element of mathematical writing. Authors already rely on L^AT_EX's label–reference system when structuring their definitions, theorems, and proofs; KnowTeX simply extends this workflow by allowing them to record the logical connections among these labeled elements through lightweight commands such as `\uses`.

By embedding these relationships directly into the writing process, authors can automatically generate graphical summaries of their work adding, for instance, a dependency graph at the end of each paper or at the conclusion of every chapter in a textbook. This integration preserves the author's natural writing habits while making the conceptual structure of the text explicit and accessible. In this sense, KnowTeX aims not to replace existing online or formal systems, but to complement them by enriching the final L^AT_EX output itself with visual representations of mathematical knowledge.

References

- [BC10] Yves Bertot and Pierre Castran. Interactive theorem proving and program development: Coq'art the calculus of inductive constructions. 2010.
- [CDPS24] Jacob Collard, Valeria De Paiva, and Eswaran Subrahmanian. Mathematical entities: Corpora and benchmarks. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 11080–11089, 2024.
- [dPGKM23] Valeria de Paiva, Qiyue Gao, Pavel Kovalev, and Lawrence S. Moss. Extracting mathematical concepts with large language models, 2023.
- [FdP21] Brendan Fong and Valeria de Paiva. Introducing the MathFoldr project, <https://topos.institute/blog/2021-07-11-introducing-mathfoldr>, 2021.
- [Lei14] Tom Leinster. *Basic category theory*, volume 143. Cambridge University Press, 2014.

- [MKA⁺15] Leonardo de Moura, Soonho Kong, Jeremy Avigad, Floris Van Doorn, and Jakob von Raumer. The lean theorem prover (system description). In *International Conference on Automated Deduction*, pages 378–388. Springer, 2015.
- [Nor09] Ulf Norell. *Dependently Typed Programming in Agda*, pages 230–266. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [WLB⁺21] Sean Welleck, Jiacheng Liu, Ronan Le Bras, Hannaneh Hajishirzi, Yejin Choi, and Kyunghyun Cho. Naturalproofs: Mathematical theorem proving in natural language. *arXiv preprint arXiv:2104.01112*, 2021.