Extracting Published Discussions of 13,616 Galaxies with Language Models

Nolan Koblischke and Jo Bovy
David A. Dunlap Department of Astronomy and Astrophysics
Dunlap Institute for Astronomy & Astrophysics
University of Toronto, 50 St. George Street, Toronto, ON, M5S 3H4, Canada
nolan.koblischke@mail.utoronto.ca

Abstract

Astronomers often have difficulty determining whether an object, such as a star or galaxy of interest, has been discussed in the literature. Existing tools enable object identification and linking in papers but do not record whether an object was genuinely discussed in a paper or just appears in a table or catalog. We address this by using language model agents to identify galaxies discussed in more than two sentences across a corpus of 156,006 arXiv astro-ph papers and extract those discussions. We extract 44,332 discussions of individual galaxies containing 85,314 quotes from 18,641 papers. With sky coordinates for 98.2% of discussions, we group them into 13,616 galaxies. We bring these results together into GalaxyWiki, where users can find literature discussing galaxies, read what was said about them, and follow links to the source papers. The code, dataset, and webpage are available at https://github.com/NolanKoblischke/galaxywiki-paper.

1 Introduction

The NASA/IPAC Extragalactic Database (NED) combines identifiers, physical measurements, and literature information for galaxies and other extragalactic objects (Mazzarella, 2020). Similarly, SIMBAD covers galaxies and stars, providing identifiers, basic data, selected measurements, and bibliographic references (Wenger et al., 2000). NED and SIMBAD also provide catalog-level object classifications like whether a galaxy is a spiral or an elliptical (Kuhn et al., 2022). Here we seek a complementary resource: a collected account of how individual galaxies have been interpreted scientifically, grounded in extended quotations from the literature.

Astronomers searching for previous discussions of a galaxy can find the object bibliographies provided by these databases (Mazzarella, 2020; Wenger et al., 2000). However, an object-to-paper link alone does not distinguish an extended discussion from an appearance in a catalog. Readers may therefore need to inspect the linked papers to determine what has actually been said about the galaxy.

Wikipedia provides this kind of scientific interpretation, although its coverage is limited. We count 3,379 English Wikipedia article pages that use the *galaxy* infobox, a proxy for its coverage of galaxies.[1] However, these pages might not provide a complete account of the literature about a given galaxy. Galaxy Zoo offers another setting for discussion, allowing volunteers to classify galaxies (Lintott et al., 2008) and discuss them on forums that bring together citizen scientists and experts (Mankowski et al., 2011).

To simplify this search, we aim to extract every galaxy discussed in more than two sentences in our corpus of arXiv papers, along with quotes. We bring these discussions together in a database that captures scientific interpretations. Astronomers can then immediately check whether a galaxy has been discussed in the extracted literature, read what was said, and identify the papers containing those discussions, as illustrated in Figure 1.

1.1 Related Work

Recent work has used language models to extract information from scientific papers and organize it into resources that are easier to search and read (Lála et al., 2023; Skarlinski et al., 2024). WikiCrow generates cited articles about human genes from the research literature using PaperQA2 agents that search papers and synthesize their findings (Skarlinski et al., 2024). In astronomy, *Lodestone* uses language models to bring together findings from multiple papers in cited encyclopedia

[1] https://linkcount.toolforge.org/?project=en.wikipedia.org&page=Template%3AInfobox+galaxy

Proceedings of the 4th Workshop on Artificial Intelligence for Scientific Publications, pages 174–179
November 10, 2026 ©2026 Association for Computational Linguistics

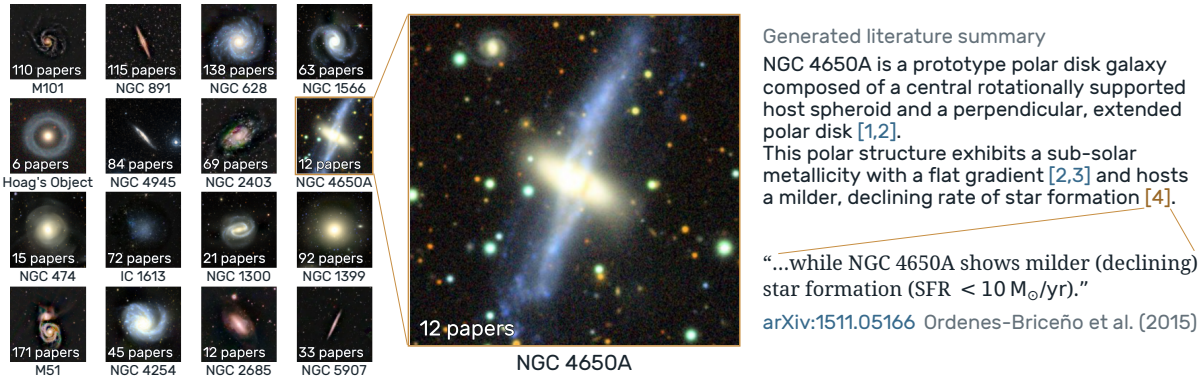


Figure 1: GalaxyWiki brings together published discussions of individual galaxies. *Left:* Selected galaxies and the number of papers discussing each galaxy in our dataset. *Right:* The two-sentence summary of the literature discussions of NGC 4650A (Iodice, 2010, 2014; Spavone et al., 2011; Ordenes-Briceño et al., 2016), along with a citation, extracted quote, and source paper.

articles (Iyer, 2026). Shao et al. (2024) extracts celestial object identifiers and telescope names from journal articles. Wang et al. (2026) constructs a galaxy knowledge graph from literature abstracts, while Li et al. (2026) resolves object mentions to SIMBAD identifiers and link them to scientific concepts. With GalaxyWiki, we focus on the literature discussions of individual galaxies, rather than just their properties or identifiers.

2 Methods

Source corpus We use preprocessed arXiv paper text from the MINT-1T dataset (Awadalla et al., 2024) that is created by parsing the source LaTeX into text (Awadalla et al., 2024). Specifically, we use `minty-astro-ph`,[2] a subset of MINT-1T focused on astro-ph papers. This contains 156,006 astro-ph papers from June 1993 to May 2024. For comparison, we count 299,824 astro-ph postings in the same time period according to the Kaggle arXiv metadata dataset.[3] This incomplete coverage appears largely attributable to the MINT-1T figure-extraction requirements, since the original corpus was designed to pair text with images. In future work, we will attempt to construct a complete astro-ph corpus. As a first processing step, we apply a filter that retains only papers containing the standalone word "galaxy" or "galaxies," leaving 87,899 papers.

Galaxy mention extraction We process each paper with the GPT-5.5[4] language model in the open-source Codex[5] agent harness, and task it with identifying mentioned galaxies and extracting quotes. In the Codex harness, the language model has access to terminal-based tools that allow it to copy quotes verbatim and selectively read the paper. For each galaxy identified, we request one to three quotes totaling three to eight sentences, and ask the language model to generate a 20-to-40-word summary supported by those quotes. We also ask it to choose up to five relevant topics drawn from a fixed subset of the Unified Astronomy Thesaurus (UAT; Frey and Accomazzi, 2018), enabling thesaurus-driven search in our GalaxyWiki. We further task the language model with extracting the galaxy's sky coordinates when explicitly provided in the paper, which is required for associating galaxies across multiple papers. We explicitly instruct the language model to ignore mentions of the Milky Way Galaxy and objects within it and instruct it to extract at most 30 galaxies per paper.

To reduce the setup and take-down overhead of the agents, we assign each agent a batch of 10 papers. We run 20 concurrent agents for approximately 71 hours with token usage corresponding to approximately $12,600 USD at standard API prices.[6] We extract 44,332 discussions of individual galaxies made of 85,314 quotes contained in 18,641 unique papers.

[2] https://huggingface.co/datasets/Smith42/minty-astro-ph

[3] https://www.kaggle.com/datasets/Cornell-University/arxiv

[4] https://developers.openai.com/api/docs/models/gpt-5.5

[5] https://github.com/openai/codex

[6] https://developers.openai.com/api/docs/pricing

Coordinate resolution for grouping discussions Access to the position of the discussed galaxy on the sky allows us to associate discussions across papers and with other catalogs. However, our paper extraction step only captures coordinate information for 3,669 mentions, or 8.3%.

Instead of relying on the coordinates from the extraction step, and to vastly improve coordinate coverage, we use the extracted galaxy identifiers if they follow standard formats such as the NGC, SDSS, and 2MASS identifiers (York et al., 2000; Skrutskie et al., 2006) to match them to an entry on the NED database (Mazzarella and NED Team, 2017). We programmatically query the NED API using galaxy identifiers that match these standard formats. Using this method, we successfully retrieve coordinates for 23,770 out of 25,206 attempted identifier queries, which populates coordinates for 53.6% of mentions.

For the remaining mentions, we task GPT-5.5 agents with finding coordinates by first instructing them to search the original paper, then query the NED API using any compatible galaxy identifier from the paper, not necessarily the formats we attempted. As a fallback, we provide internet access to allow the agents to follow links and references in the paper and retrieve resources such as VizieR catalogs published by the authors (Ochsenbein et al., 2000). We record the internet resources used so that we can audit the coordinate search and we ask the agents to report "unresolved" if they determine the task is impossible.

With agents, we find coordinates for 6,701 records in the source papers, 11,503 through NED, and 1,684 from external internet sources. These results provide coordinates for 19,888 records, or 44.9% of all mentions, leaving 674 records unresolved. Across the programmatic and agentic methods, coordinates are available for 43,546 of 44,332 mentions, or 98.2%.

We then group the 43,546 located galaxy mentions across papers using these sky coordinates. We link positions separated by less than 1" and include neighbors of neighbors. The procedure yields 13,616 entries, each of which we interpret as an individual galaxy discussed in the source papers. However, this relies on the choice of 1" and grouping by sky position does not ensure that these entries correspond to distinct galaxies.

GalaxyWiki In addition to releasing a dataset, we built a web-based wiki intended for astronomers and the general public. For each galaxy page, we generate a two-sentence summary with citations from the available extracted quotes and paper-level summaries using Gemini 3.5 Flash.[7] We programmatically verify that the source quotes and papers used for the summaries are in our dataset. Alongside this summary, each page lists the papers, their summaries, and their quotes, with direct links to the quoted passages where available through the arXiv HTML interface. We allow users to search for galaxies by sky position, UAT topic, name, or text description. To enhance user experience, we retrieve available preview images from the DESI Legacy Imaging Surveys DR10 (Dey et al., 2019), Pan-STARRS (Chambers et al., 2016), and Digitized Sky Survey 2 (McLean et al., 2000) using the coordinates.

3 Results

The extraction yields 44,332 galaxy mention records containing 85,314 quotes from 18,641 papers. The 13,616 grouped entries include 4,187 with multiple mentions and 9,429 (69.2%) with a single mention. Among entries with multiple mentions, the median is three mentions per entry. Every page in the wiki includes a preview image from Legacy Survey DR10 (9,545 pages), Pan-STARRS (3,831 pages), or DSS2 (240 pages).

To assess whether agents accurately copied quotations, we compare each extracted quote with the source paper text. We find character-for-character matches for 66,362 of the 85,314 quotes (77.8%). Normalizing Unicode, spacing, and punctuation yields another 15,971 matches, bringing the total to 82,333 (96.5%). A further 34 match after ignoring case and removing all characters other than letters and numbers. Finally, we find that 2,826 can be split at sentence punctuation into fragments that each match the source after Unicode and whitespace normalization, indicating that agents often combined non-contiguous excerpts into a single quotation. However, textual matching alone does not verify correct attribution or preservation of the source's meaning.

We manually review 120 distinct galaxy mention records. We first review 60 randomly selected records with coordinates, focusing on coordinate verification. We find support for the intended galaxy identification and correct quotations

[7] https://ai.google.dev/gemini-api/docs/models/gemini-3.5-flash

in 58 cases. Two records contained quotes that provided little substantive discussion or described a different object. In three otherwise correctly identified records, the galaxy identifiers were truncated or the paper-supplied positional coordinates were rounded, which reduced the precision of the extracted coordinates, with the largest discrepancy reaching 12.2" because an agent decoded a shortened identifier that contained coordinate information. Such errors can separate discussions of the same galaxy into different entries.

We then review another 60 records drawn at random, this time without restricting coordinate availability. During review, all 60 records identify an individual galaxy and have quotes that contain at least three distinct substantive sentences about that galaxy. Two records have passages that only discuss the galaxy jointly with other galaxies, though we determine that these passages still provide useful information for the GalaxyWiki. Four records contain source text in which the MINT-1T parsing pipeline deleted or corrupted LaTeX symbols. This omits details that could be useful to understand the discussed galaxy. Despite these issues, the review supports the conclusion that the pipeline achieves its primary goal, correctly identifying an individual galaxy and extracting its substantive discussion in 118 of 120 cases.

4 Discussion

Our GalaxyWiki is a unique dataset for the field, and has already been used for 1) determining if galaxies discovered by an anomaly detection method have been previously discussed in the literature (Cheng et al., 2026) and 2) enhancing descriptions of galaxy images for training an astronomy-specific vision-language model (Habib et al., 2026).

The GalaxyWiki dataset provides a literature-selected sample of galaxies that have received extended scientific discussion and are therefore of interest to the community. Therefore, one could use the catalog to guide similarity and semantic searches for galaxies with related features (Stein et al., 2021; Koblischke et al., 2026) and support the development of methods for detecting anomalies or identifying galaxies of scientific interest (Lochner and Bassett, 2021; Etsebeth et al., 2024). The database could also support literature surveys in astronomy, complementing semantic retrieval and corpus-wide concept extraction (Iyer et al., 2024; Ting et al., 2025). We can now track which galaxies are discussed and cited most often and how these patterns vary across subfields and years.

The same literature-processing methodology could, in principle, extend beyond galaxies to stars, supernovae, exoplanets, and even theories of dark matter. It could also support similar databases beyond astronomy, drawing on bioRxiv for genes or on ChemRxiv for materials, building on precedents for gene-focused literature synthesis (Skarlinski et al., 2024) and structured extraction of materials information (Dagdelen et al., 2024).

5 Conclusion

We create GalaxyWiki, a database that brings together scientific interpretations of galaxies and discussions of what makes them noteworthy. It allows astronomers and the general public to check whether a galaxy has been discussed in the extracted literature, read those discussions, and continue to the relevant papers. GalaxyWiki illustrates how language models can make previously impractical scientific projects feasible by performing parallelizable and verifiable tasks. Future work will expand the corpus and evaluate mention identification, coordinate resolution, and grouping to build a more complete collection.

Limitations

Even though we can verify that the extracted quotes are in the original papers, we cannot ensure that we have retrieved the mentions of every galaxy in a given paper. It is likely that the language model misses the mentions of some galaxies during the extraction step. In fact, our setup guarantees that some galaxies have been missed since we set a limit of 30 galaxies per paper, which is hit for 39 papers. Notably, in one paper the language model recorded 250 galaxies despite this prompted limit, which we inspect and find minor errors in some extracted sentence counts.

Our coverage is also limited by the source corpus, because we use only the astro-ph subset available through MINT-1T (Awadalla et al., 2024). Papers without an arXiv version are also absent. Expanding the corpus in future work would help produce a more complete dataset.

In our programmatic NED resolution, a pattern intended to capture Abell designations (Abell, 1958) also matched local identifiers like "Galaxy A1", producing false positives. After identifying

the issue, we invalidated 112 of the 156 rows resolved with this pattern.

Acknowledgments

This work was enabled by credits provided through OpenAI's ChatGPT Futures award. We also acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) [funding reference number 721697856] and the Data Sciences Institute at the University of Toronto through grant number DSI-DSFY4R1P01. We thank Gabriel Pfaffman and James Garland for valuable discussions and helpful feedback that improved this work. This work used the NASA/IPAC Extragalactic Database and the SIMBAD database.

References

George O. Abell. 1958. The Distribution of Rich Clusters of Galaxies. *The Astrophysical Journal Supplement Series*, 3:211.

Anas Awadalla, Le Xue, Oscar Lo, Manli Shu, Hannah Lee, Etash Kumar Guha, Matt Jordan, Sheng Shen, Mohamed Awadalla, Silvio Savarese, Caiming Xiong, Ran Xu, Yejin Choi, and Ludwig Schmidt. 2024. MINT-1T: Scaling Open-Source Multimodal Data by 10x: A Multimodal Dataset with One Trillion Tokens. *arXiv e-prints*, arXiv:2406.11271.

K. C. Chambers, E. A. Magnier, N. Metcalfe, H. A. Flewelling, M. E. Huber, C. Z. Waters, L. Denneau, P. W. Draper, D. Farrow, D. P. Finkbeiner, C. Holmberg, J. Koppenhoefer, P. A. Price, A. Rest, R. P. Saglia, E. F. Schlafly, S. J. Smartt, W. Sweeney, R. J. Wainscoat, and 107 others. 2016. The Pan-STARRS1 Surveys. *arXiv e-prints*, arXiv:1612.05560.

Yixuan Cheng, Nolan Koblischke, and Jo Bovy. 2026. Identifying Scientifically Interesting Galaxies with Vision-Language Models. Accepted at the Machine Learning and the Physical Sciences 2026 workshop.

John Dagdelen, Alexander Dunn, Sanghoon Lee, Nicholas Walker, Andrew S. Rosen, Gerbrand Ceder, Kristin A. Persson, and Anubhav Jain. 2024. Structured information extraction from scientific text with large language models. *Nature Communications*, 15:1418.

Arjun Dey, David J. Schlegel, Dustin Lang, Robert Blum, Kaylan Burleigh, Xiaohui Fan, Joseph R. Findlay, Doug Finkbeiner, David Herrera, Stéphanie Juneau, Martin Landriau, Michael Levi, Ian McGreer, Aaron Meisner, Adam D. Myers, John Moustakas, Peter Nugent, Anna Patej, Edward F. Schlafly, and 142 others. 2019. Overview of the DESI Legacy Imaging Surveys. *The Astronomical Journal*, 157(5):168.

V. Etsebeth, M. Lochner, M. Walmsley, and M. Grespan. 2024. Astronomaly at scale: searching for anomalies amongst 4 million galaxies. *Monthly Notices of the Royal Astronomical Society*, 529(1):732–747.

Katie Frey and Alberto Accomazzi. 2018. The Unified Astronomy Thesaurus: Semantic Metadata for Astronomy and Astrophysics. *The Astrophysical Journal Supplement Series*, 236(1):24.

Ibrahim Habib, Kevin Li, Trung Nguyen, Gabriel A. Patron, Gaurav Sekar, Nolan Koblischke, Matthieu Le Lain, Shashwat Sourav, Aman Kumar, Kartheik G. Iyer, Joshua S. Speagle, Itai Sfaradi, Michael J. Smith, and Ioana Ciucă. 2026. The Great Chat in the Sky: Aligning Exotic Astronomical Data to Language with AstroBridge. Accepted at the Machine Learning and the Physical Sciences 2026 workshop.

E. Iodice. 2010. Polar Disk Galaxies as New Way to Study Galaxy Formation: the Case of NGC4650A. In *Hunting for the Dark: the Hidden Side of Galaxy Formation*, volume 1240 of *American Institute of Physics Conference Series*, pages 379–382. AIP.

E. Iodice. 2014. Narrow Polar Rings versus Wide Polar Ring/Disk Galaxies. *arXiv e-prints*, arXiv:1402.7325.

Kartheik Iyer. 2026. Lodestone: A living astronomy encyclopedia. `https://lodestone.wiki/`. Accessed 2026-09-20.

Kartheik G. Iyer, Mikaeel Yunus, Charles O'Neill, Christine Ye, Alina Hyk, Kiera McCormick, Ioana Ciucă, John F. Wu, Alberto Accomazzi, Simone Astarita, Rishabh Chakrabarty, Jesse Cranney, Anjalie Field, Tirthankar Ghosal, Michele Ginolfi, Marc Huertas-Company, Maja Jabłońska, Sandor Kruk, Huiling Liu, and 11 others. 2024. pathfinder: A Semantic Framework for Literature Review and Knowledge Discovery in Astronomy. *The Astrophysical Journal Supplement Series*, 275(2):38.

Nolan Koblischke, Liam Parker, Francois Lanusse, Jo Bovy, Irina Espejo, and Shirley Ho. 2026. Semantic Search for 100M+ Galaxy Images Using AI-generated Captions. *The Astrophysical Journal*, 1005(2):170.

L. Kuhn, M. Shubat, and P. Barmby. 2022. Comparing NED and SIMBAD classifications across the contents of nearby galaxies. *Monthly Notices of the Royal Astronomical Society*, 515(1):807–816.

Jakub Lála, Odhran O'Donoghue, Aleksandar Shtedritski, Sam Cox, Samuel G. Rodriques, and Andrew D. White. 2023. PaperQA: Retrieval-Augmented Generative Agent for Scientific Research. *arXiv e-prints*, arXiv:2312.07559.

Jinchu Li, Yuan-Sen Ting, Alberto Accomazzi, Tirthankar Ghosal, and Nesar Ramachandra. 2026. Predicting New Concept-Object Associations in Astronomy by Mining the Literature. *arXiv e-prints*, arXiv:2602.14335.

Chris J. Lintott, Kevin Schawinski, Anže Slosar, Kate Land, Steven Bamford, Daniel Thomas, M. Jordan Raddick, Robert C. Nichol, Alex Szalay, Dan Andreescu, Phil Murray, and Jan Vandenberg. 2008. Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, 389(3):1179–1189.

M. Lochner and B. A. Bassett. 2021. ASTRONOMALY: Personalised active anomaly detection in astronomical data. *Astronomy and Computing*, 36:100481.

T. S. Mankowski, S. J. Slater, and T. F. Slater. 2011. An Interpretive Study of Meanings Citizen Scientists Make When Participating in Galaxy Zoo. In *Earth and Space Science: Making Connections in Education and Public Outreach*, volume 443 of *Astronomical Society of the Pacific Conference Series*, page 150.

J. M. Mazzarella. 2020. Science Discovery With Diverse Multi-Wavelength Datasets Fused in the NASA/IPAC Extragalactic Database (NED). In *Astronomical Data Analysis Software and Systems XXIX*, volume 527 of *Astronomical Society of the Pacific Conference Series*, page 3.

Joseph M. Mazzarella and NED Team. 2017. Evolution of the NASA/IPAC Extragalactic Database (NED) into a Data Mining Discovery Engine. In *Astroinformatics*, volume 325 of *IAU Symposium*, pages 379–384.

B. J. McLean, G. R. Greene, M. G. Lattanzi, and B. Pirenne. 2000. The Status of the Second Generation Digitized Sky Survey and Guide Star Catalog. In *Astronomical Data Analysis Software and Systems IX*, volume 216 of *Astronomical Society of the Pacific Conference Series*, page 145.

F. Ochsenbein, P. Bauer, and J. Marcout. 2000. The VizieR database of astronomical catalogues. *Astronomy and Astrophysics Supplement Series*, 143:23–32.

Yasna Ordenes-Briceño, Iskren Y. Georgiev, Thomas H. Puzia, Paul Goudfrooij, and Magda Arnaboldi. 2016. Compact stellar systems in the polar ring galaxies NGC 4650A and NGC 3808B: Clues to polar disk formation. *Astronomy and Astrophysics*, 585:A156.

Wujun Shao, Rui Zhang, Pengli Ji, Dongwei Fan, Yaohua Hu, Xiaoran Yan, Chenzhou Cui, Yihan Tao, Linying Mi, and Lang Chen. 2024. Astronomical Knowledge Entity Extraction in Astrophysics Journal Articles via Large Language Models. *Research in Astronomy and Astrophysics*, 24(6):065012.

Michael D. Skarlinski, Sam Cox, Jon M. Laurent, James D. Braza, Michaela Hinks, Michael J. Hammerling, Manvitha Ponnapati, Samuel G. Rodriques, and Andrew D. White. 2024. Language agents achieve superhuman synthesis of scientific knowledge. *arXiv e-prints*, arXiv:2409.13740.

M. F. Skrutskie, R. M. Cutri, R. Stiening, M. D. Weinberg, S. Schneider, J. M. Carpenter, C. Beichman, R. Capps, T. Chester, J. Elias, J. Huchra, J. Liebert, C. Lonsdale, D. G. Monet, S. Price, P. Seitzer, T. Jarrett, J. D. Kirkpatrick, J. E. Gizis, and 12 others. 2006. The Two Micron All Sky Survey (2MASS). *The Astronomical Journal*, 131(2):1163–1183.

M. Spavone, E. Iodice, M. Arnaboldi, G. Longo, and O. Gerhard. 2011. Chemical abundances of the PRGs UGC 7576 and UGC 9796. I. Testing the formation scenario. *Astronomy and Astrophysics*, 531:A21.

George Stein, Peter Harrington, Jacqueline Blaum, Tomislav Medan, and Zarija Lukic. 2021. Self-supervised similarity search for large scientific datasets. *arXiv e-prints*, arXiv:2110.13151.

Yuan-Sen Ting, Alberto Accomazzi, Tirthankar Ghosal, Tuan Dung Nguyen, Rui Pan, Zechang Sun, and Tijmen de Haan. 2025. AstroMLab 5: Structured Summaries and Concept Extraction for 400,000 Astrophysics Papers. *arXiv e-prints*, arXiv:2511.12353.

Jia-Qi Wang, Jin-Qu Zhang, and Zhi-Qiang Zou. 2026. An Automatic Knowledge Graph Construction Method from Astronomical Literatures via Large Language Models. *Research in Astronomy and Astrophysics*, 26(3):035011.

M. Wenger, F. Ochsenbein, D. Egret, P. Dubois, F. Bonnarel, S. Borde, F. Genova, G. Jasniewicz, S. Laloë, S. Lesteven, and R. Monier. 2000. The SIMBAD astronomical database. The CDS reference database for astronomical objects. *Astronomy and Astrophysics Supplement Series*, 143:9–22.

Donald G. York, J. Adelman, John E. Anderson, Jr., Scott F. Anderson, James Annis, Neta A. Bahcall, J. A. Bakken, Robert Barkhouser, Steven Bastian, Eileen Berman, William N. Boroski, Steve Bracker, Charlie Briegel, John W. Briggs, J. Brinkmann, Robert Brunner, Scott Burles, Larry Carey, Michael A. Carr, and 126 others. 2000. The Sloan Digital Sky Survey: Technical Summary. *The Astronomical Journal*, 120(3):1579–1587.