

CelloGraph – A Knowledge Graph System for Cellulosic Materials

Umayer Reza¹, Dr. Jinwu Wang², Dr. Torsten Hahmann¹

¹ *School of Computing and Information Science, University of Maine*

² *Forest Products Laboratory, U.S. Forest Service*

APPTI/TAPPI Meeting, Madison, Wisconsin

July 21, 2026



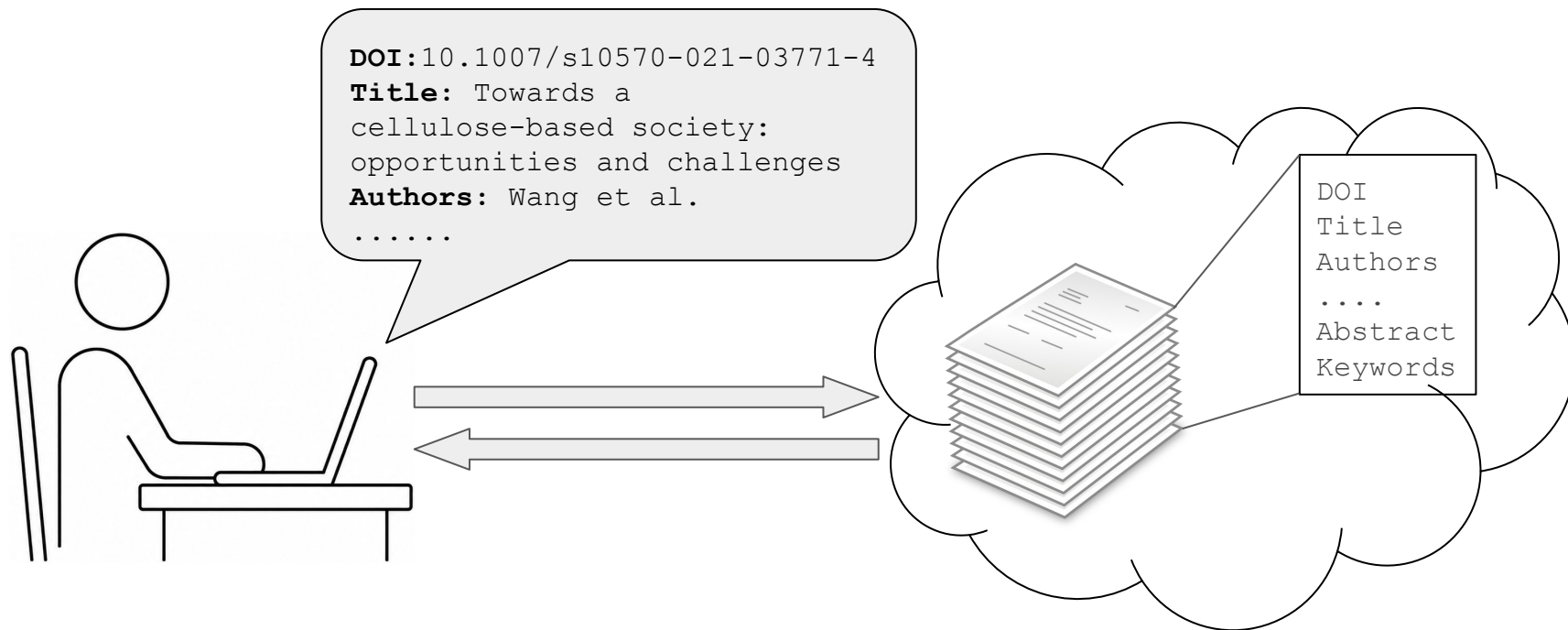
Forest Service
U.S. DEPARTMENT OF AGRICULTURE

Keeping Up-to-date with Scholarly Publications: An Ever-Increasing Challenge

- Number of publications produced per year grows 5-9% every year
- For example the number of publications in Science and Engineering published *per year* grew from 2.1M in 2012 to 3.3M in 2022
Source: <https://ncses.nsf.gov/pubs/nsb202333/publication-output-by-region-country-or-economy-and-by-scientific-field>
- Challenges of massive growth:
 - Finding relevant new knowledge about materials, processes, and properties across a fast-growing literature
 - Researchers need precise links between materials, processes, properties, and supporting publications
 - This imposes a high barrier for new or interdisciplinary researchers

Metadata for Searching Scholarly Publications

- To increase findability, publications are now commonly tagged with metadata



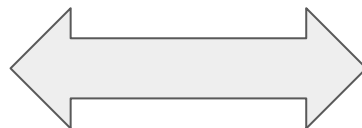
More Granular Semantic Search is Helpful (and Needed)

It remains difficult to

- Search publication content at a level of specific terms (e.g. materials, processes, properties) rather than only by title, abstract, or keywords
- Identify specific materials, properties, processes, and their relationships in the exact publications and paragraphs in where they are discussed

For example:

Which **cellulosic materials** are mentioned in conjunction with **permeability**? Where (in which publications and paragraphs)?

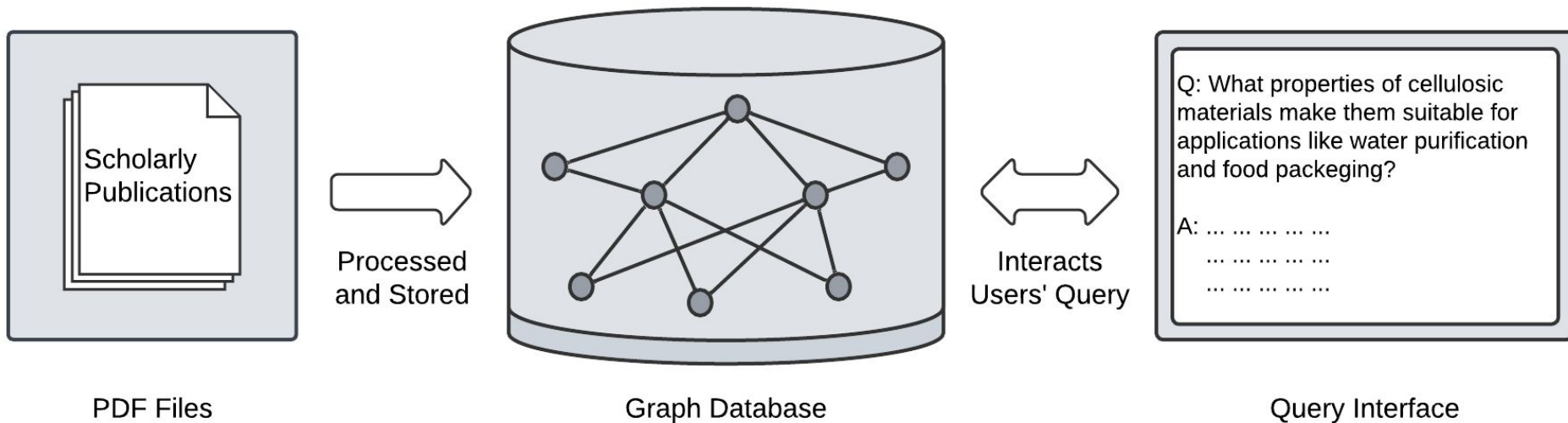


Publications

Vision: A Knowledge Graph System for Cellulosic Materials

A data-driven, knowledge-based system with

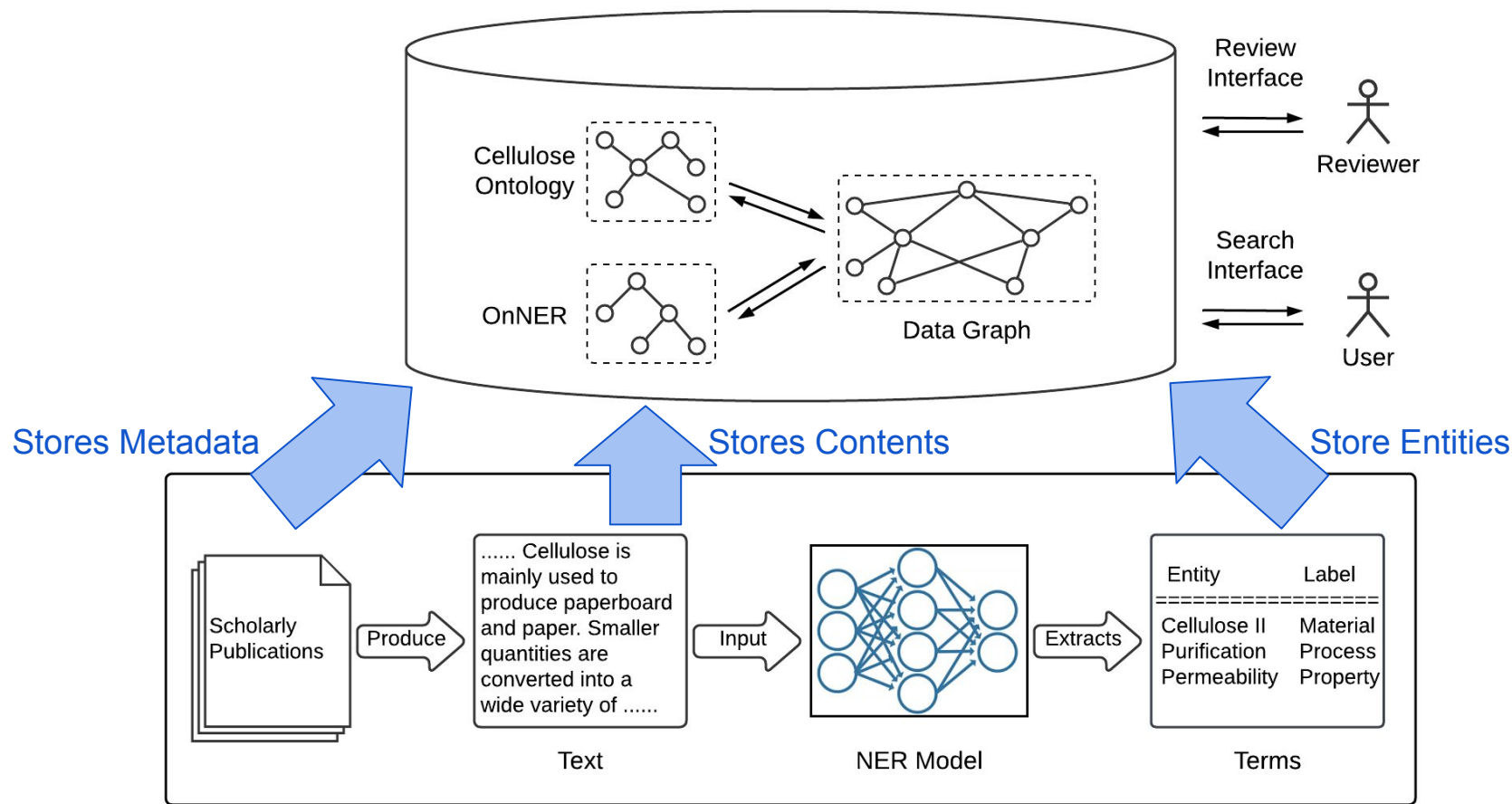
- a **knowledge graph** at its core
- ontologies for structuring the knowledge graph



Vision: Key Non-functional Requirements of the System

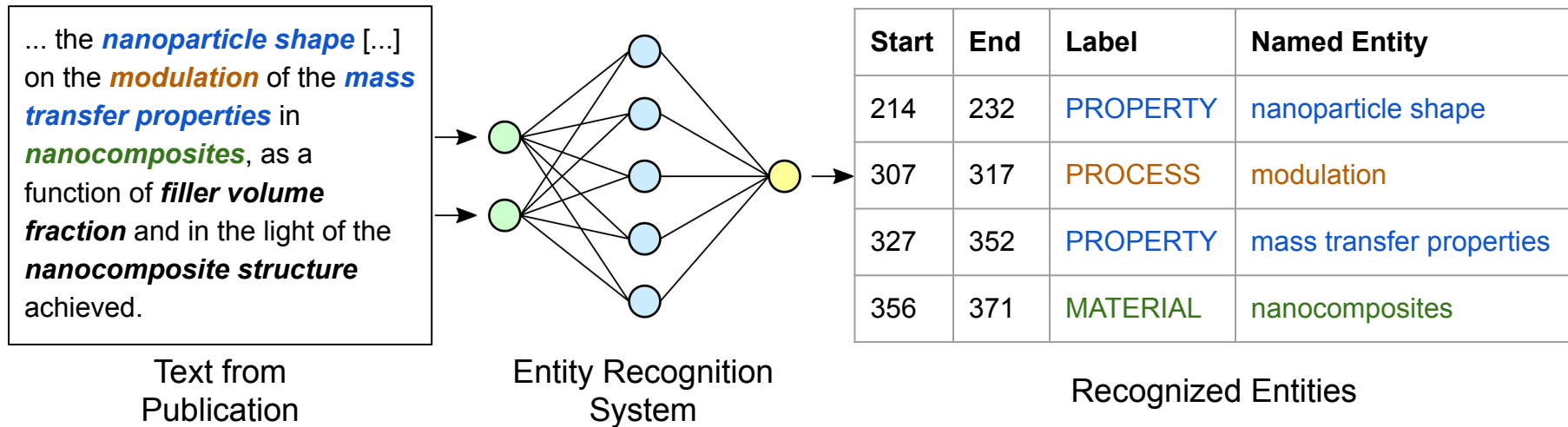
- **Scalability:** continuous accumulation of information with minimal manual intervention
- **Adaptability:** ability to support refinement of knowledge and ontological changes
- **Stability:** no major redesign of the knowledge graph
- **Provenance:** a detailed record of added information and its origin
- **Accuracy and Trustworthiness:** incorporating a human-in-the-loop approach with provenance to enhance trustworthiness

CelloGraph: Key Components



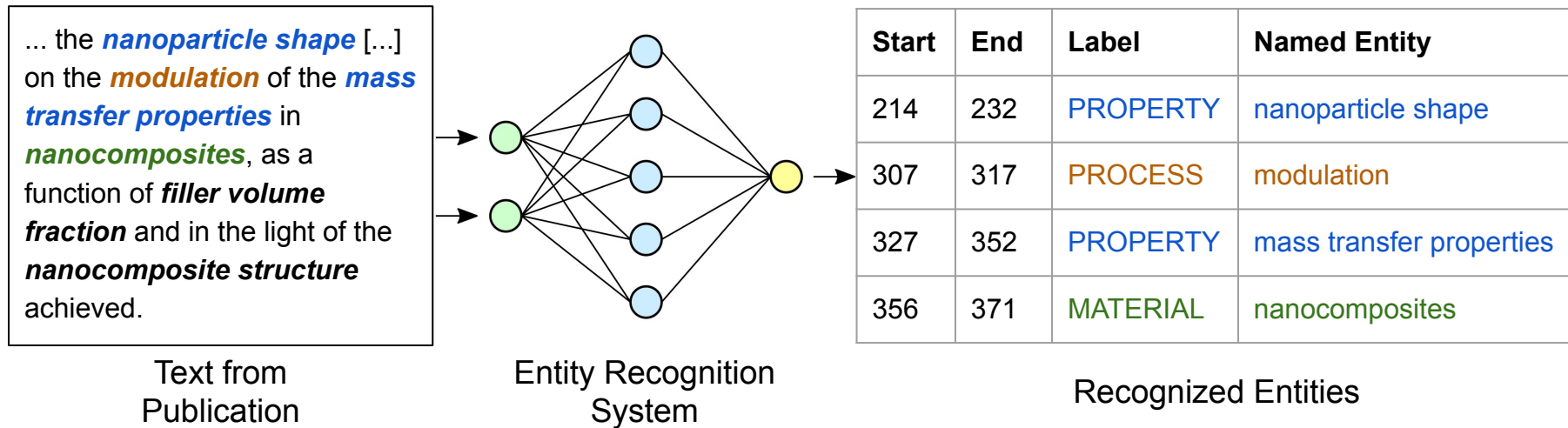
Entity Recognition: A Named Entity Recognition (NER) System

Language models are pretty good at recognizing and classifying named entities:



Entity Recognition: Limitations

Language models are pretty good at recognizing and classifying named entities:



- **Disconnected:** Recognized entities are typically stored separately for each document and are not easily searchable across publications
- **Coarse:** Very limited number of categories are typically used

Solution: A Structured and Semantic Representations of Entities

- **Disconnected:** Recognized entities are shared as lists (one per document)

1) Representation as a Knowledge Graph: **OnNER**

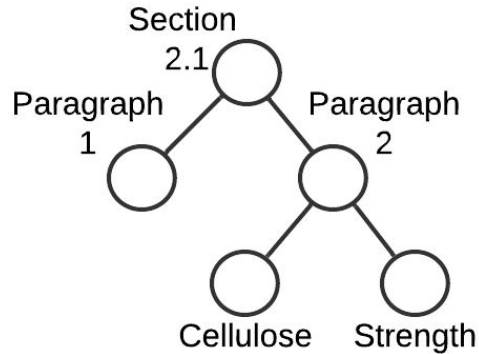
- captures the caption and context of entities
- allows querying entities more easily

- **Coarse:** Very limited number of categories are typically used

2) Linking entities and their classification labels to a **domain ontology**

Why a Knowledge Graph?

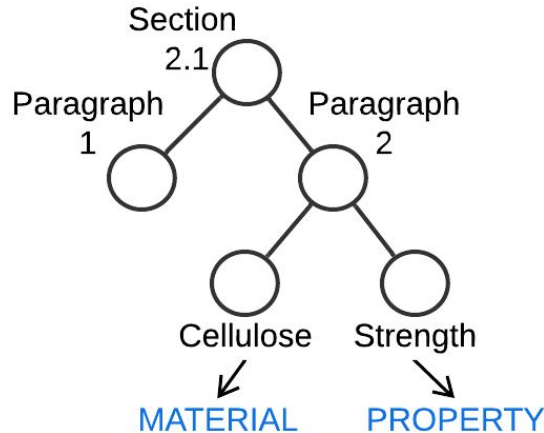
- A knowledge graph represents information as a network of connected concepts.
- Instead of isolated pieces of information, we can see how things are connected and explore relationships across data.



- **Nodes** represent *concepts*
- **Edges** represent *relationships* between nodes

Ontology: The Structure Behind a Knowledge Graph

- A structured framework to organize knowledge for consistent understanding and representation.
- An ontology defines:
 - What types of things exist
 - How different things are related



Concepts

- Document Part: Section 2.1, Paragraph 1
- Material: Cellulose
- Property: Strength

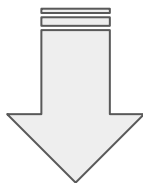
Relationships

- Cellulose $\Rightarrow$ *is-a* $\Rightarrow$ Material
- Section 2.1 $\Rightarrow$ *contains* $\Rightarrow$ Paragraph 1

OnNER's Approach: Representation of Entities

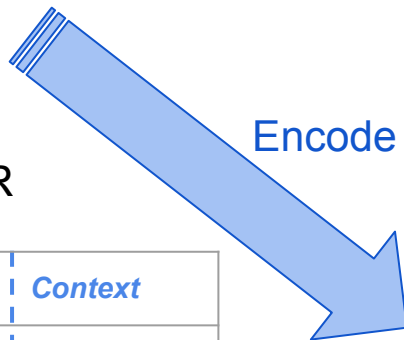


Publications
(PDFs)

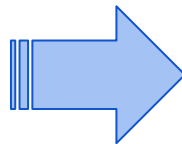


NER

Encode

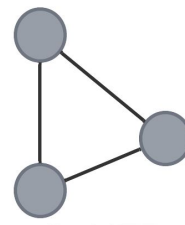


Encode



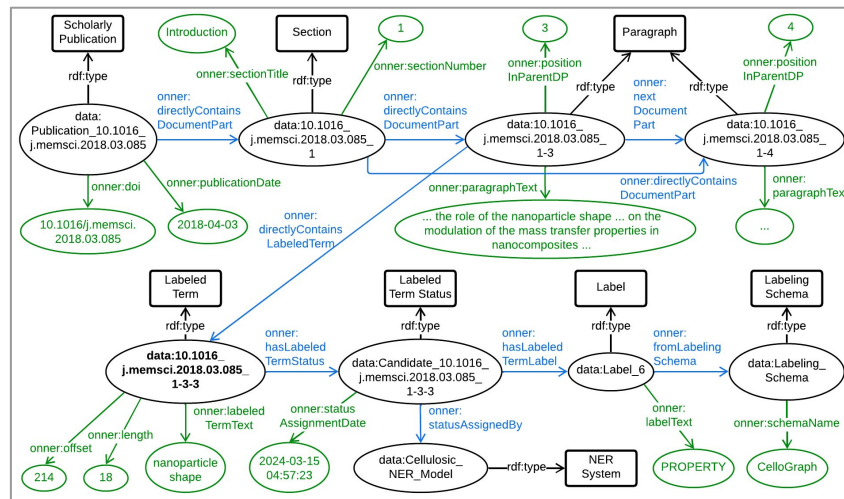
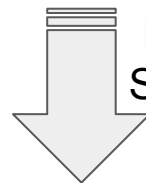
Start	End	Label	Named Entity	Context
214	232	property	nanoparticle shape	Section 1, Paragraph 1
307	317	process	modulation	Section 1, Paragraph 1
327	352	property	mass transfer properties	Section 1, Paragraph 1
356	371	material	nanocomposites	Section 1, Paragraph 1

Entities with Relative Positions in Text



OnNER

Defines
Structure



Queryable Knowledge Graph

2. Background ← Section

2.1. Dispersion of nanoparticles ← Section

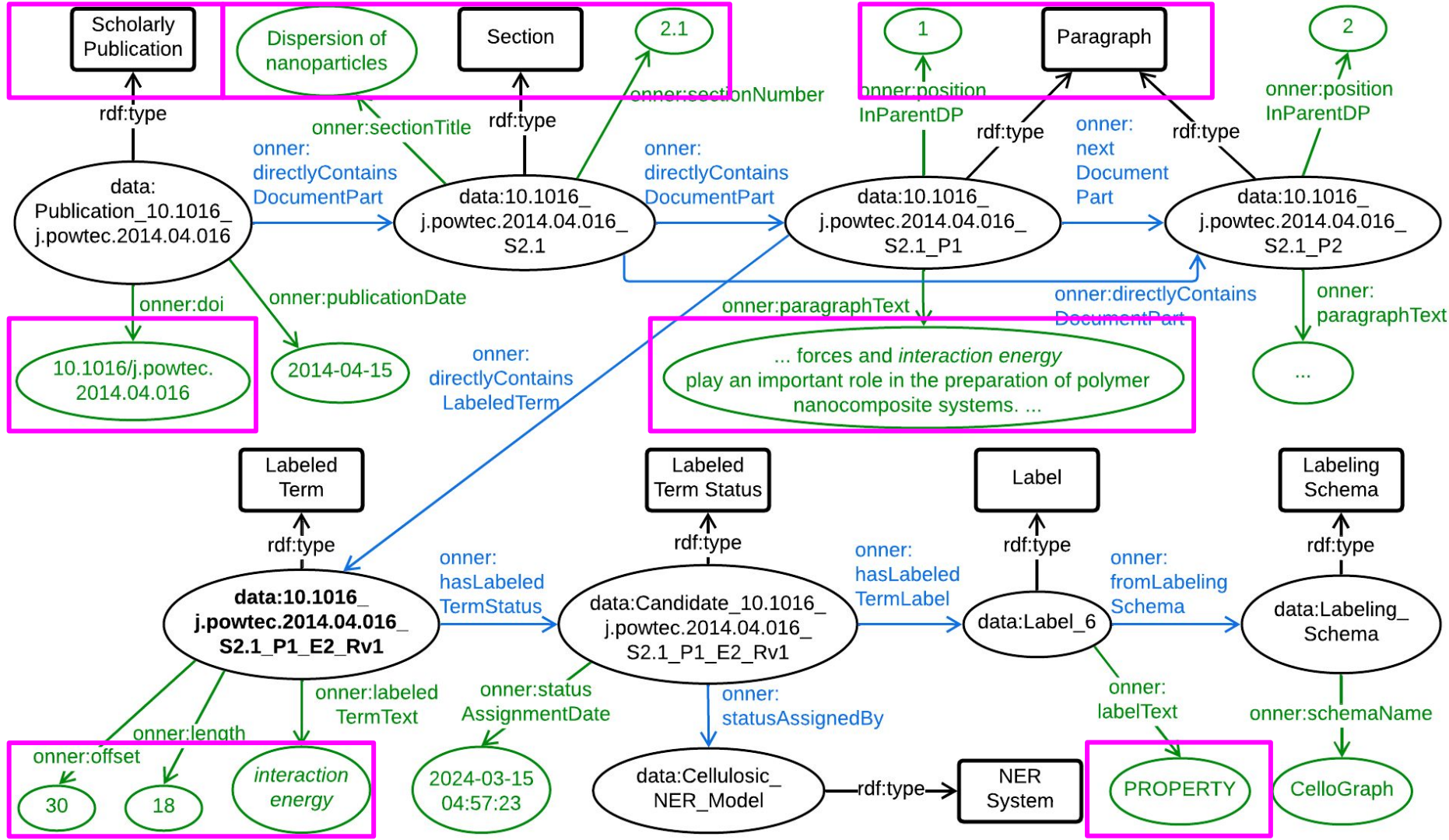
Inter-nanoparticle forces and interaction energy play an important role in the preparation of polymer nanocomposite systems. The magnitude of van der Waals interaction energy and forces between two isolated nanoparticles depends on the geometry of nanoparticles, as well as inter-particle distance. CNC can be considered as a cylindrical nanoparticle with 7 nm radius and 150 nm length [21]. Table 1 shows equations and estimates for inter-particle interaction energy and force between two cross and parallel CNC cylindrical particles. U_i , F_i , A_H , R , L , and z are the van der Waal interaction energy between particles, inter-particle force, Hamaker constant, particle radius, cylinder length, and inter-particle distance, respectively. A_H , the Hamaker constant, is defined as [22]:

$$A_H = 24\pi D_0^2 \gamma \quad (1)$$

where D_0 and γ are the theoretical molecular distance between two particles in contact, known as “cutoff”, 0.165 nm, and surface energy of CNC at 190 °C [21], respectively.

Paragraph

Khoshkava, V., & Kamal, M. R. (2014). *Effect of drying conditions on cellulose nanocrystal (CNC) agglomerate porosity and dispersibility in polymer nanocomposite s. Powder Technology, 261, 288-298.*



Curating the Knowledge Graph

- Automated knowledge graph construction is not reliable
 - Language models may **miss or incorrectly classify** specialized cellulose terms.
 - The **same term may have different meanings** depending on its context.
- That's why human review is needed
 - Human review **improves the quality and reliability** of the knowledge graph.
- We developed **AnNER**, a review tool for efficient validation of extracted entities.
 - Domain **experts can correct** these errors and confirm the extracted entities.

AnNER: A Tool for Reviewing and Validating Extracted Entities

File Edit Annotator Help

Install Application



AnNER: Named Entity Annotation and Review Tool



Open a Text or JSON file to begin

Hint: You can also drag and drop files into this window!



GITHUB



DOCUMENTATION

This tool was designed and developed in the Spatial Knowledge and Artificial Intelligence (SKAI) lab at University of Maine.

The research was supported in part by the U.S. Department of Agriculture through a National Institute of Food and Agriculture (NIFA) grant

Award #[2021-67022-34366](#)



Forest Service
U.S. DEPARTMENT OF AGRICULTURE

AnNER: Key Capabilities

File Edit Annotator Help Umayr-annotations.json - Review Mode

1 CHEMICAL 2 MATERIAL ✓ PROPERTY 4 APPLICATION

Accepted

Suggested

Rejected

Candidate

Cellulose is mainly used to produce paperboard APPLICATION and paper APPLICATION . Conversion of cellulose from energy crops MATERIAL into biofuels such as cellulosic ethanol ? CHEMICAL is under development as a renewable PROPERTY fuel source .

29 Words 189 Characters Paragraph 1/2 5 Annotations

BACK NEXT

Querying the Knowledge Graph

- Which two terms occur in the same paragraph? In which publications and paragraphs do they occur?
- What are the most recent publications that mention a specified term at least a specified number of times?
- Find all terms with a specified label from a particular publication.
 - Find all **processes** mentioned in the publication titled “*Effect of Drying Conditions on Cellulose Nanocrystal (CNC) Agglomerate Porosity and Dispersibility in Polymer Nanocomposites.*”

Browse & Filter

Frequently asked questions ?

Co-occurrence of Terms in Paragraph

Which two terms occur in the same paragraph? In which publications and paragraphs do they occur?

Term 1 Label 1

Term 2 Label 2

Execute

CelloGraph

A Knowledge Graph System for Cellulosic Materials

Output

Rows5107Columns8

TableReadable textGraphRaw JSON

	paragraph	term1	label1	term2	label2
4.016_S2.1_P1	Inter-nanoparticle	interaction energy	PROPERTY	Inter-nanoparticle forces	PROPERTY
4.016_S2.1_P1	Inter-nanoparticle	preparation	PROCESS	Inter-nanoparticle forces	PROPERTY
4.016_S2.1_P1	Inter-nanoparticle	polymer nanocomposite systems	MAT_ENT_STRUCT	Inter-nanoparticle forces	PROPERTY
4.016_S2.1_P1	Inter-nanoparticle	magnitude	PROPERTY	Inter-nanoparticle forces	PROPERTY
4.016_S2.1_P1	Inter-nanoparticle	van der Waals interaction energy a	PROPERTY	Inter-nanoparticle forces	PROPERTY
4.016_S2.1_P1	Inter-nanoparticle	nanoparticles	MAT_ENT_STRUCT	Inter-nanoparticle forces	PROPERTY
4.016_S2.1_P1	Inter-nanoparticle	geometry	PROPERTY	Inter-nanoparticle forces	PROPERTY
4.016_S2.1_P1	Inter-nanoparticle	inter-particle distance	PROPERTY	Inter-nanoparticle forces	PROPERTY

Acknowledgement
This research was supported in part by the U.S. Department of Agriculture under the Forest Service (FS) Project 20-JV-1111124-055, Agricultural Research Service (ARS) Project 0204-41510-001-98S, and the National Institute of Food and Agriculture (NIFA) Award 2021-67022-34366.

Browse & Filter

Frequently asked questions ?

Publications by Term Frequency ▼

What are the most recent publications that mention a specified term at least a specified number of times?

Term

nanoparticles

Occurrences

7

-

+

Execute

CelloGraph

A Knowledge Graph System for Cellulosic Materials

Output

Rows

1

Columns

3

Table

Readable text

Graph

Raw JSON

publication	date	doi
Effect of drying conditions on cellulose nanocrystal (CNC) agglomerate porosity and dis	2014-04-15	10.1016/j.powtec.2014.04.016

Acknowledgement

This research was supported in part by the U.S. Department of Agriculture under the Forest Service (FS) Project 20-JV-1111124-055, Agricultural Research Service (ARS) Project 0204-41510-001-98S, and the National Institute of Food and Agriculture (NIFA) Award 2021-67022-34366.

Browse & Filter

Frequently asked questions



Terms by Label and Publication



Find all terms with a specified label from a particular publication.

Title/DOI

Label

10.1016/j.powtec.2014.04.016

process

Execute

CelloGraph

A Knowledge Graph System for Cellulosic Materials

Output

Rows

16

Columns

6

Table Readable text Graph Raw JSON

	paragraph	term	label
/data#10.1016_j.powtec.2014.04.016_S2.1_P1	Inter-nanoparticle forces and interaction en	preparation	PROCESS
/data#10.1016_j.powtec.2014.04.016_S2.1_P2	The inter-particle van der Waals interaction	CNC surface modification	PROCESS
/data#10.1016_j.powtec.2014.04.016_S2.1_P2	The inter-particle van der Waals interaction	(covalent and non-covalent m	PROCESS
/data#10.1016_j.powtec.2014.04.016_S2.1_P2	The inter-particle van der Waals interaction	CNC dispersion	PROCESS
/data#10.1016_j.powtec.2014.04.016_S2.1_P2	The inter-particle van der Waals interaction	separation	PROCESS
/data#10.1016_j.powtec.2014.04.016_S2.1_P3	CNC nanoparticles with parallel orientation	pulling	PROCESS
/data#10.1016_j.powtec.2014.04.016_S2.1_P3	CNC nanoparticles with parallel orientation	peeling	PROCESS
/data#10.1016_j.powtec.2014.04.016_S2.1_P3	CNC nanoparticles with parallel orientation	pulling mechanism	PROCESS

Acknowledgement

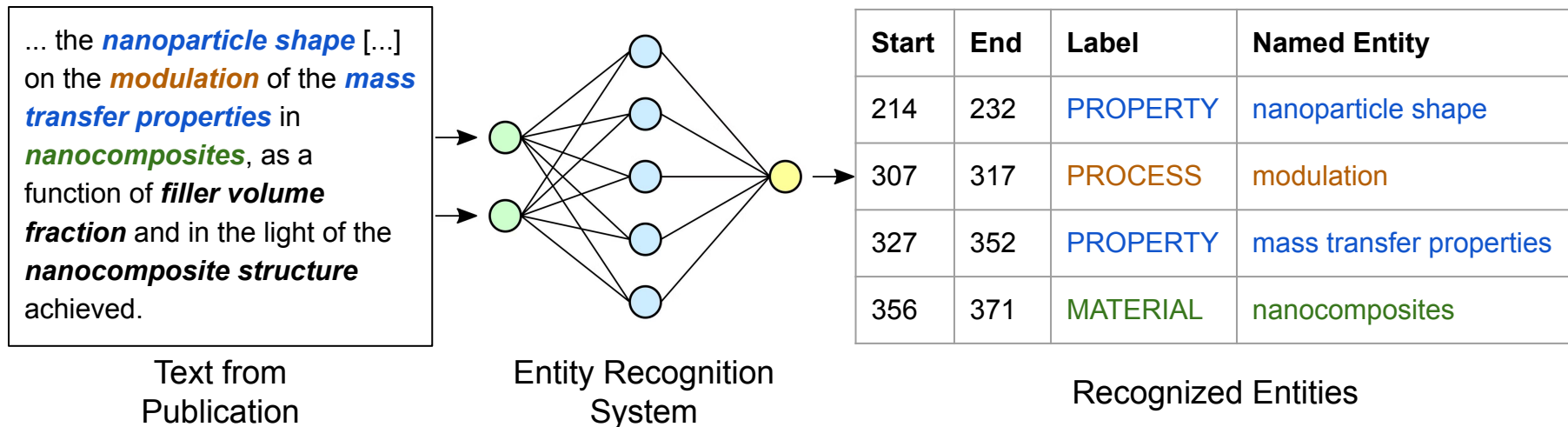
This research was supported in part by the U.S. Department of Agriculture under the Forest Service (FS) Project 20-JV-1111124-055, Agricultural Research Service (ARS) Project 0204-41510-001-98S, and the National Institute of Food and Agriculture (NIFA) Award 2021-67022-34366.

Summary and Future Work

- Scientific information on cellulosic materials is distributed across a rapidly growing body of literature.
- **CelloGraph** aims to organize scattered scientific knowledge about cellulosic materials and make it easily accessible to users.
- **OnNER** makes materials, properties, processes, and other scientific terms easier to search across publications.
- **AnNER** allows domain experts to validate automatically extracted terms.
- A domain-specific **NER model** is currently being tested to improve cellulosic entity extraction.
- Together, these components support more precise, transparent, and trustworthy knowledge discovery.

Summary and Future Work

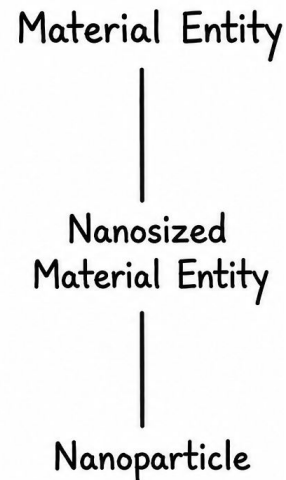
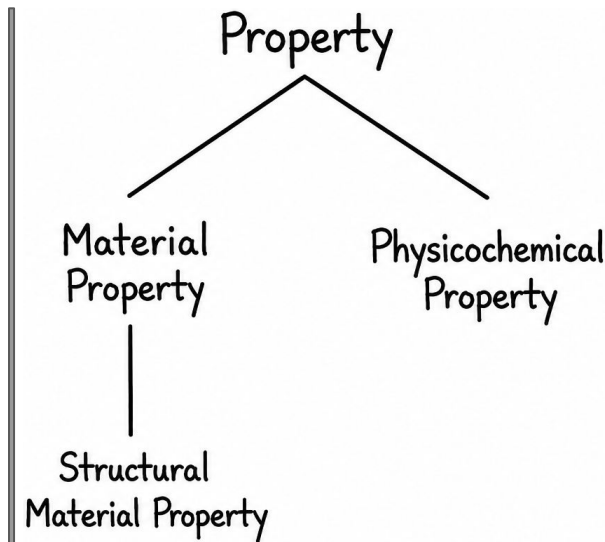
- Develop a high-performing NER model to accurately recognize specialized materials, properties, processes, and applications in scientific publications.



Summary and Future Work

➤ Develop a domain ontology to provide

- More granular categories (e.g., Nanoparticle)
- Clear structure (e.g., Material vs Physicochemical Property)
- Consistent labeling of entities
- Relationships between concepts



Acknowledgement

Thanks to our collaborators and project partners:

- Jinwu Wang, Forest Product Labs, US Forest Service
- Doug Gardner and Stephen Shaler, Sustainable Materials Technology, School of Forest Resources and Forest Bioproducts Research Institute, University of Maine

This research was supported in part by the U.S. Department of Agriculture:

- Forest Service, Project 20-JV-11111124-055
- Agricultural Research Service (ARS), Project 0204-41510-001-98S
- National Institute of Food and Agriculture (NIFA), Award 2021-67022-34366

Thank You!

Contact:

a.reza@maine.edu

References:

- Umayer Reza, Nicholas Pease, Jinwu Wang, and Torsten Hahmann. (2025). ***AnNER: Supporting Efficient Entity Annotation and Review Workflows for Knowledge Graph Construction***. In Knowledge Capture Conference 2025 (K-CAP 2025). <https://doi.org/10.1145/3731443.3771347>
- Umayer Reza, Xuelian Zhang, and Torsten Hahmann. (2024). ***OnNER: An Ontology for Semantic Representation of Named Entities in Scholarly Publications***. In 14th International Conference on Formal Ontology in Information Systems (FOIS 2024). <https://doi.org/10.3233/faia241291>