

Report on the 1st Workshop on Building Innovative Research Systems for Digital Libraries (BIRDS 2025) at TPDL 2025

Christin Katharina Kreutz

TH Mittelhessen
Gießen, Germany
ckreutz@acm.org

Hermann Kroll

TU Braunschweig
Braunschweig, Germany
krollh@acm.org

Samy Ateia, Marc-Alexis Azaïs, Wolf-Tilo Balke, Jean-Baptiste de la Broise,
Liana Ermakova, Jaap Kamps, Hajra Klair, Ari Kukkaro, Sanna
Kumpulainen, Thomas Risse, Jian Wu*

Abstract

The first workshop on **B**uilding **I**nnovative **R**esearch Systems for **D**igital Libraries (BIRDS) took place at TPDL 2025 as a full-day workshop. BIRDS addressed practitioners working in digital libraries and GLAMs as well as researchers from computational domains. Our interdisciplinary workshop focused on connecting members of both worlds. One of today's biggest challenges is the increasing information flood. Large language model-based services like ChatGPT seem to offer good performance for answering questions on the web. So, shall we just build upon that idea and use chatbots in digital libraries? Or do we need to design and develop specialized and effective access paths? Answering these questions requires to connect different communities, practitioners from real digital libraries and researchers in the area of computer science. In brief, our workshop's goal was thus to support researchers and practitioners to build the next generation of innovative and effective digital library systems.

Date: 23 September 2025.

Website: <https://ws-birds.github.io/birds2025.github.io/>.

1 Introduction

Professionals and representatives of galleries, libraries, archives, and museums (GLAM) institutes typically have deep knowledge on their collections, metadata, users and their needs. At the same time these domain experts usually lack technical expertise and resources to implement practical solutions themselves. Conversely, researchers from computer science or adjacent areas usually have the corresponding computational background but are rarely aware of real-world challenges that exist in digital libraries.

*Affiliation not shown for all authors due to space limitations (see Appendix A for details).

Time	Program point
9:10 – 9:15	Welcome
9:15 – 10:00	Keynote by Jian Wu
10:00 – 11:00	Scientific speed-dating
11:00 – 11:30	Coffee break
11:00 – 11:30	General TPDF/ADBIS keynote
11:00 – 14:00	Lunch break
14:00 – 15:30	Panel discussion
15:30 – 16:00	Breakout group discussions pt. 1
16:00 – 16:30	Coffee break
16:30 – 17:00	Breakout group discussions pt. 2
17:00 – 17:30	Summary and Closing

Table 1. BIRDS schedule.

Users’ requirements for access paths to information offered by such digital libraries are domain dependant; see e.g., [Kreutz et al. \[2023\]](#); [Kroll et al. \[2024a, 2023, 2024b\]](#). Working with real use cases and data comes with restrictions inflicted by laws on data usage or privacy and GLAMs’ budgets. The goal of the BIRDS workshop [[Kreutz and Kroll, 2025](#)] was thus to bridge these two communities, fostering collaboration leading to innovative practical applications in digital libraries and discussing current challenges.

2 Schedule

Our BIRDS workshop at TPDF spanned three blocks over a whole day (see [Table 1](#)) and attracted 25 participants. In the first session of the day all attendees were **welcomed** before the workshop schedule was introduced. Afterwards, we were given a **keynote** by Jian Wu (see [Section 3](#)). The first block concluded with an hour-long **scientific speed-dating**. Here, the workshop participants were separated into two groups A and B. Members from group A were asked to remain seated throughout the speed-dating session while members of group B switched their conversation partners after each predefined time period. Every member from one group talked to all members of the other group while members did not talk to other members of the same group. Pairs of participants were asked to introduce themselves to each other and discuss their scientific background and interests.

We continued our workshop after the lunch break with an inspiring **panel discussion** (see [Section 4](#)). Immediately afterwards, we formed two groups for **breakout discussions** (see [Section 5](#)) that continued their discussions over coffee before we concluded the day with a **summary and closing** session.



Figure 1. Left: Keynote. Right: Panel discussion.

3 Keynote

The keynote (see left part of [Figure 1](#)) titled *Beyond Retrieval: A Vision of Digital Libraries in the Large Language Model Era* was given by Jian Wu (Old Dominion University, USA).

Since 2023, there has been a surge of public and research interest in large language models (LLMs), which has significantly shifted the paradigm of information retrieval from returning keyword-based search results to the generation of natural language responses. This shift brings both challenges and opportunities for traditional digital libraries, which have served as a core infrastructure for browsing, searching, and accessing scholarly content. A critical question emerges: What role should digital libraries play in this LLM era? In this keynote, Jian shared his vision of digital libraries in the LLM era. He argued that digital libraries are still indispensable, not only as repositories for digital preservation and provenance but also for trustworthy metadata discovery and verification. The talk explored how digital libraries can evolve by integrating LLMs and structured knowledge to support advanced services such as automatic data extraction, scholarly comparison, review generation, and science communication for broader audiences. Jian shared preliminary work in this direction, including initiatives on preserving endangered open-access datasets and software [[Wu et al., 2015](#)], complex table data extraction [[Ajayi et al., 2026](#)], scientific claim verification [[Koneru et al., 2024](#)], and assessing research reproducibility [[Rajtmajer et al., 2022](#); [Obadage et al., 2024](#)].

4 Panel Discussion

For the panel our keynote speaker Jian Wu was joined by Sanna Kumpulainen (Tampere University, Finland), Liana Ermakova (Université de Bretagne Occidentale, France), Wolf-Tilo Balke (TU Braunschweig, Germany) and Thomas Risse (Goethe University Frankfurt, University Library, Germany). Hermann Kroll moderated the panel.

Our panel discussion (see right part of [Figure 1](#)) was composed of opening statements followed by three debate statements which the panellists could comment on and it ended with a hypothetical scenario.

Jian	both a prototyper and an evaluator . I build prototypes of new intelligent modules that better serve DL users and evaluate these modules in automated way or via human studies.
Sanna	I teach digital libraries and do research about them.
Liana	PI of SimpleText, member of ISTEEX, PI of ENABLE, iSchool network, TPD L'25 demo co-chair [...]
Wolf-Tilo	bringing innovation to information provisioning. How do we get the information to the user and who is willing to pay for it?
Thomas	observer, consultant, innovator, proposal writer, project manager, business model developer, ... (but not developer)

Table 2. Panellists' completions of opening statement My role in the landscape of digital libraries or library digitization is....

4.1 Opening Statements

After a brief overview of the structure of the panel discussion, panellists introduced themselves by completing three opening statements that were sent to them beforehand regarding their roles (see Table 2), what excites them about working in or with digital libraries (see Table 3) and their perceived biggest challenge for the future of the domain (see Table 4).

4.2 Debates

The following three controversial and somewhat exaggerated debate statements had not been shared with the panellists beforehand.

Our community's mission is to build research prototypes, not production-ready digital library systems. The translation of research results into production is not our task.

The panel offered diverse perspectives on the topic. One proposition was for researchers to only publish ideas while letting others implement the corresponding systems. Another view was to strive to bring fundamental research and applied sciences closer together to conceptualise and implement systems jointly. Such a connection could ensure societal impact and is increasingly required by funding bodies.

Research prototypes cannot be directly deployed as production-ready systems. The overall requirement for employing research into reality is reproducibility. Transitioning to a production system introduces technical difficulty due to scalability, rewriting code, resources for running the system and the responsibility to keep it available. Maintaining a full system requires considerable time, organisation and human resources.

At the same time waiting until technology is fully mature might hinder innovation. Both groups, researchers envisioning prototypes and maintainers of production systems, should try to collaborate more. Industry is not competitive if it is not rooting its systems in current research advancements. Additionally, benchmarks to evaluate systems against do not reflect heterogeneous user groups. A chance for collaboration and overcoming the gap between research and application lies in studying users and how they interact with existing systems and their content to ground

Jian	DLs will not be just search engines but also libraries of data and knowledge hub. DLs will make it easier to obtain information through human AI interaction .
Sanna	There are so many viewpoints how to do research of, with, or for the digital libraries. My focus is human perception – that affects everything related to digital libraries: the contents, their organization and findability of information items, the presentation of results, interfaces and interaction methods, you name it!
Liana	Preserving and transferring cultural heritage , making knowledge and cultural heritage accessible , searchable and analysable, interdisciplinarity , contributing to innovation
Wolf-Tilo	The heterogeneity of content that needs to be dealt with and the insights into a vast variety of fields!
Thomas	- Learning about new scientific areas / collections and develop ideas how to better support the scientific work with any kind of content - Possibilities given by the broader definition of “libraries” (research data, research software, etc.) and the increasing openness of information

Table 3. Panellists’ completions of opening statement *This excites me most about working in/with digital libraries or library digitization...*

design and functional decisions. Such an approach would also prevent spending time on features which are not useful for a system’s actual audience and help research not having to having to implement prototypes from scratch to study user behaviour.

Our digital libraries’ features and strategic orientation should be completely based on our users’ requirements.

Panellists agreed that while completely following users’ requirements is not feasible, following their requirements in general is desirable. We strive to provide users novel ways of interacting with content but users are often not able to formulate what they want and only have limited imagination. Only few users actively voice their complaints or wishes. So, solely basing new features on users’ input could reflect the opinion of a minority and come with the danger of incremental instead of disruptive innovation.

The question arose whether AI is the user of a digital library. Human users cannot keep up with the growing volume of publications due to limited memory and processing speed. LLMs are better at managing both. Some existing digital libraries, such as arXiv, provide their content in machine-readable formats, as chatbots are already their users today. In this context, it was questioned why authors still write full texts rather than abstracts only.

With the popularity and potence of LLMs we do not need digital libraries any more. Chatbots will replace digital libraries.

Digital libraries offer more than publications, they provide numerous artefacts such as digital models or embodied information that are difficult to integrate in an LLM. They do not only contain facts but also enable access to highly contextualised or curated content.

Jian	(1) financial interest . Funding agencies are more interested in national security, defense, and social economy, but are less interested in investing on digital infrastructures to support education and research using AI. (2) high quality data for training and evaluation. Crowdsourced data are noisy. Synthesized data are less representative. Expert labels are not scalable.
Sanna	funding and sustainability for some specific collections. Also, even that standardization is a good thing in general, one size does not fit all, necessarily
Liana	access to the multimedia content, meta-data and discoverability, intellectual property , AI plagiarism detection, deepfake, etc., different biases , including representation bias, polarization, fewer people think that it is useful to access primary sources
Wolf-Tilo	the elephant in the room: LLMs. Do we still need DLs for information discovery in future or will we just ask some LLM..?!
Thomas	- Available resources (money, people) - Available time - Bureaucracy (e.g. all the EU acts like GDPR, Resilience Act, AI Act)

Table 4. Panellists’ completions of opening statement **The biggest challenges for the future of digital libraries or library digitization are....**

With the culture of publish or perish, prioritising quantity over quality researchers are incentivised to publish more. This leads to the problem of humans no longer being able to read all the papers. This process of over-accumulating information is a cyclic occurrence that has already been an issue thirty years ago when switching to automatic lookup instead of manual search by librarians.

Despite LLMs giving users speed, full documents give them depth. While LLMs excel in quick answers, summaries, and short explanations, users need to have knowledge to formulate their questions. Full documents provide information and perspectives that are overlooked inherently by LLMs targeting short answers. Full documents are curated texts with a specific message, while LLMs aim to provide broad but surface-level answers. Full documents, including books and scientific articles, expose a user to unexpected ideas and a variety of styles, which is inherently impossible in LLMs.

If LLMs take over the whole research life cycle, even the component of peer review, we run into the problem of trust. Applied over digital libraries, LLMs summarise entire collections and these summaries lose lots of information. To verify chatbots’ outputs we still require original texts.

A less controversial application of LLMs in a digital library could be the first point of entry. A retrieval-augmented generation (RAG) system could replace a digital librarian in trying to help specify or define users’ visceral information needs (see Taylor [1962]). If a chatbot then points towards potentially relevant information instead of summarising it, a digital library would become more accessible.

In general, it is important to consider the dangers of using LLMs in the context of digital libraries. Chatbots lower the critical thinking required to answer questions and could hinder or prevent students’ learning processes. It would be desirable for chatbots to pose questions to users to force them to reflect their thinking.

4.3 Hypothetical Scenario

We asked our panellists to assume the following scenario: `The director of our digital library wants to incorporate a chatbot. What would be your course of action?`

The first step would be to ask for money to enable the project. Next, computational requirements should be assessed to ensure a suitable performance of the chatbot. Additionally, security should be considered as a chatbot should not disregard copyright and ethical considerations. Ensuring these properties can only be assessed via extensive evaluations.

A user-centred approach is important to determine why users would want a chatbot as requirements change with the provision of different assess paths. Here, existing research or benchmarks on desirable and undesirable components could be helpful in composing a framework modified for the current case.

In general, it is important to determine the purpose of incorporating the chatbot: the goal of commercialising a RAG-style chatbot comes with a different considerations, including availability and deployment, compared to offering it to users for research purposes.

5 Breakout Group Discussions

5.1 Group 1: Evaluating digital library systems

The **first session** of this group focused on discussing the case of a participant in which they described their current problem of evaluating summarisation and categorisation of electronic theses and dissertations.

Assessment of summaries and categories should be conducted by domain experts from the individual areas, not by a small group of non-domain experts. A pilot user study could try to employ authors of theses as assessors of summaries and categories. Here, incentives for human assessors could help increase participation. Categories can be evaluated against ground truth using metrics such as precision, recall, and accuracy. However, when the ground truth is more subjective, defining the *goodness* of categories (like a *great* summarisation) becomes less straightforward.

For instance, fully automating the pipeline—from category extraction to classification and evaluation—can create a sense of lost control for practitioners, who generally prefer to refine and steer both the categories and the assignments based on their domain expertise [Schofield et al., 2025]. More pragmatically, the trade-offs between classical approaches such as LDA and LLM-based methods—in terms of human and budgetary costs, processing time, and knowledge gain—are not clearly favourable to LLMs, with LDA remaining a free and sustainable tool [Li et al., 2025]. Nevertheless, LLM-based evaluation becomes increasingly popular, supported by frameworks that simplify the deployment of the LLM-as-a-judge paradigm [Kim et al., 2024].

It will become increasingly important to consider the target audience for whom the summary is created, in this specific case it is laypeople. Therefore, this summarisation can be considered a specific type of text simplification for non-experts (see e.g., the SimpleText lab at CLEF [Ermakova et al., 2025]).

The **second session** of this group discussion after the coffee break stayed on a more conceptual level. The question arose if it is enough to assess if users are happy with our digital library system and to compare against what users intend to achieve. Users misunderstanding our access paths

to information indicates room for improvement. But if we only have a small community of users we are not able to identify the long tail of problems as obscure materials or eclectic literature is only rarely accessed.

The issue of funding, especially for digitizing content that is currently stored away in boxes in institutes or libraries, was brought up in context with the trend towards funding breakthrough research rather than maintenance. This again links back to our panel discussion on research prototypes vs. production-ready systems.

5.2 Group 2: Designing user-centered digital library systems

The second group focused its discussion on the gap between what digital libraries could offer and what users really want. In the discussion phase, a quote from Henry Ford was cited:

If I had asked people what they wanted, they would have said faster horses.

On the one hand, digital libraries need and want to move forward by providing effective access paths to users. This way, they strive to create something new and useful. On the other hand, digital libraries might not know how useful access paths could look like. Asking users then seems to be the first choice. However, users of digital libraries might not know what is possible with today's technology. They are familiar with the existing access paths of a digital library and thus might rather suggest smaller improvements (evolution rather than revolution). However, to really progress and build the next generation of digital library systems, novel interfaces combined with effective and state-of-the-art methods are worth exploring. Beyond that, asking a larger number of users may result in many different ideas. Distilling what would be really worth exploring is not always easy and can hinder the development.

Solutions could be using prototyping, hackathons, and vibe-coding. These methods allow digital libraries to co-create useful access paths together with their users. Researchers and developers can then show users what is possible with today's technology. Users can directly explain to developers what they really need and what challenges they face when using existing systems. Still, digital libraries will rather create prototypes than production-ready systems. For instance, data restrictions might get ignored in a prototype phase, but hinder the development of production-ready systems in the end. Thus, compromises have to be made.

In summary, the development of novel digital library systems must be built upon a close cooperation between developers, researchers, and users.

6 Reflection and Summary

6.1 Reflection

While participants of the workshop expressed their initial scepticism towards the scheduled scientific speed-dating, they were pleasantly surprised by the session afterwards. As a recommendation for further iterations it could help to let attendees of the workshop know in advance what such a program point consists of. Additionally it would be desirable to have the opportunity to scientifically speed-date all other participants, not only those that were in the opposing group. Regarding the allocated time for discussions of pairs we experienced six minutes being a good length such that

pairs got to introduce themselves and start talking more in-depth about their interests without running out of discussion points.

For breakout group discussions guiding questions, a more narrow scope and smaller groups would help participants engage more. Our group 1 had more than 10 members, which led to reluctance to participate actively, especially among members with lower academic seniority.

We identified funding and the decreasing funding opportunities for research and work in the field of digital libraries as one of the overarching shared challenges throughout our discussions.

6.2 Summary

Our BIRDS workshop tackled the question of how to design innovative research systems for digital libraries. The best answer is to support a close cooperation between developers, researchers, and users. Opportunities lie in the usage of today's technology, e.g., as demonstrated by the usage of LLMs in our keynote. Our panellists agreed that while chatbots are useful, they will not replace digital libraries at all. Answers of chatbots can be pure hallucinations, and that is why they need to be rechecked/fact-checked. And here, digital libraries provide well-curated collections, and thus, trust, which is not offered by using pure LLMs. While trust is a central argument for digital libraries, systems need to move forward. Novel and effective access paths are desired, but developing them remains challenging: The design of user-centered digital library systems comes with issues like users not knowing what is possible today and developers not knowing what is beneficial for users. And then, evaluating these systems requires practical definitions on how to measure the usefulness of new systems. Beyond that, challenges such as data restrictions, limited funding, limited computation power, missing human resources, and expertise are common.

A Authors and Affiliations

Workshop organisers:

- Christin Katharina Kreutz; TH Mittelhessen - University of Applied Sciences & Herder Institute, Germany; ckreutz@acm.org; <https://orcid.org/0000-0002-5075-7699>
- Hermann Kroll; TU Braunschweig, Germany; krollh@acm.org; <https://orcid.org/0000-0001-9887-9276>

Keynote speaker and panellists:

- Wolf-Tilo Balke; TU Braunschweig, Germany; balke@ifis.cs.tu-bs.de; <https://orcid.org/0000-0002-5443-1215>
- Liana Ermakova; HCTI, Université de Bretagne Occidentale, France; liana.ermakova@univ-brest.fr; <https://orcid.org/0000-0002-7598-7474>
- Sanna Kumpulainen; Tampere University, Finland; sanna.kumpulainen@tuni.fi; <https://orcid.org/0000-0002-7016-257X>
- Thomas Risse; Goethe University Frankfurt, University Library, Germany; t.risse@ub.uni-frankfurt.de; <https://orcid.org/0000-0001-6248-1709>
- Jian Wu; Old Dominion University, USA; jwu@cs.odu.edu; <https://orcid.org/0000-0003-0173-4463>

Other authors:

- Samy Ateia; University of Regensburg, Germany; Samy.Ateia@ur.de; <https://orcid.org/0009-0000-2622-9194>
- Marc-Alexis Azais; La Rochelle University, France; marc-alexis.azais@univ-lr.fr; <https://orcid.org/0009-0007-3812-1341>
- Jean-Baptiste de la Broise; MDPI, Switzerland; jeanbaptiste.delabroise@mdpi.com; <https://orcid.org/0009-0001-2048-4888>
- Jaap Kamps; University of Amsterdam, The Netherlands; kamps@uva.nl; <https://orcid.org/0000-0002-6614-0087>
- Hajra Klair; Virginia Tech, USA; khajra23@vt.edu; <https://orcid.org/0009-0008-2428-7401>
- Ari Kukkaro; Tampere University, Finland; ari.kukkaro@tuni.fi; <https://orcid.org/0009-0005-8461-1873>

References

- Kehinde Ajayi, Yi He, and Jian Wu. Uncertainty-aware complex scientific table data extraction. In *ICDAR '25*, pages 56–73, 2026. URL <https://doi.org/10.1145/3641525.3663628>.
- Liana Ermakova, Hosein Azarbondy, Jan Bakker, Benjamin Vendeville, and Jaap Kamps. Overview of the CLEF 2025 simpletext track - simplify scientific text (and nothing more). In *CLEF '25*, volume 16089, pages 436–463, 2025. URL https://doi.org/10.1007/978-3-032-04354-2_23.
- Seungone Kim, Juyoung Suk, Shayne Longpre, Bill Yuchen Lin, Jamin Shin, Sean Welleck, Graham Neubig, Moontae Lee, Kyungjae Lee, and Minjoon Seo. Prometheus 2: An open source language model specialized in evaluating other language models. In *EMNLP '24*, pages 4334–4353, 2024. URL <https://aclanthology.org/2024.emnlp-main.248/>.
- Sai Koneru, Jian Wu, and Sarah Rajtmajer. Can large language models discern evidence for scientific hypotheses? case studies in the social sciences. In *LREC-COLING '24*, pages 2787–2797, 2024. URL <https://aclanthology.org/2024.lrec-main.248/>.
- Christin Katharina Kreutz and Hermann Kroll. First workshop on building innovative research systems for digital libraries (BIRDS 2025). *CoRR*, abs/2509.26001, 2025. URL <https://doi.org/10.48550/arXiv.2509.26001>.
- Christin Katharina Kreutz, Martin Blum, Philipp Schaer, Ralf Schenkel, and Benjamin Weyers. Evaluating digital library search systems by using formal process modelling. In *JCDL '23*, pages 1–12, 2023. URL <https://doi.org/10.1109/JCDL57899.2023.00058>.
- Hermann Kroll, Christin Katharina Kreutz, Pascal Sackhoff, and Wolf-Tilo Balke. Enriching simple keyword queries for domain-aware narrative retrieval. In *JCDL '23*, pages 143–154, 2023. URL <https://doi.org/10.1109/JCDL57899.2023.00029>.

-
- Hermann Kroll, Christin Kreutz, Mathias Jehn, and Thomas Risse. Requirements for a digital library system: A case study in digital humanities (technical report). *CoRR*, abs/2410.22358, 2024a. URL <https://doi.org/10.48550/ARXIV.2410.22358>.
- Hermann Kroll, Jan Pirklbauer, Jan-Christoph Kalo, Morris Kunz, Johannes Ruthmann, and Wolf-Tilo Balke. A discovery system for narrative query graphs: entity-interaction-aware document retrieval. *Int. J. Digit. Libr.*, 25(1):3–24, 2024b. URL <https://doi.org/10.1007/S00799-023-00356-3>.
- Zongxia Li, Lorena Calvo-Bartolomé, Alexander Hoyle, Paiheng Xu, Daniel Stephens, Alden Dima, Juan Francisco Fung, and Jordan Boyd-Graber. Large language models struggle to describe the haystack without human help: A social science-inspired evaluation of topic models. In *ACL '25*, pages 7583–7604, 2025. URL <https://aclanthology.org/2025.acl-long.375/>.
- Rochana R. Obadage, Sarah M. Rajtmajer, and Jian Wu. Short: Can citations tell us about a paper’s reproducibility? a case study of machine learning papers. In *ACM REP '24*, pages 96–100, 2024. URL <https://doi.org/10.1145/3641525.3663628>.
- Sarah Rajtmajer, Christopher Griffin, Jian Wu, Robert Fraleigh, Laxmaan Balaji, Anna Squicciarini, Anthony Kwasnica, David Pennock, Michael McLaughlin, Timothy Fritton, Nishanth Nakshatri, Arjun Menon, Sai Ajay Modukuri, Rajal Nivargi, Xin Wei, and C. Lee Giles. A synthetic prediction market for estimating confidence in published work. *AAAI '22*, 36(11): 13218–13220, 2022. URL <https://doi.org/10.1609/aaai.v36i11.21733>.
- Alexandra Schofield, Siqi Wu, Theo Bayard de Volo, Tatsuki Kuze, Alfredo Gomez, and Sharifa Sultana. ”my very subjective human interpretation”: Domain expert perspectives on navigating the text analysis loop for topic models. *Proc. ACM Hum. Comput. Interact.*, 9(1):1–30, 2025. URL <https://doi.org/10.1145/3701201>.
- Robert S. Taylor. The process of asking questions. *American Documentation*, 13(4):391–396, 1962. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.5090130405>.
- Jian Wu, Kyle Mark Williams, Hung-Hsuan Chen, Madian Khabsa, Cornelia Caragea, Suppawong Tuarob, Alexander G. Ororbia, Douglas Jordan, Prasenjit Mitra, and C. Lee Giles. Citeseerx: Ai in a digital library search engine. *AI Magazine*, 36(3):35–48, 2015. URL <https://doi.org/10.1609/aimag.v36i3.2601>.