Machine Learning, Text Summarization, and Optimizing Scholarship for Citizen Audiences and Discovery

Leila Sterman
Jason A. Clark
Justin Shanks
Daniel Laden

Montana State University
Coalition for Networked Information, Spring 2022
What’s Ahead

- Research Motivation
- Survey
  - How do scholars think about outreach and translation?
- Summarizing and restructuring scholarship
  - Optimizing Scholarship for Citizen Audiences and Discovery
  - Challenges and Opportunities
- Research Implications
  - Interdisciplinary communication
Research Motivation
Access is essential to understanding.
Translation takes effort.
First, we wanted to understand the current practices and motivations.

Survey questions
Survey responses
Survey
Evaluating the Utility of Species Distribution Models in Informing Climate Change-Resilient Grassland Restoration Strategy

Nicholas J. Lyon*, Diane M. Debinski1,2 and Imitiaz Rangwala3,4,5

1 Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, United States, 2 Department of Ecology, Montana State University, Bozeman, MT, United States, 3 Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, United States, 4 Physical Sciences Division, Earth System Research Laboratory (ESRL), Boulder, CO, United States, 5 Department of Interior North Central Climate Science Center, Fort Collins, CO, United States

Tallgrass prairie ecosystems in North America are heavily degraded and require effective restoration strategies if prairie specialist taxa are to be preserved. One common management tool used to restore grassland is the application of a seed-mix of native prairie plant species. While this technique is effective in the short-term, it is critical that species’ resilience to changing climate be evaluated when designing these mixes. By utilizing species distribution models (SDMs), species’ bioclimatic envelopes—and thus the geographic area suitable for them—can be quantified and predicted under various future climate regimes, and current seed-mixes may be modified to include more climate resilient species or exclude more affected species. We evaluated climate response on plant functional groups to examine the generalizability of climate response among species of particular functional groups. We selected 14 prairie species representing...
"How do you target your intended audience?"

"I probably don't do as good a job of targeting the uptake audience because my job counts productivity as scientific articles + grant dollars"  

"I just expect researchers to find my papers"
Understanding requires translation.
"Which audience would you like to reach and do not currently?"

"... there are so many publications it is even difficult to keep up with ... one field, so articles/podcasts, or other media outlets that highlight our research efforts help it gain more visibility"
Interdisciplinarity takes effort.
"What impact do you think your research has?"

"On a scholarly level, builds new knowledge within specific niches; cross-fertilizes between fields, opens up new avenues of discussion; reveals the historical roots of current issues (e.g., refugee crises); preserves historical materials for future generations; brings "hidden stories" to public light; educates the public; contributes to peace and reconciliations processes in post-trauma societies."

- "it moves ideas around within whatever academic field it lands in"
The ideal

"Often our peer-reviewed research is then crafted into a policy brief that is disseminated to the government officials and policymakers. Additionally, our communication administrator crafts new articles highlighting the topic. Lastly, economists present their work to the public."

Reality

- "I publish in reputable journals"
How might we create a “snapshot” of a scholarly article - a concise and readable summarization of the thesis, methods, findings, major figures of an article - for a broad, reading audience.
Research Motivation - Access

There are a number of tools in this area:

- **Scholarcy**
- **GROBID**
- **ScienceBeam**
- **CERMINE**
- **ContentMine**
- **The up-goer text editor**
Applying the Survey
21. If you had to select certain sections of your articles to summarize, which sections would you prioritize?

☐ Abstract
☐ Introduction
☐ Literature review
☐ Methods
☐ Results
☐ Discussion
☐ Limitations
☐ Further Study
☐ Other

[Other section input field]
New forms for Scholarship
Evaluating the Utility of Species Distribution Models in Informing Climate Change-Resilient Grassland Restoration Strategy

Nicholas J. Lyon*, Diane M. Debinski1,2 and Imtiaz Rangwala3,4,5

1 Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, United States; 2 Department of Ecology, Montana State University, Bozeman, MT, United States; 3 Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, United States; 4 Physical Sciences Division, Earth System Research Laboratory (NOAA), Boulder, CO, United States; 5 Department of Interior North Central Climate Science Center, Fort Collins, CO, United States

**OPEN ACCESS**

**Edited by:** Anouschka Th. Hof, Wageningen University & Research, Netherlands

**Reviewed by:** Bernd Panissidi, Laimburg Research Centre, Italy; Jaime A. Collazo, North Carolina State University, United States

*Correspondence:* Nicholas J. Lyon njlyon@iastate.edu

Tallgrass prairie ecosystems in North America are heavily degraded and require effective restoration strategies if prairie specialist taxa are to be preserved. One common management tool used to restore grassland is the application of a seed-mix of native prairie plant species. While this technique is effective in the short-term, it is critical that species resilience to changing climate be evaluated when designing these mixes. By utilizing species distribution models (SDMs), species' bioclimatic envelopes—and thus the geographic area suitable for them—can be quantified and predicted under various future climate regimes, and current seed-mixes may be modified to include more climate resilient species or exclude more affected species. We evaluated climate response on plant functional groups to examine the generalizability of climate response among species of particular functional groups. We selected 14 prairie species representing...
Minimum Viable Unit for Expressing Scholarship
What is a Nanopublication?

A nanopublication is the smallest unit of publishable information. This information can be about anything, for example a relation between a gene and a disease or an opinion. Nanopublications are fully expressed in a formal and machine-interpretable way. With nanopublications, it is possible to disseminate individual data as independent publications with or without an accompanying research article. Furthermore, because nanopublications can be attributed and cited, they provide incentives for researchers to make their data available in standard formats that drive data accessibility and interoperability. Nanopublications have the following general structure:
Summarizing Research – Creating a "snapshot" article

Example encoded as RDFa embedded in HTML.

```html
<!-- An article citation in MLA format, using a 'flat' approach that simplifies markup by not specifying an explicit relationship between the periodical, volume, and issue -->
<div vocab="https://schema.org/" typeof="ScholarlyArticle">
  <span property="author">Carlyle, Allyson.</span>
  "Understanding FRBR as a Conceptual Model: FRBR and the Bibliographic Universe".
  <div property="isPartOf" typeof="Periodical">
    <em>Library Resources and Technical Services</em>  
  </div>
  v. <span property="volumeNumber">50</span>,  
  <span property="issueNumber">4</span> 
  (<time datetime="2006-10" property="datePublished">October 2006</time>): 
</div>
```

Print.
Scripting and Programming

Summarization
Summarizing Research – Machine Processes

Data Mining

Harvesting
+
Normalizing

Corpus for analysis
- PubMed (STEM)
- Web of Science (STEM)
- IEEE Xplore (STEM)
- Academic Search Complete (Social Sciences, Humanities)
- JSTOR (Social Sciences, Humanities)
Data Mining

Harvesting
+
Normalizing

CERMINE Java Library

scripts/Cermine-Cleanup.py
Summarizing Research – Machine Processes

Text Processing + Analysis

Natural Language Processing + Evaluating Language Accessibility

Snapshot article
Text Processing + Analysis

Natural Language Processing +
POS tagging, Vectors for ranking sentences

Python Natural Language Toolkit (NLTK)

/scripts/Sumzor.py
Summarizing Research – Creating a "snapshot" article


ScholarlyArticle, NewsArticle
Summarizing Research – Creating a "snapshot" article

Example encoded as RDFa embedded in HTML.

<!-- An article citation in MLA format, using a 'flat' approach that simplifies markup by not specifying an explicit relationship between the periodical, volume, and issue -->
<div vocab="https://schema.org/" typeof="ScholarlyArticle">
  <span property="author">Carlyle, Allyson.</span>
  "Understanding FRBR as a Conceptual Model: FRBR and the Bibliographic Universe".
  <div property="isPartOf" typeof="Periodical">
    <em>Library Resources and Technical Services</em>
  </div>
  v. 50, no. 4 (October 2006):
  264–273.
</div>
2 to 3 Sentence Summary

Readability Score of 60

Search Engine Optimization
Evaluating the Utility of Species Distribution Models in Informing Climate Change-Resilient Grassland Restoration Strategy

Nicholas J. Lyon 1*, Diane M. Dabinski 1,2 and Imtiaz Rangwala 3,4,5

1 Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, United States, 2 Department of Ecology, Montana State University, Bozeman, MT, United States, 3 Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, United States, 4 Physical Sciences Division, Earth System Research Laboratory (NOAA), Boulder, CO, United States, 5 Department of Interior North Central Climate Science Center, Fort Collins, CO, United States

OPEN ACCESS

Edited by:
Anouschka R. Hof, Wageningen University & Research, Netherlands

Reviewed by:
Benedikt Panasesti, Limsburg Research Centre, Italy
Jaime A. Colazo, North Carolina State University, United States

*Correspondence:
Nicholas J. Lyon
nlyon@iastate.edu

Tallgrass prairie ecosystems in North America are heavily degraded and require effective restoration strategies if prairie specialist taxa are to be preserved. One common management tool used to restore grassland is the application of a seed-mix of native prairie plant species. While this technique is effective in the short-term, it is critical that species’ resilience to changing climate be evaluated when designing these mixes. By utilizing species distribution models (SDMs), species’ bioclimatic envelopes—and thus the geographic area suitable for them—can be quantified and predicted under various future climate regimes, and current seed-mixes may be modified to include more climate resilient species or exclude more affected species. We evaluated climate response on plant functional groups to examine the generalizability of climate response among species of particular functional groups. We selected 14 prairie species representing...
Evaluating the Utility of Species Distribution Models in Informing Climate Change-Resilient Grassland Restoration Strategy

Summary: In this study, the authors use generated scenarios to map the change of habitat for plant species in the prairie. They have two main scenarios that predict difference in temperature, plant species, as well as CO2 emissions. They find that both models predict a decline in habitat by 2040 and by 2080 all species will decline. Due to these findings we can find the plant species most at risk from climate change and consider steps forward in protecting them.

Important Sentences: The A1B scenario predicts an increase in temperature from 1.4 to 6.4 °C, whereas the A2 scenario predicts temperature increases from 2 to 5.4 °C and much greater CO2 emissions than the A1B scenario. However, both models predict the majority of species will experience declines in habitat by 2040. Models run using the A2 scenario predict declines in habitat for just four species by 2040, but models predict that by 2080, habitat suitability will decline for all species.

Keyword(s): seed-mixtures functional group climate change species distribution (niche) model bioclimatic envelope
Optimizing Scholarship for Citizen Audiences and Discovery

Accessible Research
Translating Concepts
Findable Research
Challenges + Opportunities
Challenges + Opportunities

Challenges...

Evaluating the NLP model
- Are the summarizations correct?
- Is it finding important sentences?
- Can we adjust reading levels?

Data sources for articles
Opportunities...

New partnerships
• Applying summarization and markup model to MSU Research News

Benchmarking the snapshot article
• Impacts and reach
• Referrals to original research
Research Implications
Research Implications + Next Steps

Implications...

Changing how the library is viewed or integrated into the research enterprise.

Understanding how or why scholars summarize their work.

Creating interdisciplinary communication through simplifying research expression.
Positioning the library as a partner throughout the research process—acknowledging that scholarship is an ongoing conversation.
**Thank you**

Jason A. Clark
Leila Sterman
Daniel Laden
References and Follow-up Resources

MSU Researchers - Article Summarization Practices and Preferences
• [Survey questions](#)
• [Survey responses](#)

MSU Article Summarization - Working Code and Supporting Documents
• [https://github.com/msulibrary/msu-article-summarization](#)


Article Summarization Development - Tools and Getting Started

- Nanopublications
  - [https://nanopub.org/wordpress/](https://nanopub.org/wordpress/)

- Readability Testing
  - [https://www.online-utility.org/english/readability_test_and_improve.jsp](https://www.online-utility.org/english/readability_test_and_improve.jsp)

- Understand Text Summarization and Create Your own Summarizer in Python
  - [https://towardsdatascience.com/understand-text-summarization-and-create-your-own-summarizer-in-python-b26a9f09fc70](https://towardsdatascience.com/understand-text-summarization-and-create-your-own-summarizer-in-python-b26a9f09fc70)