Interactive Taxonomy Development with Hybrid Methods

Jiaming Qu
Amazon
Seattle, USA
qjiaming@amazon.com

Madhu Gopinathan
Amazon
Seattle, USA
madhugop@amazon.com

Shayan Ali Akbar
Amazon
Seattle, USA
shayaakb@amazon.com

Omar Alonso
Amazon
Santa Clara, USA
omralon@amazon.com

Abstract

Taxonomies organize knowledge into hierarchical structures that support effective information seeking behaviors. However, developing taxonomies in fast-evolving domains like e-commerce remains a labor-intensive process. In this paper, we present an interactive system that assists users in expanding taxonomies through automated knowledge discovery from large text corpora. On the back end, our *hybrid methods* combine topic modeling and large language models (LLMs) to uncover emerging concepts, generate concise summaries, and suggest mappings to taxonomy nodes. On the front end, we develop an interactive web-based interface that supports iterative, human-in-the-loop taxonomy expansion. We demonstrate the system's versatility through two scenarios using publicly available datasets: amplifying a preliminary taxonomy in the e-commerce domain and refining a mature taxonomy in the medical domain.

CCS Concepts

• **Human-centered computing → Interactive systems and tools**; **Visualization systems and tools**.

Keywords

Taxonomy Development, Interface Design, Topic Modeling, LLMs

ACM Reference Format:
Jiaming Qu, Madhu Gopinathan, Shayan Ali Akbar, and Omar Alonso. 2026. Interactive Taxonomy Development with Hybrid Methods. In *2026 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '26), March 22–26, 2026, Seattle, WA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3786304.3787912

1 Introduction

Taxonomies are hierarchical structures that organize knowledge into meaningful categories, supporting a wide range of information seeking tasks such as search, recommendation, and analytics [12]. They are fundamental to how people and systems navigate large volumes of data and have been applied across domains. For instance, e-commerce companies use taxonomies to organize product catalogs [8]; researchers utilize taxonomies such as Medical Subject Headings (MeSH) [18], Gene Ontology [5], and the ACM Computing Classification System [4] to categorize scientific knowledge.

While these taxonomies are high-quality outcomes of domain expertise, critical challenges remain: *how can we ensure that a taxonomy reflects the latest knowledge structure*? And when it does not, *how can we evolve it in step with emerging concepts and data*? An outdated taxonomy can negatively affect downstream systems: users encounter misplaced products in search results, business operations rely on obsolete knowledge, and novel concepts goes untracked. However, maintaining taxonomies is often time-consuming and resource-intensive, especially when the hierarchy contains thousands of nodes or the domain evolves rapidly with new data. Although previous work has explored expanding taxonomies with external knowledge bases [6, 23, 24], few has effectively captured insights from large-scale data directly.

To address this challenge, we developed a human-in-the-loop system. We began with a user-centered design process involving two ontologists working on taxonomy-related applications (Section 3.1). Through contextual inquiries, we examined how they expand taxonomies from raw text corpora. The sessions revealed two major pain points: (1) extensive manual effort required to identify emerging concepts, and (2) the absence of an interactive tool to integrate new insights into existent taxonomies. These findings have motivated us to develop approaches that facilitate taxonomy expansion while preserving expert oversight over quality.

At a high level, our system comprises two key components (Figure 1). On the back end (Section 3.2), we employ hybrid methods—combining topic modeling and LLMs—to uncover concepts from large text corpora, generate concise summaries, and suggest mappings to taxonomy nodes. On the front end (Section 3.4), we develop an interactive web-based interface that enables users to expand a taxonomy with LLM-generated insights.

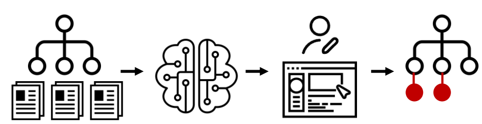

Figure 1: Overview of our system design. From left to right: a taxonomy and raw documents (input), a back-end system (Section 3.2), a front-end web-based interface (Section 3.4), and the updated taxonomy with new nodes (output).

We demonstrate the system using two publicly available datasets that represent distinct usage scenarios (Section 4): (1) amplifying a preliminary taxonomy, where concepts are ambiguous and require

ACM ISBN 979-8-4007-2414-5/2026/03
https://doi.org/10.1145/3786304.3787912

immediate clarification, and (2) refining a mature taxonomy, where well-defined nodes can still be further expanded with nuanced insights. Preliminary results show that our system generalizes effectively across domains and supports both early-stage taxonomy construction and incremental refinement.

Our contributions are twofold. First, we propose a hybrid approach that integrates topic modeling and LLMs to automatically reveal hidden-yet-meaningful patterns in large text corpora and generate data-driven insights for taxonomy expansion. Second, we develop an interactive, human-in-the-loop system that allows experts to scrutinize LLM-generated suggestions during taxonomy expansion. Our system not only supports human–AI collaboration in taxonomy development, but also provides design principles to broader knowledge organization tasks as well.

2 Related Work

Taxonomy Development. Constructing taxonomies has traditionally relied on expert efforts [12]. While such approaches often produce high-quality hierarchies, they are time-consuming, resource-intensive, and difficult to scale [14]. These limitations have motivated the development of automated taxonomy induction methods. Early studies have explored pattern-based hypernym extraction [11] and graph-based techniques leveraging term co-occurrence [25]. Recent advances in natural language processing have enabled automated methods that can identify and hierarchically organize concepts from large text corpora [19, 20].

In addition to constructing taxonomies from scratch, a complementary line of research focuses on taxonomy expansion, i.e., enriching existing hierarchies with new concepts and relationships. A common approach is to extract candidate concepts from external knowledge bases first, and then employ techniques such as graph propagation [23], embedding-based representations [24], or prompting LLMs [6] to predict optimal insertions. Our work is inspired by prior studies that use topic modeling to discover relevant keywords for a given taxonomy node [1]. However, we use topic modeling to uncover emerging themes across the corpus, and leverage LLMs to summarize them and suggest mappings to existing taxonomy nodes. More importantly, we develop a front-end system following the human-in-the-loop paradigm that integrates automated knowledge discovery with expert validation in taxonomy development [14, 20].

Topic Modeling. Topic modeling seeks to uncover latent semantic structures within document collections. Early topic modeling approaches like Probabilistic Latent Semantic Analysis [13] and Latent Dirichlet Allocation (LDA) [2] identify topics mainly based on word co-occurrences. The key idea in such generative topic models is to consider each document as a mixture of topics, where each topic represents a probability distribution over words. Recent advancements have used document embeddings to enhance the coherence and semantic relationship between topics [7, 10, 27]. In this paper, we adopt Semantic Signal Separation (S^3) [15] for topic modeling. Unlike traditional topic models using bag-of-words representations, S^3 conceptualizes topics as independent semantic directions within a continuous embedding space and disentangles latent dimensions using independent component analysis [17]. This embedding-based topic model captures nuanced semantic themes to generate diverse and coherent topics. For each topic, we passed keywords and documents to an LLM for topic interpretation.

3 System Design

3.1 User-Centered Design Process

We followed a user-centered, iterative design process—consisting of contextual inquiry, prototyping, and expert feedback—methods commonly adopted in human–computer interaction research [9]. Two Amazon ontologists who were engaged in an internal taxonomy project voluntarily participated throughout the process.

The design process proceeded as follows. The first author conducted initial contextual inquiries with each ontologist to understand their workflows for expanding taxonomies with new insights from large text corpora. The sessions revealed two key challenges: (1) substantial manual effort to identify emerging concepts not yet captured in the taxonomy, and (2) lack of appropriate tools to support the taxonomy expansion workflow. Consequently, the ontologists expressed a need for a system that could automatically uncover new data patterns and visualize them through an interactive interface. Insights from these inquiries informed the early system sketches. Two of the authors then conducted follow-up interviews with the same ontologists to refine system design.

Guided by these design principles, we developed a system comprising two components: a back-end system that employs topic modeling and LLMs to generate data-driven insights (Section 3.2), and a web-based front-end interface that enables users to interactively review LLM-generated recommendations (Section 3.4). We describe these two components in detail below.

3.2 Back-end: Data-driven Insights Generation

The back-end system automates the discovery of emerging concepts from large text corpora and mapping to taxonomy nodes. Hence, it operates in three sequential stages: topic discovery, topic interpretation, and topic-to-taxonomy mapping. For all LLM-related components, we use Claude 4 via the Amazon Bedrock API[1] with default hyper-parameters. We implement Chain-of-Thought prompts [26] through the DSPy library [16].

Topic Discovery: We employ S^3 for topic modeling. Since S^3 identifies latent themes in the embedding space rather than word co-occurrences, we use SBERT [21]—a transformer-based encoder—to encode input documents into embeddings.

Topic Interpretation: For each discovered topic, we extract the top 20 keywords and 5 representative documents with the highest scores (computed by S^3). These are passed to the LLM for interpretation. Each time, the LLM is prompted to generate a short phrase summarizing the topic and a concise explanation sentence based on a list of keywords and documents.

Topic-to-Taxonomy Mapping: We provide the short phrase for each topic, its explanation, and the full list of taxonomy nodes to the LLM, and prompt it to identify up to three most possible mappings with a confidence score (1–10) for each suggestion.

In practice, the system is *semi-automated.* Users first specify the desired number of topics and execute the topic discovery and interpretation stages via command-line scripts to generate candidate topics. If satisfied with the results, users can import them into a web-based interface to review each topic and obtain mapping suggestions from the LLM. We describe the interface below.

[1] https://aws.amazon.com/bedrock/

Table 1: Examples of data-driven insights for taxonomy expansion. We first prompted the LLM to summarize topics into concise phrases based on 20 keywords and 5 representative documents. A second LLM then mapped each topic to taxonomy nodes.

Dataset	Keywords (top 5)	LLM-summarized Topic	LLM-suggested Mapping (top 1)
ABCD	cotton, sponge, fabric, wool, polyester	Product Material Information	Single Item Query
	premium, benefits, policies, secure, comfort	Premium Membership Benefits	Store-wide Query → Membership
	membership, levels, silver, bronze, platinum	Membership Level Information	Store-wide Query → Membership
MeSH	trachoma, trachomatis, africa, asia, arabia	Trachoma Epidemiology	Eye Infections, Bacterial
	demodex, blepharitis, demodicosis, descemet, lashes	Demodex Blepharitis	Eye Infections, Parasitic

3.3 Datasets for Demonstration

While the system was originally developed to assist ontologists in an internal taxonomy project, the data cannot be released due to confidentiality restrictions. Nevertheless, the underlying ideas are *generalizable across domains*. To demonstrate this, we showcase our system using two publicly available datasets with similar structures—each containing a hierarchical taxonomy associated with document collections. We select these two datasets because they represent two *distinct usage scenarios*: one involves a preliminary taxonomy with ambiguous concepts, and the other features a well-maintained taxonomy that can be further refined.

First, we used the Action-Based Conversation Dataset (ABCD) [3], which contains 10K customer–agent dialogues in e-commerce and a two-layer taxonomy summarizing major contact intents in customer service (e.g., account or order issue). The taxonomy[2] includes 10 top-level and 55 second-level nodes. For each dialogue, we concatenated all turns into a single transcript. The second dataset is MeSH, which provides a comprehensive hierarchical vocabulary for indexing biomedical literature. Because MeSH is extensive, we focused on one representative branch: *Diseases → Eye Diseases → Eye Infections*. This branch[3] has 6 top-level and 14 second-level nodes. Using the query "Eye Infections[MeSH]", we retrieved 6k articles (considering only titles and abstracts) from PubMed.

We applied the same topic modeling and LLM analysis pipeline on both datasets with the number of topics set to 50 and 30, respectively. Table 1 presents examples of LLM-summarized topics and suggested topic-to-taxonomy mappings on the two datasets.

3.4 Front-end: Interactive Taxonomy Expansion

The front-end system provides an interactive web-based interface for users to expand a taxonomy with topic modeling outputs[4]. Users begin by uploading two input files: a taxonomy and a topic modeling output containing candidate topics and explanations (as detailed in Section 3.2). Once uploaded, the system parses both files and loads the interface shown in Figure 2. The interface comprises a control panel at the top and a main workspace below. The workspace supports two synchronized views—*Topic View* (shown by default) and *Taxonomy View*—which users can toggle seamlessly for topic exploration and taxonomy visualization.

In the *Topic View*, uploaded topics are displayed in a searchable table. Users can browse topics freely or issue queries (e.g., "subscription") to focus on topics of interest. Each topic entry includes its explanation and a `Get AI Suggestions` button. Clicking the button triggers a backend LLM call that generates up to three topic-to-taxonomy mappings, each with a confidence score shown in different opacity levels. For each topic, users can take one of three actions after scrutinizing LLM-suggested mappings: (1) *Ignore*—dismiss the topic as irrelevant; (2) *Insert as a New Node*—add it as a new taxonomy node if it represents a novel concept; or (3) *Insert as an Example*—attach it as an illustrative example to clarify an existing node. For insertion, users may either adopt an LLM-recommended path or manually select a destination node using a navigation tool. The `Insert Topics` button allows users to add selected topics to the taxonomy or saving them as examples.

Users can switch to the *Taxonomy View* to visualize the updated structure (Figure 3). The taxonomy visualization is implemented with the D3.js library[5], which renders the taxonomy in an interactive tree structure. Newly added nodes are highlighted in red for visual identification. Users can zoom, pan, and expand or collapse branches to inspect hierarchical relationships in the taxonomy. Topics inserted as examples are stored separately in another JSON file. Finally, users can click the `Download Taxonomy` button to export the updated taxonomy and the file containing topic examples.

4 Usage Scenarios and Demo Plan

Scenario I: Amplifying a Preliminary Taxonomy (ABCD). As illustrated in Table 1, our system can expand an early-stage taxonomy with vague concepts such as that in the ABCD dataset. Consider an ontologist exploring customer–agent dialogues labeled with the original ABCD taxonomy. During review, the ontologist notices frequent mentions of words such as *cotton*, *sponge*, and *polyester* in certain transcripts and begins to wonder whether customers frequently ask about product materials—an intent not covered in the existing taxonomy. Using our system, this new concept *Product Material Information* is automatically surfaced through topic modeling and mapped to the *Single Item Query* node, which reduces the manual effort of hypothesis generation and validation.

Similarly, the topics *Membership Level Information* and *Premium Membership Benefits* are both mapped to *Store-wide Query → Membership* node. By examining these LLM-suggested mappings, the ontologist could have realized that the former reflects inquiries about tier requirements, whereas the latter captures questions about benefits and rewards after achieving a tier. Using the interface, the ontologist could effectively expand the original *Membership* node into two sub-nodes that represent two different concepts.

Scenario II: Refining a Mature Taxonomy (MeSH). Now, consider a biomedical researcher uses our system to refine the MeSH branch *Diseases → Eye Diseases → Eye Infections*. As shown in Table 1, the system identifies new concepts such as *Trachoma*

[2]https://github.com/asappresearch/abcd/blob/master/data/ontology.json
[3]https://www.ncbi.nlm.nih.gov/mesh/68015817
[4]The demo video is available at https://jiamingqu.com/CHIIR26_demo/.
[5]https://d3js.org/

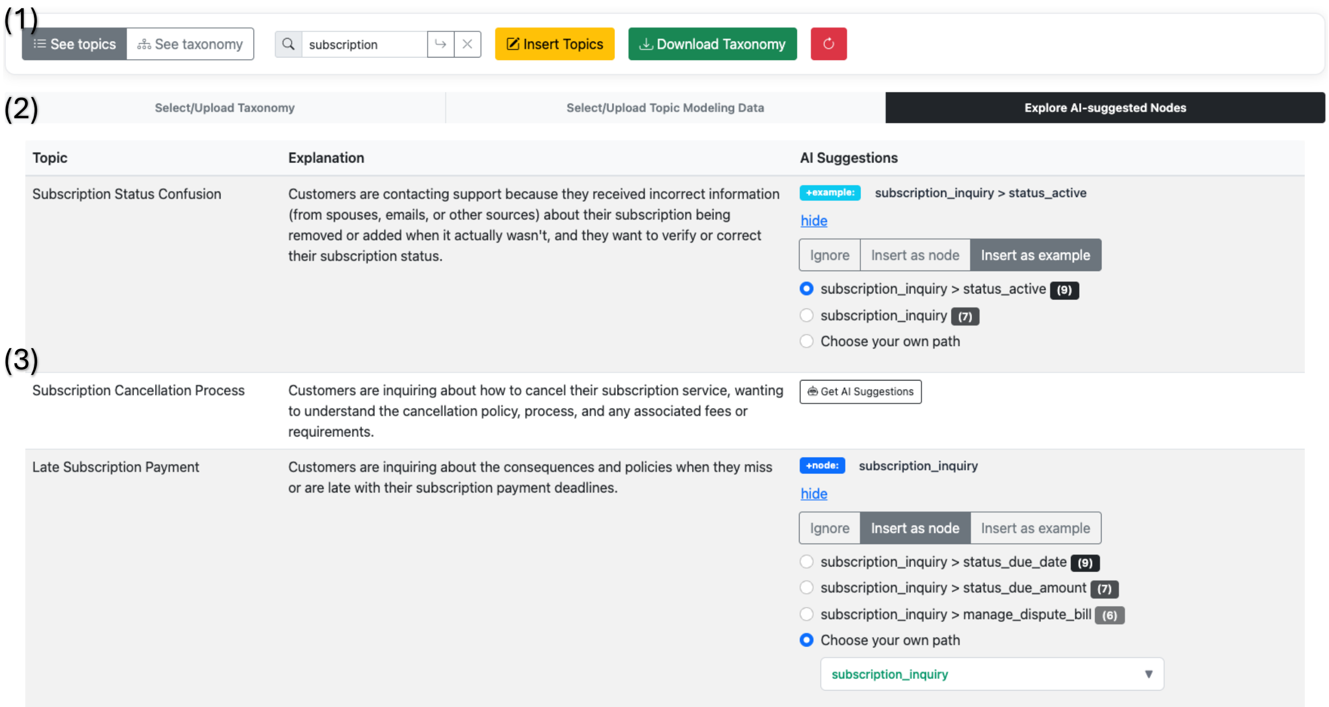

Figure 2: System interface after a taxonomy and topic modeling outputs are uploaded. From top to bottom: (1) the control panel, which includes buttons for switching views, searching for topics, inserting topics as nodes or examples, downloading the updated taxonomy and file of examples, and resetting the session; (2) a progress bar indicating the current step; and (3) the main workspace displaying topics. Each topic panel contains a `Get AI Suggestions` button that triggers LLM calls for topic-to-taxonomy mappings. For each topic, users can ignore it or insert it as a new node or an illustrative example. For insertion, users can accept LLM-suggested mappings or select their desired path in the taxonomy.

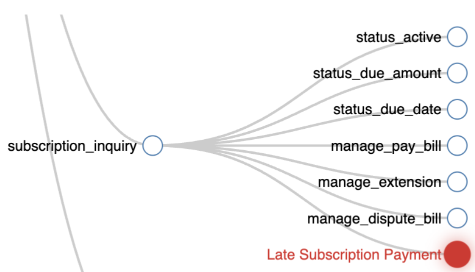

Figure 3: *Taxonomy view* after topic insertion. The newly added nodes are highlighted in red in the interactive tree.

Epidemiology and *Demodex Blepharitis*, which were mapped to *Eye Infections, Bacterial* and *Eye Infections, Parasitic*, respectively. These additions capture meaningful variations—such as region-specific disease prevalence and pathogen-specific etiologies—that are absent from the original taxonomy node. By integrating these nuanced patterns as child nodes or examples, domain experts can incrementally refine an already mature taxonomy without disrupting its structure.

Demonstration Plan: We will demonstrate the system live at the CHIIR conference. Participants will be encouraged to upload their own datasets or use the provided ABCD and MeSH examples. Participants will experience how our system can support two different usage scenarios (i.e., amplifying a preliminary taxonomy and refining a mature taxonomy) in taxonomy development.

5 Discussion

Genralizability: Preliminary results on the ABCD and MeSH datasets suggest that our system can generalize across domains and usage scenarios. More importantly, the overall system architecture is model-agnostic. On the back end, domain-specific pre-trained language models can be incorporated to generate embeddings that further enhance the quality of topics discovered by S^3. Beyond S^3, future topic modeling approaches—those with superior performance—can be seamlessly integrated as well. Likewise, one can use newer LLMs and conduct prompt engineering for more accurate topic interpretation and topic-to-taxonomy mapping results. On the front end, the interface provides a user-friendly workspace for examining machine-generated, data-driven insights. This design can be readily extended to a broader range of human–AI collaboration scenarios in knowledge discovery and information organization.

Limitations and Future Work: As a demonstration paper, our primary goal has been to showcase system functionality rather than to conduct a comprehensive empirical evaluation. As an immediate next step, we plan to conduct a user study comparing how ontologists develop taxonomies using our system versus traditional spreadsheet-based workflows. Such a study will help assess whether our system improves both *efficacy* (e.g., quality of new concepts) and *efficiency* (e.g., time and cognitive effort) during taxonomy expansion. On the system side, we aim to integrate an algorithmic evaluation module that leverages established metrics such as topic diversity and coherence [22] to provide feedback for topic modeling. This module will guide users in selecting an appropriate number of topics and make the system fully automated.

References

[1] Anton Bakalov, Andrew McCallum, Hanna Wallach, and David Mimno. 2012. Topic models for taxonomies. In *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries*. 237–240.

[2] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *The Journal of machine learning research* 3 (2003), 993–1022.

[3] Derek Chen, Howard Chen, Yi Yang, Alex Lin, and Zhou Yu. 2021. Action-Based Conversations Dataset: A Corpus for Building More In-Depth Task-Oriented Dialogue Systems. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021*. Association for Computational Linguistics, Online, 3002–3017. https://www.aclweb.org/anthology/2021.naacl-main.239

[4] Generate Code. 1995. Acm computing classification system. (1995).

[5] Gene Ontology Consortium. 2019. The gene ontology resource: 20 years and still GOing strong. *Nucleic acids research* 47, D1 (2019), D330–D338.

[6] Simone D'Amico, Alessia De Santo, Mario Mezzanzanica, and Fabio Mercorio. 2025. Taxonomy Expansion through Collaborative LLM Mapping. In *Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing*. 1961–1968.

[7] Adji Bousso Dieng, Francisco J. R. Ruiz, and David M. Blei. 2019. Topic Modeling in Embedding Spaces. *Transactions of the Association for Computational Linguistics* 8 (2019), 439–453. https://api.semanticscholar.org/CorpusID:195886143

[8] eBay. 2025. Shop by Category: All Categories. https://www.ebay.com/n/all-categories. Accessed: 2025-10-10.

[9] Claude Ghaoui. 2005. *Encyclopedia of human computer interaction*. IGI global.

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

[11] Marti A. Hearst. 1992. Automatic Acquisition of Hyponyms from Large Text Corpora. In *Proceedings of the 14th conference on Computational linguistics–Volume 2* (Nantes, France) *(COLING '92)*. Association for Computational Linguistics, Nantes, France, 539–545. doi:10.3115/992133.992154

[12] Heather Hedden. 2016. *The accidental taxonomist*. Information Today, Inc.

[13] Thomas Hofmann. 1999. Probabilistic latent semantic indexing. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Berkeley, California, USA) *(SIGIR '99)*. Association for Computing Machinery, New York, NY, USA, 50–57. doi:10.1145/312624.312649

[14] Meng Jiang. 2020. Chapter 2 Part 1: Taxonomy Construction. In *KDD Tutorial*. http://www.meng-jiang.com/tutorial-kdd20-scikg/part3-1_shang.pdf

[15] Márton Kardos, Jan Kostkan, Kenneth Enevoldsen, Arnault-Quentin Vermillet, Kristoffer Nielbo, and Roberta Rocca. 2025. S^3 - Semantic Signal Separation. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (Eds.). Association for Computational Linguistics, Vienna, Austria, 633–666. doi:10.18653/v1/2025.acl-long.32

[16] Omar Khattab, Arnav Singhvi, Paridhi Maheshwari, Zhiyuan Zhang, Keshav Santhanam, Sri Vardhamanan, Saiful Haq, Ashutosh Sharma, Thomas T. Joshi, Hanna Moazam, Heather Miller, Matei Zaharia, and Christopher Potts. 2024. DSPy: Compiling Declarative Language Model Calls into Self-Improving Pipelines. *The Twelfth International Conference on Learning Representations*.

[17] Te-Won Lee. 1998. Independent component analysis. In *Independent component analysis: Theory and applications*. Springer, 27–66.

[18] Carolyn E Lipscomb. 2000. Medical subject headings (MeSH). *Bulletin of the Medical Library Association* 88, 3 (2000), 265.

[19] Yuning Mao, Yue Sun, and Jiawei Han. 2018. End-to-End Reinforcement Learning for Automatic Taxonomy Induction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*. https://arxiv.org/abs/1805.04044

[20] Luca Pantera and Roberto Navigli. 2023. A Short Survey on Taxonomy Learning from Text Corpora. In *Proceedings of EMNLP*. https://aclanthology.org/D17-1123/

[21] 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 and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 3982–3992.

[22] Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*. 399–408.

[23] Wei Shen, Jianyong Wang, Ping Luo, and Min Wang. 2012. A graph-based approach for ontology population with named entities. In *Proceedings of the 21st ACM international conference on Information and knowledge management*. 345–354.

[24] Nikhita Vedula, Patrick K Nicholson, Deepak Ajwani, Sourav Dutta, Alessandra Sala, and Srinivasan Parthasarathy. 2018. Enriching taxonomies with functional domain knowledge. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 745–754.

[25] Paola Velardi, Roberto Navigli, and Valentina De Cerbo. 2013. Taxonomy induction methodologies: A survey. *Comput. Surveys* 47, 2 (2013), 1–44.

[26] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* 35 (2022), 24824–24837.

[27] Xiaobao Wu, Thong Nguyen, and Anh Tuan Luu. 2024. A survey on neural topic models: methods, applications, and challenges. *Artificial Intelligence Review* 57, 2 (Jan. 2024). doi:10.1007/s10462-023-10661-7