

# The Scientific Publication Ontology (SciPub): Linking publications' content to the domain ontologies

577

Abstract No.

Umayer Reza, Xuelian Zhang, Dr. Torsten Hahmann

School of Computing and Information Science, University of Maine

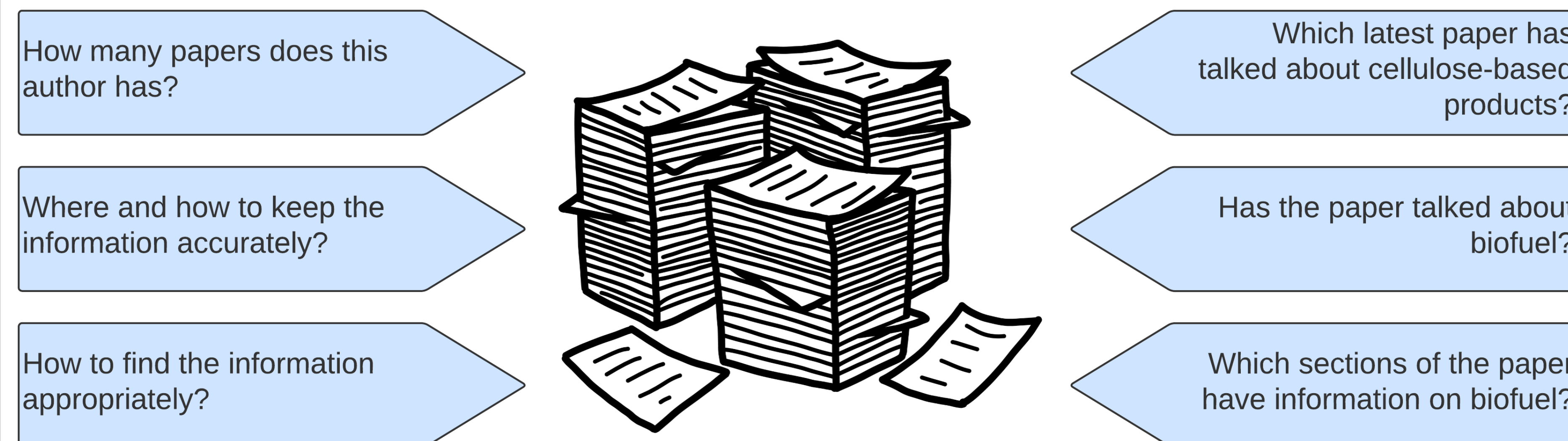
{a.reza, xuelian.zhang, torsten.hahmann}@maine.edu

## Motivation

The growing volume of scientific publications is making scientific knowledge challenging in various areas such as

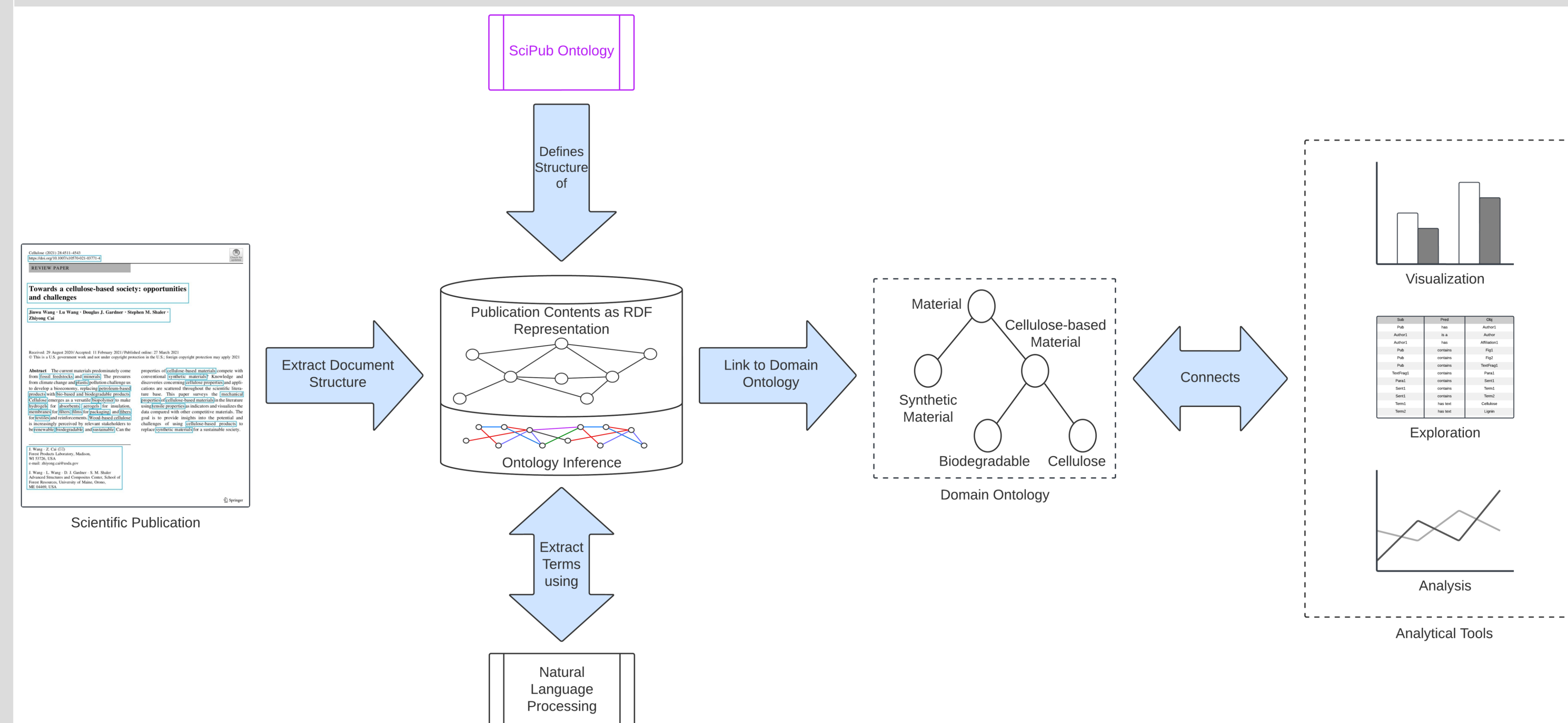
- Extraction
- Management
- Reusability
- Integrated analysis

It is also time-consuming and inefficient for humans to perform these tasks manually.

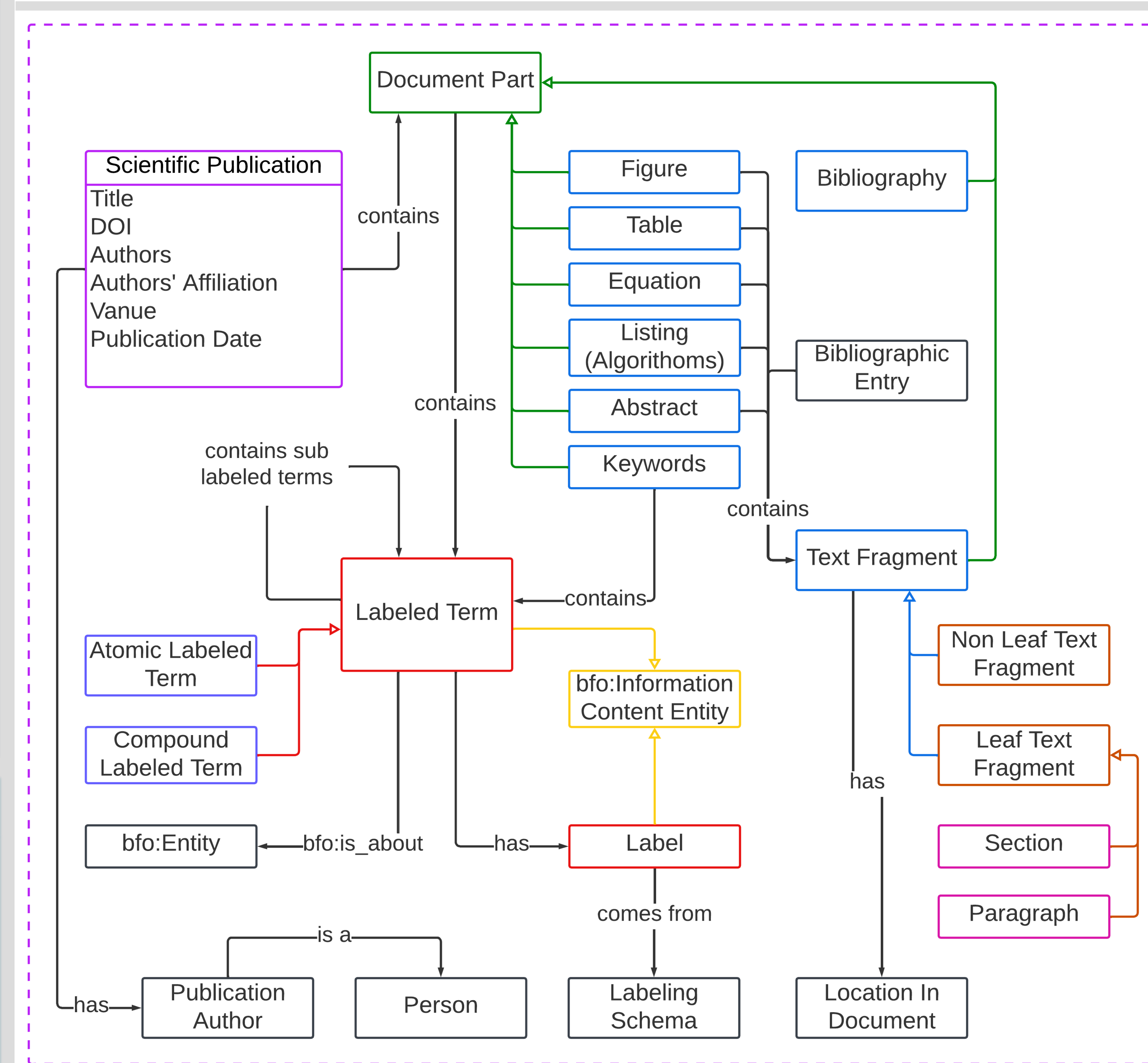


## Research Objective

- Develop SciPub, that can enable computer-readable and structured representation of domain knowledge by logically linking the contents of scientific articles to existing domain ontologies
- Assist in automating the knowledge extraction process and provide a systematic way of feeding domain knowledge
- Build a knowledge-based system that allows users to search and analyze scientific knowledge more holistically



## Methodology



1. Create competency questions as guide
2. Identify concepts at core of SciPub
3. Evaluate which existing ontologies to reuse
  - DoCO (Document Components Ontology)
  - CCO (Common Core Ontology)
  - CiTO (Citation Typing Ontology)
4. Connect to ontologies
5. Axiomatize SciPub
6. Instantiate SciPub with sample data
7. Check consistency using OWL Reasoners
8. Deploy ontology and dataset in GraphDB
9. Test answering competency questions using SPARQL

## Evaluation

**Competency Question:** Give me all sentences from publications that talk about **Lignin** (term CHEBI:6457 from the existing domain ontology ChEBI).

**Data**

```

:Figure1 rdf:type sp:Figure;
sp:title 'Main structure of plant cell wall in lignocellulosic biomass which is consisted of lignin, hemicellulose, and cellulose.'^^xsd:string;
sp:figureNumber '1'^^xsd:nonNegativeInteger;
sp:hasStartingLocation :Fig1Loc .
          
```

**SPARQL Query**

```

select ?pub, ?date, ?dp
where {
  ?pub sp:contains ?dp ;
  sp:published ?date .
  ?dp rdf:type sp:Text .
  ?dp sp:contains ?t .
  ?t rdf:type sp:LabeledTerm ;
  sp:is_about CHEBI:6457 .
}
          
```

| pub                                                                                                               | date | dp                                                                                                                                           |
|-------------------------------------------------------------------------------------------------------------------|------|----------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="http://cellograph.com/SciPub/2022v1.0/Para4Sent1">http://cellograph.com/SciPub/2022v1.0/Para4Sent1</a>   | 2018 | Cell wall structure of lignocellulosic biomass mainly consists of three kinds of polymer, i.e. <b>lignin</b> , hemicellulose, and cellulose. |
| <a href="http://cellograph.com/SciPub/2022v1.0/Para4Sent4">http://cellograph.com/SciPub/2022v1.0/Para4Sent4</a>   | 2018 | In plant cell walls, <b>lignin</b> serves as the binder which holds between and around cellulose and hemicellulose complex (Fig. 4).         |
| <a href="http://cellograph.com/SciPub/2022v1.0/Para22Sent3">http://cellograph.com/SciPub/2022v1.0/Para22Sent3</a> | 2018 | Firstly, the non-cellulosic components, such as <b>lignin</b> , hemicellulose, and other compounds, are removed by the pretreatment.         |

**Result**

- **Implementation:** OWL/RDF ontology
- **Use Case:** Populating the ontology with data from one scientific publication
- **Evaluation:** 24 competency questions were expressed and answered with SPARQL query

## Acknowledgement

- Thanks to Jinwu Wang, Douglas Gardner, and Stephen Shaler for fruitful discussions of this work.
- This research was supported in part by the U.S. Department of Agriculture:
  - Forest Service, Project 20-JV-1111124-055
  - USDA Agricultural Research Service (ARS), Project 0204-41510-001-98S
  - National Institute of Food and Agriculture (NIFA), Award 2021-67022-34366