TYPE Original Research
PUBLISHED 29 January 2026
DOI 10.3389/fcomp.2026.1636097

Check for updates

OPEN ACCESS

EDITED BY
Marc Bertin,
University of Franche-Comté, France

REVIEWED BY
Ariadne de Andrade Costa,
Universidade Federal de Jataí, Brazil
Joel Oduro-Afriyie,
OriGen AI, United States

*CORRESPONDENCE
Mark Daly
✉ mark.daly@outlook.de

RECEIVED 27 May 2025
REVISED 14 December 2025
ACCEPTED 05 January 2026
PUBLISHED 29 January 2026

CITATION
Daly M, Aslam MA, Froelian R and Schimmler S (2026) OntoTrack: a linked open data based solution to track mutual citation networks in research publications.
*Front. Comput. Sci.* 8:1636097.
doi: 10.3389/fcomp.2026.1636097

# OntoTrack: a linked open data based solution to track mutual citation networks in research publications

Mark Daly[1]*, Muhammad Ahtisham Aslam[2], Ronja Froelian[1] and Sonja Schimmler[2]

[1]Technische Universität Berlin, Faculty IV, Information Systems Management, Berlin, Germany,
[2]Fraunhofer Institute for Open Communication Systems (FOKUS), Digital Public Services (DPS), Berlin, Germany

**Context:** Scientific publications are vital for a researcher's scientific career. Citations of scientific work by other researchers are considered as evidence of scientific and technical strength and acceptance of one's scientific contributions. The higher citation count directly raises the "h-index," which is evidence of a strong scientific profile. Due to its direct impact on scientific profiles, citation manipulation (unnatural citations) is becoming a major concern in academia and industry.

**Methods:** To address this challenge, we present OntoTrack, an ontology-based solution that can be used to detect and identify potential unnatural citations and citation networks within the scientific literature. In this paper, we present the complete architecture of the OntoTrack solution. We also present the OntoTrack data model and ontology with its key attributes and parameters, which play an important role in tracking citation networks. OntoTrack ontology is equipped with a comprehensive set of rules that are defined using the Semantic Web Rule Language (SWRL). These rules enhance the reasoning capabilities of OntoTrack and facilitate smart identification of unnatural citation indicators. A proof-of-concept dataset is produced as part of this work and used to evaluate the effectiveness and precision of the OntoTrack Solution in detecting citation anomalies.

**Results:** We also present the evaluation of the OntoTrack Solution by defining a comprehensive set of Competency Questions (CQs) and executing these against the OntoTrack SPARQL Protocol and RDF Query Language (SPARQL) Endpoint. The results of the evaluations show that OntoTrack can successfully identify various forms of unnatural citations, including self-citations, citation cartels, and citation manipulation among researchers. The results also show that an ontology-based approach provides a sustainable and efficient alternative to traditional machine-learning methods, which often require extensive computational resources.

**Discussion:** The findings suggest that ontology-based systems such as OntoTrack can enhance transparency and integrity in academic research by providing a robust mechanism for monitoring citation practices.

KEYWORDS

**linked open data (LOD), machine reasoning, ontologies, scientific citations, semantic web, semantic web rule language (SWRL), SPARQL Queries**

# 1 Introduction

Career building is a key focus for every professional in almost every field and domain. In the IT research industry, career building is closely dependent on the scientific profile of a researcher, which depends on several key factors, and the citation count of a researcher's scientific work is one of them. A citation in the scientific literature is a formal acknowledgment of a source of information referenced and used by the citation literature (Ashikuzzaman, 2018). A researcher's standing is often measured by the number of publications and the frequency with which they are cited. Increasing the number of citations can be used to increase one's recognition and hence enhance the scientific profile, which provides direct support in career building.

To gain recognition within the scientific community, citation count is more decisive than the number of publications, as it reflects the impact and reach of a researcher's work (Smith, 1981). This is one of the main reasons for the increase in the number of non-essential (non-related) citations in scientific articles. The findings of one study show that 41% and another show that 20% of the references in the scientific literature are not essential (Nicolaisen, 2007). There are also cases where the citation is used with less academic motivation, by including citations solely to promote a colleague's or friend's work without any scientific relevance (Smith, 1981). This practice, known as citation manipulation or citation stacking (COPE Council, 2019), leads to unnatural citations and can occur at both the individual and among journals, contributing to journal cartel citations. Tracking and identifying such unnatural citations and citation networks is a key challenge in true scientific profiling.

To address the above-mentioned challenges, in this paper, we present our solution (named OntoTrack). The OntoTrack Solution uses Linked Open Data (LOD) techniques and machine reasoning to identify unnatural citation patterns. In this paper, we present the complete architecture of the OntoTrack Solution with its different phases, such as (i) Requirements specifications, (ii) the OntoTrack ontology model, (iii) Data preparation, and (iv) SPARQL Endpoint. By representing citation networks semantically and leveraging the reasoning capabilities of ontologies, it becomes possible to model, track, and analyse citation relationships systematically. Our ontology-based reasoning approach demonstrates how logical rules and publicly available metadata, such as authorship and keyword overlap, can be used to detect anomalies in citation patterns. Furthermore, as an integral part of our solution, we developed and used the OntoTrack ontology and evaluated key attributes, such as thematic relevance and co-authorship, and their effectiveness in distinguishing natural citations from manipulated ones.

This study presents the following key contributions:

- OntoTrack Solution and its architecture, with its various phases.
- OntoTrack data model and its implementation as a Web Ontology Language (OWL) ontology with identification of key attributes such as academic work, author information, and keywords to track citation networks.
- A Comprehensive set of rules, their implementation in OntoTrack Ontology by using SWRL and the role of these rules in reasoning and identification of unnatural citations.
- A list of Competency Questions (CQs) to target citation validity, providing a structured approach to detect citation anomalies.
- A use case scenario and a proof-of-concept dataset to demonstrate the system's effectiveness.
- Implementation of OntoTrack SPARQL Endpoint for smart competency question answering purposes to identify, retrieve and show the unnatural citation data.

The rest of the paper is structured as follows: Section 2 provides an overview of the relevant background, including a detailed discussion on citations in the scientific context and the role of ontologies. This section also reviews related work in unnatural citation detection and presents relevant ontologies in this domain. Section 3.1 describes the detailed architecture (with various phases) of the OntoTrack Solution, outlining the ontology-based approach for tracking and identifying unnatural citation patterns using machine reasoning. Furthermore, this section also describes the technical implementation of the OntoTrack Solution, including the Competency Questions (CQs) and rules defined by using Semantic Web Rule Language (SWRL). Section 4.1 presents the evaluation results using a proof-of-concept dataset, discussing the effectiveness and limitations of OntoTrack in detecting unnatural citations. Finally, Section 7 summarizes this work's contributions and suggests future work directions.

# 2 Background and related work

## 2.1 Citation background

A citation in research papers is considered unnatural when it lacks proper justification or relevance, such as referencing a research colleague's work solely based on personal relationships (Smith, 1981). Citation manipulation typically aims to inflate an academic work's perceived significance artificially, thereby undermining the integrity of scientific research (Smith, 1981). Proper citations, in contrast, acknowledge the source of shared information from where the information is shared (Ashikuzzaman, 2018). This transparency is essential in scientific literature to avoid plagiarism. The distinction between a citation and a reference is also important, with a citation serving as a pointer to the source of information and a reference providing the complete bibliographic framework (Garfield, 1955).

A citation network establishes semantic links between cited and citing documents, forming a structure that can be represented as a directed graph where nodes represent documents and edges represent citations (Garfield, 1955). In this graph, each node represents a document, and the edges indicate a citation from one document to another. Citation networks are generally sparsely linked but contain clusters of densely linked regions that correlate to specific subject areas or expertise (Small, 1999). Such networks are useful for many forms of citation analysis, including identifying unnatural citations.

Several approaches to identifying unnatural citations have employed various analytical techniques, including graph analysis and machine learning. Graph Neural Networks, for instance, have been used to incorporate both the semantic content and citation structure of papers to detect anomalous citations (Liu et al., 2022). Another approach involves identifying reliable citations by learning different graph embeddings and examining permutations of the citation graph (Avros et al., 2023). CIDRE, a degree-corrected stochastic block model, specifically targets cartel citations by comparing randomized versions of citation networks with the original to detect manipulative citation behavior between journals (Kojaku et al., 2021). In addition, Wren and Georgescu (2020) proposed using the Gini Index to detect reference list manipulation, while Zhao and Strotmann (2015) outlined three commonly used measures in citation analysis inter-citation counts, co-citation counts, and bibliographic coupling frequencies (BCFs), to assess the strength of relationships between two publications. **Inter-citation counts** measure the number of times two documents have directly cited each other, representing reciprocal acknowledgment and mutual influence between works. **Co-citation counts** refer to the number of times two papers are cited together by subsequent publications, indicating that they are perceived as related or complementary in later research. In contrast, **bibliographic coupling frequency (BCF)** captures how many references two papers share in their respective bibliographies, reflecting similarity in intellectual foundations or research domains (Zhao and Strotmann, 2015). These measures together provide a perspective of the connectedness of scientific documents and suggest that documents with low scores across all three are less likely to have contextual overlap, indicating a higher chance of unnatural citation.

## 2.2 Ontologies in the publication field

In knowledge engineering, ontologies have become a formal tool for understanding specific domains (Chandrasekaran et al., 1999). An ontology specifies key concepts, their properties and relationships in a structured way, making domain knowledge machine-readable for better communication and reasoning (Vickery, 1997). They facilitate effective communication and reasoning across different systems and applications (Uschold and Gruninger, 1996). This machine-readable format clarifies the knowledge structure and provides a systematic representation that improves understanding and communication, enables knowledge sharing by providing a common vocabulary and semantics, and promotes collaboration, interoperability and consistency (Uschold and Gruninger, 1996).

In addition, ontologies are beneficial to information systems such as search engines and digital libraries (Aslam and Aljohani, 2016, 2017), which use domain ontologies to organize and guide searches. Ontologies are also valuable for artificial intelligence, where they can help systems understand, reason and make decisions based on structured knowledge (Uschold and Gruninger, 1996). They can be used as data models in databases, with SPARQL being used to query these models. Compared to traditional relational databases, ontologies excel at managing complex relationships and inferring new knowledge. Reasoning engines can derive new knowledge from existing data by applying logical rules and relationships defined in the ontology. Additionally, ontologies provide greater flexibility and scalability, as they are not constrained by their static table structures, making it easier to scale and manage evolving data models horizontally.

The FRBR-aligned Bibliographic Ontology (FaBiO) is an ontology for describing entities that are published or potentially publishable (e.g., journal articles, conference papers, and books) and that contain or are referred to by bibliographic references. It focuses on the description of bibliographic entities, which might include anything from published works to those that could be published (SPAR Ontologies, 2019). FaBiO provides a comprehensive list of potential media but cannot track metadata information about the authors of these media, severely limiting its stand-alone usage for citation tracking.

With its aim to help make in-text references machine-readable, the Citation Counting and Context Characterization Ontology (C40) provides properties and classes that help identify these. This allows a user to identify the number of times another work is directly referenced and provides a more fine-grained overview than the Citation Typing Ontology as it captures all references within a text (SPAR Ontologies, 2024). Similar to CiTO, C40 focuses heavily on the citation tracking aspect on an even more fine-grained scale. Therefore, it has similar limitations when used alone for unnatural citation identification, namely, a lack of ways to track author and academic work metadata.

The Citation Typing Ontology (CiTO) is an ontology that enables the characterization of the nature or type of citations. It is used for articulating both the factual and rhetorical aspects of citations. This ontology facilitates the detailed characterization of citations (SPAR Ontologies, 2018). CiTO is heavily focused on the citation aspects and has no classes or properties that enable a user to track author information or metadata information about the academic works that are cited.

The OpenCitations Ontology (OCO) aggregates ontological entities from various existing ontologies to support the OpenCitations Data Model, which describes metadata for the OpenCitations Corpus, OpenCitations Indexes, and other datasets (Peroni, 2024). While OCO is the most comprehensive ontology in this section, it lacks certain aspects we consider important to identify unnatural citations. Such as a way to track the topics of academic works via keywords, and is limited in its abilities to track the social network between the different authors.

In this section, we described several existing ontologies, with varying focuses, in the publication field. Table 1 provides a short overview of the most relevant ontologies for unnatural citation identification. Entries marked by X contain the described attribute, while entries marked with (X) only partly contain it and would require additional ontologies to fulfill the needs of OntoTrack completely. Ontology-based solutions like CiTO, FaBiO, and OCO describe bibliographic relationships but do not reason about citation intent or integrity. As summarized in Table 1, existing ontologies do not consider attributes, such as institutional affiliations, keywords, and journal dynamics within a unified framework, which play an important role in reasoning

TABLE 1 Comparison of key attributes across the different ontologies.

| Ontology | Multiple types of documents | Lists of authors | Citation relationship | Extended citation information | Journal information | Keywords | Research field |
|---|---|---|---|---|---|---|---|
| FaBiO | X | X | X | | | | |
| C40 | | | X | X | | | |
| CiTO | | X | X | X | X | | |
| OCO | X | X | X | | | | |
| OntoTrack | X | X | X | X | X | X | X |

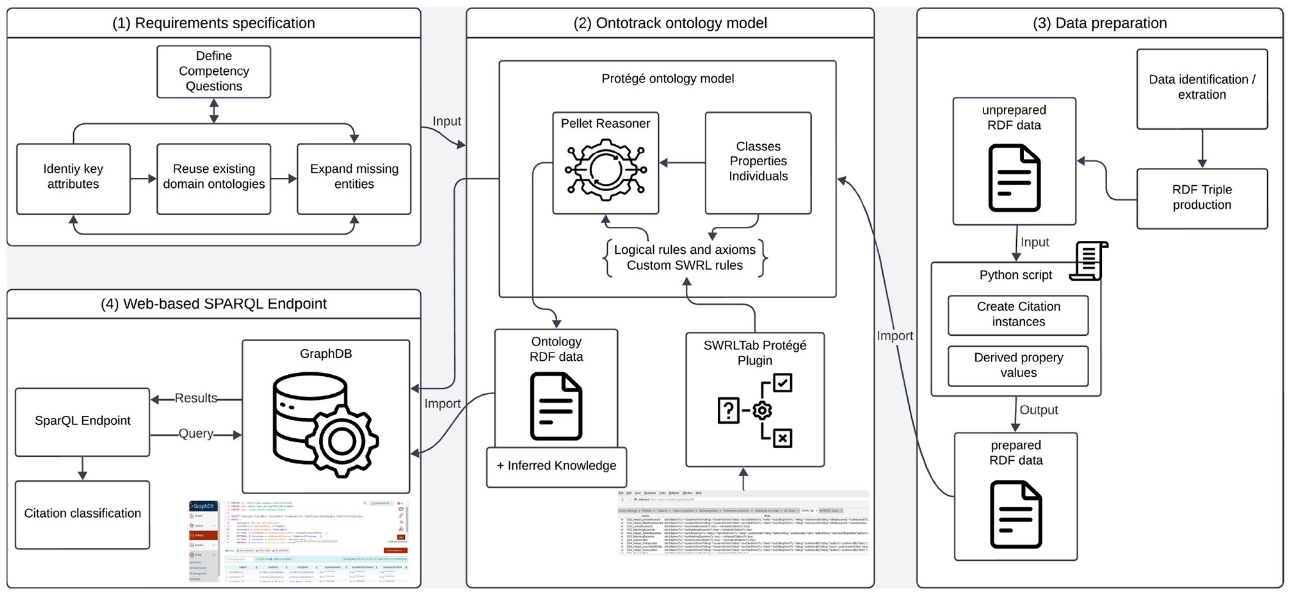


FIGURE 1
Architecture of the OntoTrack solution.

and identifying potentially unnatural citations. OntoTrack bridges this gap by combining Linked Open Data (LOD) with semantic reasoning, using SWRL rules to infer thematic relevance, collaboration history, and self-citation behavior and by considering the above-mentioned entities and parameters. This reasoning-driven design promotes transparency, interpretability, and reproducibility through a public SPARQL endpoint, distinguishing OntoTrack as an extensible framework for maintaining citation integrity.

# 3 Methodology

## 3.1 Architecture of the OntoTrack solution

The solution diagram in Figure 1 shows the architecture and functionality of OntoTrack, demonstrating how different interconnected phases build a structured workflow to identify unnatural citations. The ontoTrack approach leverages machine reasoning and logical rules to enhance citation analysis. The workflow comprises of four main interfeeding phases: (1) Requirements Specification, where key attributes are identified, existing ontologies are reused, and competency questions guide ontology is developed; (2) OntoTrack Ontology Model, which integrates logical rules and custom SWRL rules to enable reasoning about classes, properties, and individuals; (3) Data Preparation, where citation-related Resource Description Framework (RDF) data is enriched with additional properties and structured relationships; and (4) Web-based SPARQL Endpoint, which uses GraphDB to query and classify citations and uses inferred data for analysis. This approach combines ontology modeling, machine reasoning and data enrichment to effectively detect and analyse unnatural citation patterns. The following sections describe each phase of the OntoTrack ontology-based solution in detail.

### 3.1.1 Requirements specification

The initial step in designing the ontology-based solution is defining competency questions (CQs), which guide the requirement specification process and ensure that key attributes for citation analysis are identified from the outset. A comprehensive list of these CQs, covering aspects such as collaboration networks, thematic overlap, and citation legitimacy, is provided in Section

TABLE 2 Key attributes of OntoTrack ontology.

| Attribute | Reason for inclusion |
|---|---|
| Academic Work | To be able to identify academic works and their metadata. |
| Author Information | Tracking Author information allows us to model the social component of the citation network. |
| Affiliation | Provides more social context to the citation network and is used to support the author information. |
| Publication Information | Publication Information includes where an academic work is published and allows us, for example, to track how different journals cite each other. (see text footnote [1]). |
| Keywords | Provide a way to track the subject matter of a paper. |
| Research Field | Allows modeling more information about the authors and with which subject matters they engage. (see text footnote [2]). |

4.1.2 and serves as the conceptual basis for constructing both the ontology and the reasoning framework. The CQs define the core analytical needs that the OntoTrack should be able to answer, ensuring that the ontology can represent the information necessary to analyse citation behavior and classify citations effectively. For example, the CQ to identify self-citation behavior can be defined as: "Does an academic work cite another work authored by the same person?" To achieve this goal, the OntoTrack ontology requires entities to represent academic work, the author, including attributes for unique identification such as name and affiliation, and attributes to represent citations between the literature. It must be determined how these attributes are defined in the ontology, as a class, object property or data property, and which characteristics the ontology entities should have. Expanding the ontology with additional attributes is a continuous, iterative process as new CQs are defined. The defined set of competency questions developed from this specification is presented in Section 4.1.2, where they are further used to derive reasoning rules that infer new data and enhance citation behavior analysis.

Table 2, shows some key attributes identified for the OntoTrack ontology. Including Academic Work in the OntoTrack ontology as an entity is essential, as it forms the foundation of the citation network being tracked. Additionally, Author Information is crucial for modeling the social dynamics of the citation network, such as determining who has worked with whom. Authors with shared affiliations, whether at the same company or university, could likely cite one another, which may indicate potential citation manipulation. The Affiliation attribute further supports this by tracking institutional connections that might influence citation behavior. Publication journal Information helps track how different works within journals cite each other, addressing issues such as citation cartels or citation manipulation at the publication level. Finally, the attributes of keywords and area of expertise are used to capture the subject matter of academic works and authors' research interests, allowing the detection of potential unnatural citations based on thematic overlap. These attributes, among others, form a comprehensive framework for monitoring citation patterns and identifying irregularities in citation networks.

When designing an ontology, it is important to identify and reuse existing ontology constructs as much as possible. This helps ensure consistency, avoid redundancy, and take advantage of the expertise and best practices in established ontologies. Reusing existing constructs also enhances interoperability, enabling seamless integration with other ontology-based systems and datasets that rely on the same or compatible ontologies. It also saves time and effort in the design process and helps maintain consistency across domains. In Section 2.2, multiple existing ontologies in the publication field were identified. For the development of OntoTrack the Open Citations Ontology (OCO) has been reused. This ontology fits the requirements best and already reuses constructs imported from the previous publication ontologies. Various attributes specific to citation analysis and classification, describing different citation attributes and relationships between entities, have been built upon the OCO ontology.

### 3.1.2 OntoTrack ontology model

The ontology model is developed by using Protégé as an ontology editor, incorporating classes, properties, individuals, and rules [implemented by using Semantic Web Rule Language (SWRL)] (see Section 3.1.2.3). The Pellet reasoner ensures consistency and enables the production of RDF data with inferred knowledge. Rules are developed with the SWRL plugin for Protégé and are incorporated as an integral part of the OntoTrack solution.

#### 3.1.2.1 Ontology entities and definitions

The OntoTrack ontology data model is designed in such a way that it systematically organizes and represents data about academic authors, their publications, and the relationships between these entities. By leveraging existing constructs from existing domain ontologies, OntoTrack promotes interoperability, standardization, and semantic consistency while minimizing redundancy. This design choice allows the ontology to benefit from domain expertise and improve extensibility for future applications in tracking and analyzing citation behaviors.

The OntoTrack ontology focuses on modeling the relationships between *academic works*, *authors*, *affiliations*, *citations*, and other related key entities. For example, the *AcademicWork* class represents scientific publications (such as journal articles, conference papers, or books), which is crucial for tracking citation patterns. Relationships, such as *author* of a published paper, are modeled using the *hasPublished* object property, which links an *Author* to an *AcademicWork*. Similarly, *PublishedIn* connects *AcademicWork* to their corresponding *journal* or *proceeding volume*, ensuring accurate tracking of publication venue.

Additionally, the ontology tracks areas of expertise through the *hasAreaOfExpertise* property, which links an *Author* to areas of expertise. This is essential when analyzing the expertise overlap between citing and cited works, especially when determining the relevance of a citation. The *Keyword* entity, linked via *hasKeyword*, allows papers to be tagged with specific terms, improving the granularity and accuracy of citation analysis by enabling keyword-based reasoning.

A key feature of the OntoTrack ontology is its ability to track relationships between *citations* and their associated *academic works*.

For instance, each *Citation* is linked to the citing and cited works using the *hasCitingWork* and *hasCitedWork* properties. This allows *OntoTrack* to model and reason how citations are used and whether they exhibit unusual patterns, such as self-citations or citation cartels.

To support this model, several object properties have been newly created or extended within *OntoTrack*. These include *hasWorkedWith*, which tracks collaboration between *authors*, and *isLinkedTo*, which connects related *keywords*. These relationships are significant for reasoning about co-authorship networks and detecting citation anomalies.

Table 3 outlines some key classes, including *AcademicWork*, *Author*, and *Journal*, and their respective attributes and relationships. Table 4 presents data properties like *authorID*, *isSelfCitation*, and *hasUnnaturalIndicator*, which are critical for analyzing citation behavior and identifying irregular patterns. Table 5 provides a technical specification of some important object properties, detailing their domains, ranges, characteristics, constraints, and implementation status, while Table 6 offers a conceptual overview, explaining the purpose and reasoning behind some object properties in the ontology. The OntoTrack ontology incorporates data properties that enable a detailed analysis of citation relationships. For instance, the *isSelfCitation* property identifies self-citation cases, where the citing and cited works share the same author. Similarly, the *hasMatchingKeyword* property determines whether the keywords associated with the citing and cited academic works align, providing insights into the thematic relevance of a citation. These properties, among others, enable OntoTrack to reason about citation patterns and assess their legitimacy.

### 3.1.2.2 Role of reasoner in ontology-based citation analysis

When working with ontologies, reasoning plays an important role in ensuring the logical consistency of context information and deriving high-level, implicit context from low-level, explicit data. Logical reasoning ensures consistency, where every axiom $A$ in the ontology satisfies:

$$\forall x\, A(x) \rightarrow \text{true}. \tag{1}$$

By applying logical rules, the reasoner checks for inconsistencies, which can be represented as:

$$\exists x\, \neg A(x), \tag{2}$$

This helps to maintain the integrity of the context model. Logical reasoning can be extended by adding specific user-defined logical rules to the ontology using SWRL. Section 3.1.2.3 demonstrates how these rules extend the logical reasoning and support the analysis of citations. The Pellet reasoner is used, which is not the default reasoner for the Protégé version and requires a plugin. However, it offers native support for W3C SWRL specifications and does not require manual execution of Drools in the SWRLTab plugin in Protégé (DeBellis, 2020). Another example is HermiT 1.4.3 Reasoner, which does not support built-in atoms used by SWRL.

The classes and object properties also influence the reasoning capabilities of an ontology. Table 7 provides the reasoning context behind some example object properties to enhance the logical consistency of the ontology. For example, in the OntoTrack project, the *isLinkedTo* property between keywords is characterized as symmetric, transitive, and reflexive. These properties allow the reasoner to recognize bidirectional relationships:

$$\text{Symmetric:} \quad \forall x, y\, (R(x, y) \rightarrow R(y, x)), \tag{3}$$

$$\text{Transitive:} \quad \forall x, y, z\, ((R(x, y) \wedge R(y, z)) \rightarrow R(x, z)), \tag{4}$$

$$\text{Reflexive:} \quad \forall x\, R(x, x). \tag{5}$$

This extends the network of linked keywords and prevents misclassification when works share the same keyword. Each property characteristic, such as symmetry, transitivity, and reflexivity, significantly impacts the reasoner's functionality. Symmetry allows the reasoner to recognize bidirectional relationships, ensuring that if one element is linked to another, the reverse is also true. Transitivity enables the reasoner to infer indirect relationships through intermediate elements, expanding the network of related entities. Reflexivity ensures each element is linked to itself, preventing misclassifications in overlapping cases. Tables 5, 6 provide an overview of the assigned characteristics of further object properties.

### 3.1.2.3 Custom reasoning rules with semantic web rule language

**SWRL** is a language for expressing rules that can be combined with the Web Ontology Language (OWL) ontologies. SWRL rules are expressed in terms of OWL classes, properties, and individuals and provide a way to infer logical consequences based on these entities. The syntax of SWRL resembles rules in logic programming, functioning as a series of if-then statements. The "if" part specifies the conditions, while the "then" part specifies the conclusions.

SWRL rules in OntoTrack are designed to identify and flag citation patterns by inferring logical relationships from the underlying data. Rather than rendering definitive judgements, these rules highlight characteristics that may indicate natural or potentially problematic citation behavior. These rules allow logical implications to be applied to infer new knowledge, particularly distinguishing between appropriate citation patterns and potential citation manipulation. The SWRL rules are based on competency questions, each aimed at capturing specific characteristics of citation behavior. The framework divides the reasoning process into two types of rules: helper rules and indicator rules. A full list of the SWRL rules in OntoTrack is provided in the Github repository.[1]

**Helper rules:** Helper rules decompose complex reasoning processes into smaller, interpretable components. By inferring intermediate properties of academic works, such as co-citation, shared keywords, or author overlap. These rules enrich the ontology with structured relational information without relying on unsupported constructs such as negation or aggregation. Splitting the logic into smaller helper rules also allows the reasoner to restrict the search space earlier and reuse intermediate results, improving both reasoning speed and transparency. Helper rules supply the structured evidence on which indicator rules subsequently build to

[1] https://github.com/MarkEMD95/OntoTrack

TABLE 3 Ontotrack Classes.

| Class name | Text description | Parent class | Child class | Disjoint class | Same class | Origin | Status (New/ adopted/extended) |
|---|---|---|---|---|---|---|---|
| AcademicWork | Class to represent scientific publications | Thing | | | OCO:Expression | OntoTrack | Extension |
| AcademicProceedings | Academic proceedings are events such as conferences or workshops | AcademicWork | Onto:Book/Thesis | | OCO:AcademicProceedings | OntoTrack | Extension |
| ProceedingsPaper | Specific academic work that is published via academic proceedings | AcademicProceedings | | | OCO:ProceedingsPaper | OntoTrack | Extension |
| ProceedingsVolume | Media entity collecting all proceedings papers that are published together | AcademicProceedings | ProceedingsVolume, ProceedingsPaper | Journal, Book, Thesis | | OntoTrack | New |
| Book | A scientific book | AcademicWork | BookChapter | | OCO:Book | OntoTrack | Extension |
| Citation | Class to represent relationships between citing and cited academic works | Thing | | | cito:Citation | OntoTrack | New |
| Author | Represents a person who creates or contributes to an academic work | Thing | | | FOAF:Person | OntoTrack | New |
| InTextReferencePointer | Denoting a single bibliographic reference, that is embedded in the text of a document | Thing | | | | C4O | re-used |
| AreaOfExpertise | Area of expertise of an Author | Thing | | | | OntoTrack | New |
| BookSeries | A sequence of books having certain characteristics in common that can be grouped together | fabio:Expression | | fabio:BookSet | | fabio | re-used |

TABLE 4 Ontotrack data properties.

| Data property name | Domain | Range | Text description | Property characteristics | Origin | Status (New/ adopted/ extended) |
|---|---|---|---|---|---|---|
| authorID | Author | Integer | Unique Id to identify individual authors | Functional | OntoTrack | New |
| bibliographicCouplingFrequency | Citation | Integer | The Bibliographic Coupling Frequency of a specific citation. I.e the two works that are part of that citation | Functional | OntoTrack | New |
| co-citationCount | Citation | Integer | The Co-citation Count of a specific citation. I.e the two works that are part of that citation | Functional | OntoTrack | New |
| institutionName | Affiliation | string | The official name of the workplace. |  | OntoTrack | New |
| hasPublishedDate | AcademicWork | Date | The date on which the paper was published. |  | OntoTrack | Extended |
| keywordValue | Keyword | String | Keyword Title, for string-based comparison |  | OntoTrack | New |
| selfCitationCount | Author | Integer | The number of times an author cited themself in any of their publications |  | OntoTrack | New |
| hasMatchingExpertise | Citation | Boolean | Indicates if the citation matches the expertise of the author | Functional | OntoTrack | New |
| isSelfCitation | Citation | Boolean | Indicates if the citation is by the same author as the citing paper | Functional | OntoTrack | New |
| isCollaboratorCitation | Citation | Boolean | Indicates if the citation is from a collaborator of the author | Functional | OntoTrack | New |
| isJournalSelfCitation | Citation | Boolean | Indicates if the citation is within the same journal as the citing paper | Functional | OntoTrack | New |
| hasUnnaturalIndicator | Citation | Boolean | Boolean value to identify a citation as unnatural or natural | Functional | OntoTrack | New |
| hasNaturalIndicator | Citation | Boolean | Indicates if a citation is classified as natural based on analysis | Functional | OntoTrack | New |
| InTextReferencePointer | Citation | integer | Indicates number of times a particular documented is cited |  | OntoTrack | New |
| selfCitationCount | Situation/ Citation | integer | Indicates number of times citation of document with same author |  | OntoTrack | New |
| hasSameAuthor | Citation | Boolean | Indicates if an author is the same author who is mentioned in another document |  | OntoTrack | New |

assign preliminary classifications, signaling when a citation exhibits behavior that may warrant closer examination.

For example, the following helper rule determines whether two academic works have been co-cited, meaning they appear together in the reference list of the same citing work. This inference simplifies later classification tasks related to citation analysis:

$$\begin{aligned}&Citation(?c1) \wedge Citation(?c2) \wedge AcademicWork(?citing)\\&\wedge AcademicWork(?work1) \wedge AcademicWork(?work2)\\&\wedge hasCitedWork(?c1, ?work1) \wedge hasCitedWork(?c2, ?work2)\\&\wedge hasCitingWork(?c1, ?citing) \wedge hasCitingWork(?c2, ?citing)\\&\rightarrow coCitedWith(?work1, ?work2)\end{aligned} \tag{6}$$

**Indicator rules:**

Indicator rules use the relationships inferred by the helper rules to flag citation patterns that may be natural or unnatural. A citation is flagged as likely natural when the ontology detects meaningful connections such as shared keywords, thematic overlap, or aligned areas of expertise between the citing and cited works. Conversely, a citation is flagged for further analysis when no such connection is identified or when patterns commonly associated with questionable practices such as self-citation or citation cartels are observed. By clearly separating the detection of relevant features from the assignment of these indicators, the framework maintains a structured and transparent approach to analyzing citation behavior.

For instance, the following indicator rule flags a citation is as possibly natural based on the presence of a matching keyword:

$$\begin{aligned}&Citation(?c) \wedge hasMatchingKeyword(?c, true)\\&\rightarrow hasNaturalIndicator(?c, true)\end{aligned} \tag{7}$$

While SWRL provides a robust framework for extending OWL ontologies with rule-based reasoning, it has certain limitations. One significant limitation is that SWRL is inherently monotonic, meaning it does not support negation or retraction. This poses challenges when expressing the absence of relationships between entities, such as identifying cases where no link exists between

TABLE 5 Ontotrack object properties Part 1.

| Object property | Domain | Range | Characteristics | Constraints | Origin | Status (New/ adopted/ extended) |
|---|---|---|---|---|---|---|
| belongsToArea | Keyword | HasAreaOfExpertise | Transitive, reflexive, inverse asociatedsKeyword | - | OntoTrack | New |
| hasAreaOfExpertise | Author | HasAreaOfExpertise | - | owl:minCardinality - 1 | OntoTrack | New |
| coCitedWith | AcademicWork | AcademicWork | Symmetric, Irreflexive | - | OntoTrack | New |
| asociatedKeyword | area_of_expertise | Keyword | Inverse BelongsToArea | - | OntoTrack | New |
| hasKeyword | AcademicWork | Keyword | Asymmetric, irreflexive | - | OntoTrack | New |
| HasPublished | author | AcademicWork | inverse publishedBy | owl:minCardinality - 1 | OntoTrack | New |
| hasAffiliation | author | Affiliation | - | - | OntoTrack | New |
| HasWorkedWith | author | Author | Symmetric, irreflexive | - | OntoTrack | New |
| isLinkedTo | Keyword | Keyword | Symmetric, Transitive, Irreflexive | - | OntoTrack | New |
| PublishedBy | AcademicWork | Author | inverse HasPublished | owl:minCardinality - 1 | OntoTrack | New |
| PublishedIn | AcademicWork, PrePrint, JournalArticle, ProceedingsPaper | AcademicWork, Journal, AcademicProceedings | Functional, inverse hasContent | owl:minCardinality - 1, owl:maxCardinality -1 | OntoTrack | New |
| HasInstance | Journal,AcademicProceedings | JournalIssue, ProceedingsVolume | Inverse InstanceOf | - | OntoTrack | New |
| InstanceOf | JournalIssue, ProceedingsVolume | Journal,AcademicProceedings | Inverse HasInstance | - | OntoTrack | New |
| HasCitingWork | Citation | AcademicWork | Inverse HasCitedWork, Functional | owl:minCardinality - 1, owl:maxCardinality -1 | OntoTrack | Extension:<br>oco:has citing entity |
| HasCitedWork | Citation | AcademicWork | InverseHasCiting Work, Functional | owl:minCardinality - 1, owl:maxCardinality -1 | OntoTrack | Extension:<br>oco:has cited entity |

keywords or citations. For example, SWRL cannot handle negated expressions like:

$$Keyword(?a) \wedge Keyword(?b) \wedge \neg isLinkedTo(?a, ?b) \quad (8)$$

Additionally, SWRL struggles with counting and aggregation operations, such as determining whether a publication has exactly one author, due to the open-world assumption, potentially leading to inconsistencies when new information becomes available. Furthermore, running SWRL rules can be computationally expensive, especially for large datasets or complex rules, and often requires manual consistency checks by users (O'Connor and Das, 2006).

### 3.1.3 Data preparation

Citation triples commonly represent a direct relationship between two academic works through an object property like `oco:cites` with the academic works as object and subject. Citation instances are created to enable a better analysis of the citation itself in OntoTrack. New triples are created from the original citation triples, ensuring each citation is represented as an object with explicit links to both the citing and cited works, rather than as a direct relationship property as follows:

```
Paper A   oco:cites   Paper B
->
Citation 1   ont:hasCitingWork   Paper A    (9)
Citation 1   ont:hasCitedWork   Paper B
```

TABLE 6 Ontotrack object properties part 2.

| Object property | Text description | Reasoning |
|---|---|---|
| belongsToArea | Connects a keyword to an area of expertise | Facilitates the classification of keywords within specific areas of expertise. |
| hasAreaOfExpertise | An author possesses expertise in a certain area | Used to relate authors to their fields of expertise. |
| hasContent | Links the content of a journal, such as journal articles to the journal | Enables linking different forms or editions of journals.<br>Ensures that the content within an academic work is uniquely<br>associated with it, maintaining consistency and integrity of academic data. |
| asociatedKeyword | An area of expertise is associated with specific keywords | Used for tagging areas with relevant keywords. |
| hasKeyword | An AcademicWork may involve multiple keywords | Helps in the categorization and searchability of papers. |
| HasPublished | An author has published an Academic Work | Links authors to their publications. |
| hasAffiliation | An author has worked at a particular workplace | Useful for tracking the employment history of authors. |
| HasWorkedWith | An author has collaborated with another author | Important for analyzing collaboration networks. |
| isLinkedTo | Links areas of expertise with keywords bi-directionally | Facilitates navigation and understanding of relationships between areas and keywords. |
| PublishedBy | An academic work is published by an author | Connects papers directly to their authors. |
| PublishedIn | An academic work is published in a journal | Denotes the publication outlet for an academic work. |
| HasInstance | A journal may have instances, such as special editions or issues. | Allows to link each journal to its instances and trace metadata |
| InstanceOf | Each published instance of a journal can be linked to the journal | Allows for better citation cartel identification |
| HasCitingWork | Each citation has one citing work | Add logical reasoning: cited work cannot be published after citing work |
| HasCitedWork | Each citation has one cited work | Add logical reasoning: cited work cannot be published after citing work |

TABLE 7 Logical description of object properties in OntoTrack.

| Object property | Text description | Reasoning context |
|---|---|---|
| belongsToArea | Connects a keyword to an area of expertise | Keywords $\in$ AcademicWork<br>Academic Work $\subseteq$ Area of Expertise<br>Keywords $\in$ Area of Expertise |
| hasAreaOfExpertise | An author possesses expertise in a certain area | Academic Work $\subseteq$ Author<br>Keywords $\in$ Academic Work<br>Keywords $\in$ Area of Expertise<br>$\therefore$ Area of Expertise $\subseteq$ Author |
| hasContent | Links the content of a journal, such as journal articles to the journal | $\because$ Academic Work $\subseteq$ Journal<br>$\therefore$ Journal $\subseteq$ Academic Work |
| asociatedKeyword | An area of expertise is associated with specific keywords | Keywords $\in$ Area of Expertise<br>Area of Expertise $\subseteq$ Author<br>$\because$ AcademicWork $\subseteq$ Author<br>$\therefore$ Keywords $\subseteq$ Academic Work |
| hasKeyword | An AcademicWork may involve multiple keywords | Area of Expertise $\in$ Paper<br>$\because$ AcademicWork $\subseteq$ Author<br>AcademicWork $\subseteq$ Publication Keywords<br>$\because$ Keywords $\in$ AcademicWork<br>$\therefore$ Keywords $\in$ Area of Expertise |
| HasPublished | An author has published an Academic Work | $\because$ Author $\subseteq$ Academic Work<br>$\therefore$ Academic Work $\subseteq$ Author |

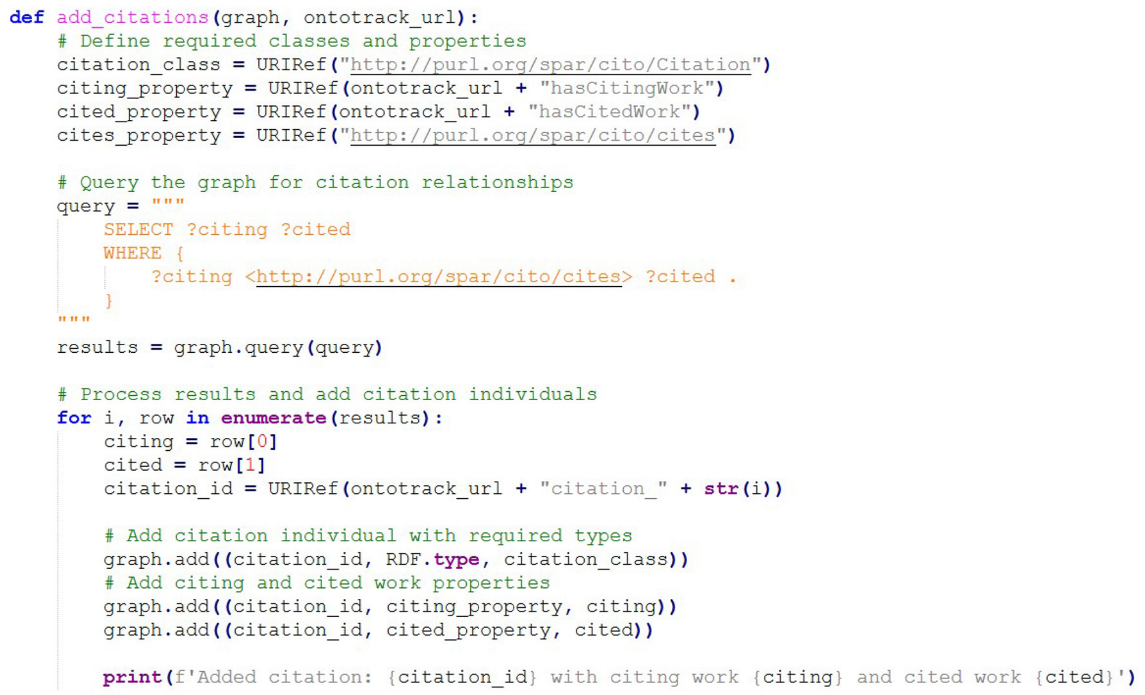
This representation not only aligns with the OntoTrack model but also facilitates the creation of SWRL rules based on citation instances.

Further transformations are performed to enable reasoning through attribute comparisons, such as authors or keywords. Since direct comparisons between the URIs of the instances are not feasible, literal or integer values are created as data properties of the corresponding instances. These enhancements improve the ontology's capacity for reasoning by enabling effective comparisons of citation-related attributes.

The transformation and enrichment process is automated using a Python script. The `addcitations` script function (Figure 2) demonstrates how citation instances are generated and enriched. It queries the ontology to identify existing citation relationships, creates citation instances, and links them to their citing and cited works using ontology-specific properties. Each instance is assigned a unique identifier, ensuring structured and consistent citation data in the ontology. These transformations prepare the dataset for reasoning tasks and form a robust foundation for applying SWRL rules to detect unnatural citation patterns.

### 3.1.4 OntoTrack SPARQL endpoint

To support experimentation and query execution, a SPARQL Endpoint has been established for OntoTrack, which can also be made available on request. In the future, the OntoTrack SPARQL Endpoint will be made publicly accessible as the GraphDB database is expanded by integrating bigger citation datasets. OntoTrack SPARQL Endpoint plays a central role in enabling advanced querying capabilities and seamless integration



```python
def add_citations(graph, ontotrack_url):
    # Define required classes and properties
    citation_class = URIRef("http://purl.org/spar/cito/Citation")
    citing_property = URIRef(ontotrack_url + "hasCitingWork")
    cited_property = URIRef(ontotrack_url + "hasCitedWork")
    cites_property = URIRef("http://purl.org/spar/cito/cites")

    # Query the graph for citation relationships
    query = """
        SELECT ?citing ?cited
        WHERE {
            ?citing <http://purl.org/spar/cito/cites> ?cited .
        }
    """
    results = graph.query(query)

    # Process results and add citation individuals
    for i, row in enumerate(results):
        citing = row[0]
        cited = row[1]
        citation_id = URIRef(ontotrack_url + "citation_" + str(i))

        # Add citation individual with required types
        graph.add((citation_id, RDF.type, citation_class))
        # Add citing and cited work properties
        graph.add((citation_id, citing_property, citing))
        graph.add((citation_id, cited_property, cited))

        print(f'Added citation: {citation_id} with citing work {citing} and cited work {cited}')
```

FIGURE 2
Sample Python implementation for triplication.

of semantic reasoning. Unlike the native Protégé SPARQL Editor, GraphDB supports executing complex SPARQL queries over inferred triples. For experimental purposes, OntoTrack ontology and the corresponding sample dataset are loaded into the GraphDB database. For the citation classification, the queries focus on identifying citations with specific data properties or those flagged with the properties *hasUnnaturalIndicator* or *hasNaturalIndicator*, forming the basis for detecting irregular citation patterns. The query execution results are displayed directly in the Query editor and can be exported for further analysis. The citation classification provides important information to determine the legitimacy of the citations.

# 4 Implementation

## 4.1 Use case implementation

This section demonstrates the application of OntoTrack to identify an unnatural citation through a use case involving academic works from a Proof of Concept (POC) dataset. The primary question addressed is whether a given citation is unnatural, determined by evaluating factors such as thematic overlap and keyword matching, which assess the alignment of research areas and relevance between the citing and cited works. To evaluate multiple individual influencing factors, a set of Competency Questions (CQs) was formed, from which rules for the logical reasoner were derived to be evaluated. In the first step of the analysis process, a Python script is run to prepare the POC data in OntoTrack. A citation object is created with the properties *hasCitedWork* and *hasCitingWork*. Then, additional property values are generated by the logical reasoner, which applies the SWRL helper rules to infer relationships such as the *isLinkedTo* or *hasWorkedTogether* properties based on keyword, author data or research area similarities. These rules produces the inferred data used to directly infer the citation indicator data properties required for the subsequent analysis.

### 4.1.1 Proof of concept dataset

As an integral part of the OntoTrack Solution, and for evaluation and experimental purposes, a sample dataset was required that is structured according to the OntoTrack data model. For this purpose, we investigated various existing citation networks such as citeseer (Rossi and Ahmed, 2015) or AMiner (Tang et al., 2008), which provide extensive citation networks but often lack detailed information such as keywords, author information, etc. A similar problem exists with comprehensive semantic graphs such as SemOpenAlex (Färber et al., 2023). Even though these repositories are noticeably more enriched in metadata and provide extensive author information, they lack keywords and area of expertise information, which are the key attributes of the citation tracking dataset. This issue led to the creation of a POC dataset to fully assess and utilize the potential of the OntoTrack Solution. The dataset consists of academic works based on the connected papers graph (Connected Papers, 2024) with the topic of Natural Language Processing and Generative Pre-trained Transformers. All entities were collected together with their corresponding object properties such as *cites* and *isCitedBy* to construct the citation graph.

### 4.1.2 Competency questions

Building upon the requirements specification outlined in Section 3.1.1, a set of Competency Questions (CQs) has been formalized to evaluate the core capabilities of the OntoTrack solution. These questions ensure that the ontology is adequately structured to address the complex criteria associated with citation analysis and the detection of unnatural citation behavior. In addition to defining the informational scope of the ontology, these CQs also serve as the foundation for designing rule-based reasoning that enables the detection of unnatural citation behavior.

Table 8 presents the list of CQs that OntoTrack is designed to answer. Each question targets a specific aspect of citation behavior relevant to identifying potential irregularities or biases in academic referencing.

CQ0 serves as a general query, asking the ontology about "which academic work has unnaturally cited another academic work." The further questions target different criteria for identifying unnatural citations. CQ1 assesses whether the citations are technically sound. CQ2 and CQ3 focus on the thematic overlap between the scientific papers, such as keywords and matching areas of expertise. CQ4 and CQ5 are used to track specific citation patterns, including repeated citations from the same journal, self-citations, and citations from previous collaborators. Finally, CQ6, CQ7 and CQ8 measure the interrelationship strength between two documents using bibliographic coupling frequency, co-citation, and inter-citation counts. As discussed in Section 2.1, there are no recommended values for what constitutes a high or low interrelationship strength, so these metrics must be defined based on the specific use case. The SWRL rules implemented in OntoTrack are directly derived from these CQs, ensuring that the ontology not only stores descriptive information but also infers new relationships and classifies citations as natural or unnatural based on logical reasoning.

TABLE 8 List of competency questions (CQs) that can be answered with OntoTrack.

| CQ | Competency questions |
|---|---|
| CQ 0 | Which Academic Work has unnaturally cited another Academic Work? |
| CQ 1 | Is the citation technically sound, ensuring all in-text reference pointers have corresponding bibliographic entries and vice versa? |
| CQ 2 | Which Citing Academic Works do not have matching or indirectly linked keywords with their cited counterparts? |
| CQ 3 | Which Citing Academic Works do not have matching areas of expertise and/or keywords with their cited counterparts? |
| CQ 4 | Which academic works cite multiple works from the same journal, by the same author, or from previous colleagues? |
| CQ 5 | For which citing academic works have the authors worked with their cited counterparts either at the same institution or on a previous paper? |
| CQ 6 | What is the bibliographic coupling frequency between authors of citing and cited works? |
| CQ 7 | What is the co-citation count of two academic works? |
| CQ 8 | What is the inter-citation count between two academic works? |

### 4.1.3 SWRL rule implementation

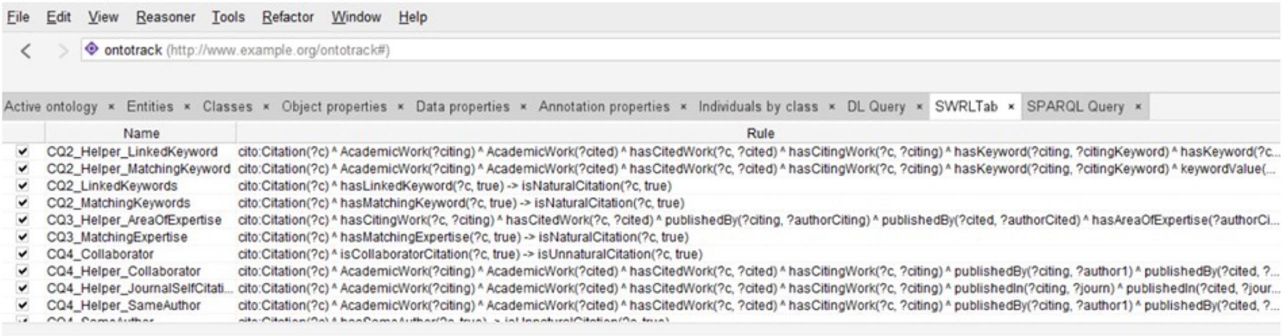
In the context of OntoTrack, SWRL rules were implemented to identify and flag citation patterns by inferring relationships from the underlying data. These rules do not produce definitive classifications, but generate indicators that highlight when a citation exhibits characteristics associated with natural or potentially unnatural citation behavior. The SWRLTab Protégé plugin is used for the rule implementation, providing a development environment for working with SWRL rules and allowing us to create rules directly in the Protégé editor, as shown in Figure 3. Here, a snapshot of the custom rules in the SWRLTab is shown. These rules are then integrated into the ontology RDF file.

Rules in OntoTrack serve distinct but complementary purposes. Helper rules first establish intermediate relationships between entities in the dataset, such as matching keywords, shared expertise, or bibliographic coupling, that supply the evidential basis for subsequent reasoning steps.These inferred properties are then used by the indicator rules, which assign preliminary signals regarding whether a citation aligns with features typically interpreted as natural (e.g., thematic overlap) or with features that may warrant further examination (e.g., absence of topical similarity or recurrent self-citation).** An example is shown in Table 9. This table outlines the logic of a helper rule for detecting keyword matches, followed by an indicator rule that uses the inferred information to flag the corresponding citation. These rules assert the Boolean data properties *hasNaturalIndicator* and *hasUnnaturalIndicator* when their respective conditions are satisfied, providing structured signals for downstream analysis rather than definitive judgements about citation behavior.

## 5 Results and evaluation

### 5.1 Data enrichment

Before the evaluation process, the first step was to enrich the gathered data in the ontology with additional inferred information. For this purpose, a Python script which imports the SPARQLWrapper module to parameterise the queries and run them dynamically and efficiently was built. The script first queries the RDF store to retrieve all existing citation relationships between academic works. For each identified citation relationship, a new RDF triple is constructed to represent the citation as an individual entity. Additionally, it sets the various Object Properties such as *hasCitedWork* and *hasCitingWork* to enable querying from academic work to citation. These newly created citation entities are then inserted back into the RDF store using SPARQL updates. This process enriched the original RDF data by adding explicit citation entities, facilitating more detailed and structured citation analysis. Furthermore, queries are run to generate data properties which hold Author IDs and keyword values and various counts are calculated, such as citation count and the bibliographic coupling frequency, which are necessary for the analysis.



| | Name | Rule |
|---|---|---|
| ✓ | CQ2_Helper_LinkedKeyword | cito:Citation(?c) ^ AcademicWork(?citing) ^ AcademicWork(?cited) ^ hasCitedWork(?c, ?cited) ^ hasCitingWork(?c, ?citing) ^ hasKeyword(?citing, ?citingKeyword) ^ hasKeyword(?c... |
| ✓ | CQ2_Helper_MatchingKeyword | cito:Citation(?c) ^ AcademicWork(?citing) ^ AcademicWork(?cited) ^ hasCitedWork(?c, ?cited) ^ hasCitingWork(?c, ?citing) ^ hasKeyword(?citing, ?citingKeyword) ^ keywordValue(... |
| ✓ | CQ2_LinkedKeywords | cito:Citation(?c) ^ hasLinkedKeyword(?c, true) -> isNaturalCitation(?c, true) |
| ✓ | CQ2_MatchingKeywords | cito:Citation(?c) ^ hasMatchingKeyword(?c, true) -> isNaturalCitation(?c, true) |
| ✓ | CQ3_Helper_AreaOfExpertise | cito:Citation(?c) ^ hasCitingWork(?c, ?citing) ^ hasCitedWork(?c, ?cited) ^ publishedBy(?citing, ?authorCiting) ^ publishedBy(?cited, ?authorCited) ^ hasAreaOfExpertise(?authorCi... |
| ✓ | CQ3_MatchingExpertise | cito:Citation(?c) ^ hasMatchingExpertise(?c, true) -> isNaturalCitation(?c, true) |
| ✓ | CQ4_Collaborator | cito:Citation(?c) ^ isCollaboratorCitation(?c, true) -> isUnnaturalCitation(?c, true) |
| ✓ | CQ4_Helper_Collaborator | cito:Citation(?c) ^ AcademicWork(?citing) ^ AcademicWork(?cited) ^ hasCitedWork(?c, ?cited) ^ hasCitingWork(?c, ?citing) ^ publishedBy(?citing, ?author1) ^ publishedBy(?cited, ?... |
| ✓ | CQ4_Helper_JournalSelfCitati... | cito:Citation(?c) ^ AcademicWork(?citing) ^ AcademicWork(?cited) ^ hasCitedWork(?c, ?cited) ^ hasCitingWork(?c, ?citing) ^ publishedIn(?citing, ?journ) ^ publishedIn(?cited, ?jour... |
| ✓ | CQ4_Helper_SameAuthor | cito:Citation(?c) ^ AcademicWork(?citing) ^ AcademicWork(?cited) ^ hasCitedWork(?c, ?cited) ^ hasCitingWork(?c, ?citing) ^ publishedBy(?citing, ?author1) ^ publishedBy(?cited, ?... |
| ✓ | CQ4_SameAuthor | cito:Citation(?c) ^ hasSameAuthor(?c, true) -> isUnnaturalCitation(?c, true) |

FIGURE 3
A sample snapshot of SWRL rules implementation by using Protégé SWRLTab as rules editor.

TABLE 9 CQ2 rule logic.

**Competency question**
CQ2: Which citing academic works have matching or indirectly linked keywords with their cited counterparts?

**Rule textual description**
If citing and cited Academic Works have matching Keywords, then the Citation is natural.
If citing and cited Academic Works have linked Keywords, then the Citation is natural.

**Helper rules**

```
Citation(?c)
^AcademicWork(?citing) ^AcademicWork(?cited)
^hasCitedWork(?c, ?cited) ^hasCitingWork(?c, ?citing)
^hasKeyword(?citing, ?citingKeyword) ^keywordValue(?citingKeyword, ?Value1)
^hasKeyword(?cited, ?citedKeyword) ^keywordValue(?citedKeyword, ?Value2)
^swrlb:equal(?Value1, ?Value2) → hasMatchingKeyword(?c, true)

 Citation(?c) ^AcademicWork(?citing) ^AcademicWork(?cited)
^hasCitedWork(?c, ?cited) ^hasCitingWork(?c, ?citing)
^hasKeyword(?citing, ?citingKeyword) ^hasKeyword(?cited, ?citedKeyword)
^isLinkedTo(?citingKeyword, ?citedKeyword) → hasLinkedKeyword(?c, true)
```

**Classification rules**

```
Citation(?c) ^hasLinkedKeyword(?c, true) → hasNaturalIndicator(?c, true)
Citation(?c) ^hasMatchingKeyword(?c, true) → hasNaturalIndicator(?c, true)
```

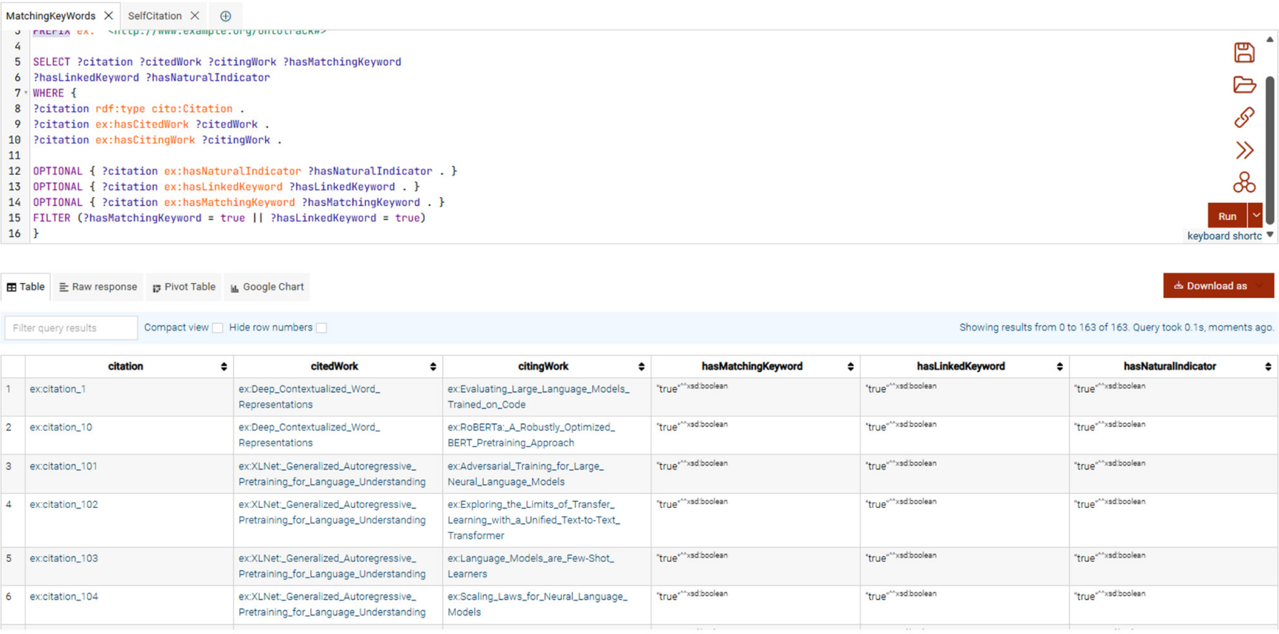
**Reasoning result**
The reasoning identifies natural citations by applying rules that classify citations as natural when the citing and cited academic works had matching or indirectly linked keywords. The rule asserts the presence of a match between keywords of the two works, ensuring that citations with aligned topics are classified as natural. This confirms that the citation relationship is likely based on thematic relevance, supporting the legitimacy of the citation.

**SPARQL query**

```
SELECT ?citation ?citedWork ?citingWork ?hasMatchingKeyword ?hasLinkedKeyword ?hasNaturalIndicator
?hasUnnaturalIndicator
WHERE {
?citation rdf:type cito:Citation .
?citation ex:hasCitedWork ?citedWork .
?citation ex:hasCitingWork ?citingWork .
OPTIONAL { ?citation ex:hasUnnaturalIndicator ?hasUnnaturalIndicator . }
OPTIONAL { ?citation ex:hasNaturalIndicator ?hasNaturalIndicator . }
OPTIONAL { ?citation ex:hasLinkedKeyword ?hasLinkedKeyword . }
OPTIONAL { ?citation ex:hasMatchingKeyword ?hasMatchingKeyword . }
FILTER (?hasMatchingKeyword = true || ?hasLinkedKeyword = true)
}
```

## 5.2 Reasoning SWRL rules

The Pellet reasoner was utilized to apply the SWRL rules and automatically generate the inferred relationships required for the indicator-based analysis, such as the *hasUnnaturalIndicator* and *hasNaturalIndicator* properties. The Pellet reasoner is particularly beneficial because it natively supports the W3C SWRL specification and apply SWRL rules as part of the reasoning process (DeBellis, 2020). The SWRL rules mentioned in Section 4.1.3 are executed by using the reasoner, and the inferred data properties are added to the ontology. The reasoning process derived from the SWRL rules provided the inferred evidence required for evaluating the CQs. The rules successfully inferred indicators associated with citation behavior, such as thematic similarity (CQ2), expertise alignment



```
PREFIX ex:  <http://www.example.org/ontotrack#>

SELECT ?citation ?citedWork ?citingWork ?hasMatchingKeyword
?hasLinkedKeyword ?hasNaturalIndicator
WHERE {
?citation rdf:type cito:Citation .
?citation ex:hasCitedWork ?citedWork .
?citation ex:hasCitingWork ?citingWork .

OPTIONAL { ?citation ex:hasNaturalIndicator ?hasNaturalIndicator . }
OPTIONAL { ?citation ex:hasLinkedKeyword ?hasLinkedKeyword . }
OPTIONAL { ?citation ex:hasMatchingKeyword ?hasMatchingKeyword . }
FILTER (?hasMatchingKeyword = true || ?hasLinkedKeyword = true)
}
```

Showing results from 0 to 163 of 163. Query took 0.1s, moments ago.

| | citation | citedWork | citingWork | hasMatchingKeyword | hasLinkedKeyword | hasNaturalIndicator |
|---|---|---|---|---|---|---|
| 1 | ex:citation_1 | ex:Deep_Contextualized_Word_Representations | ex:Evaluating_Large_Language_Models_Trained_on_Code | "true"^^xsd:boolean | "true"^^xsd:boolean | "true"^^xsd:boolean |
| 2 | ex:citation_10 | ex:Deep_Contextualized_Word_Representations | ex:RoBERTa:_A_Robustly_Optimized_BERT_Pretraining_Approach | "true"^^xsd:boolean | "true"^^xsd:boolean | "true"^^xsd:boolean |
| 3 | ex:citation_101 | ex:XLNet:_Generalized_Autoregressive_Pretraining_for_Language_Understanding | ex:Adversarial_Training_for_Large_Neural_Language_Models | "true"^^xsd:boolean | "true"^^xsd:boolean | "true"^^xsd:boolean |
| 4 | ex:citation_102 | ex:XLNet:_Generalized_Autoregressive_Pretraining_for_Language_Understanding | ex:Exploring_the_Limits_of_Transfer_Learning_with_a_Unified_Text-to-Text_Transformer | "true"^^xsd:boolean | "true"^^xsd:boolean | "true"^^xsd:boolean |
| 5 | ex:citation_103 | ex:XLNet:_Generalized_Autoregressive_Pretraining_for_Language_Understanding | ex:Language_Models_are_Few-Shot_Learners | "true"^^xsd:boolean | "true"^^xsd:boolean | "true"^^xsd:boolean |
| 6 | ex:citation_104 | ex:XLNet:_Generalized_Autoregressive_Pretraining_for_Language_Understanding | ex:Scaling_Laws_for_Neural_Language_Models | "true"^^xsd:boolean | "true"^^xsd:boolean | "true"^^xsd:boolean |

FIGURE 4
Results of CQ2 query when executed over the OntoTrack SPARQL Endpoint.

TABLE 10 CQ4 rule logic.

**Competency question**
CQ4: Which academic works cite multiple works from the same journal, by the same author, or from previous colleagues?

**Rule textual description**
If the citing and cited papers have the same authors, it is a self-citation.
If the citation is a self-citation, then the citation is possibly unnatural.

**Helper rule**

```
Citation(?c)
^AcademicWork(?citing) ^AcademicWork(?cited)
^hasCitedWork(?c, ?cited) ^hasCitingWork(?c, ?citing)
^publishedBy(?citing, ?author1) ^publishedBy(?cited, ?author2)
^authorID(?author1, ?id) ^authorID(?author2, ?id)
→ isSelfCitation(?c, true)
```

**Classification rule**

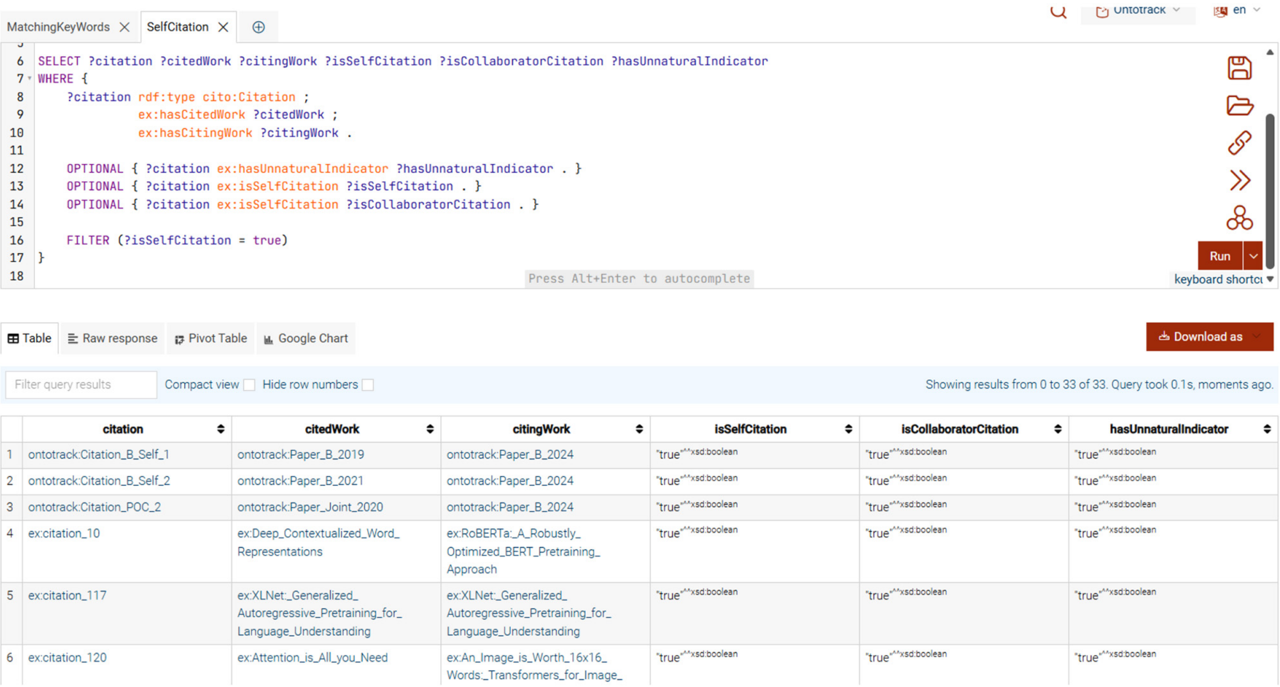
```
Citation(?c) ^isSelfCitation(?c, true)
→ hasUnnaturalIndicator(?c, true)
```

**Reasoning result**
The reasoner detects unnatural citations by identifying instances of potential self-citation, such as when a citing paper references multiple works by the same author or from former colleagues. By identifying these patterns, the reasoner highlights cases where citation behavior could be seen as potentially biased or unnatural due to close professional relationships rather than based on the academic merit of the cited work.

**SPARQL query**

```
SELECT ?citation ?citedWork ?citingWork ?isSelfCitation ?hasNaturalIndicator ?hasUnnaturalIndicator
WHERE {
?citation rdf:type cito:Citation .
?citation ex:hasCitedWork ?citedWork .
?citation ex:hasCitingWork ?citingWork .
OPTIONAL { ?citation ex:hasUnnaturalIndicator ?hasUnnaturalIndicator . }
OPTIONAL { ?citation ex:hasNaturalIndicator ?hasNaturalIndicator . }
OPTIONAL { ?citation ex:isSelfCitation ?isSelfCitation . }
FILTER (?isSelfCitation = true)
}
```

(CQ3), and author-related patterns (CQ4). These rules provided a comprehensive review of citation practices and ensured that unnatural citation patterns were effectively flagged. Overall, the reasoning process enriched the ontology with inferred semantic relationships and supported a nuanced, indicator-driven analysis of citation behavior.



```
SELECT ?citation ?citedWork ?citingWork ?isSelfCitation ?isCollaboratorCitation ?hasUnnaturalIndicator
WHERE {
    ?citation rdf:type cito:Citation ;
              ex:hasCitedWork ?citedWork ;
              ex:hasCitingWork ?citingWork .

    OPTIONAL { ?citation ex:hasUnnaturalIndicator ?hasUnnaturalIndicator . }
    OPTIONAL { ?citation ex:isSelfCitation ?isSelfCitation . }
    OPTIONAL { ?citation ex:isSelfCitation ?isCollaboratorCitation . }

    FILTER (?isSelfCitation = true)
}
```

Showing results from 0 to 33 of 33. Query took 0.1s, moments ago.

| | citation | citedWork | citingWork | isSelfCitation | isCollaboratorCitation | hasUnnaturalIndicator |
|---|---|---|---|---|---|---|
| 1 | ontotrack:Citation_B_Self_1 | ontotrack:Paper_B_2019 | ontotrack:Paper_B_2024 | "true"^^xsd:boolean | "true"^^xsd:boolean | "true"^^xsd:boolean |
| 2 | ontotrack:Citation_B_Self_2 | ontotrack:Paper_B_2021 | ontotrack:Paper_B_2024 | "true"^^xsd:boolean | "true"^^xsd:boolean | "true"^^xsd:boolean |
| 3 | ontotrack:Citation_POC_2 | ontotrack:Paper_Joint_2020 | ontotrack:Paper_B_2024 | "true"^^xsd:boolean | "true"^^xsd:boolean | "true"^^xsd:boolean |
| 4 | ex:citation_10 | ex:Deep_Contextualized_Word_Representations | ex:RoBERTa:_A_Robustly_Optimized_BERT_Pretraining_Approach | "true"^^xsd:boolean | "true"^^xsd:boolean | "true"^^xsd:boolean |
| 5 | ex:citation_117 | ex:XLNet:_Generalized_Autoregressive_Pretraining_for_Language_Understanding | ex:XLNet:_Generalized_Autoregressive_Pretraining_for_Language_Understanding | "true"^^xsd:boolean | "true"^^xsd:boolean | "true"^^xsd:boolean |
| 6 | ex:citation_120 | ex:Attention_is_All_you_Need | ex:An_Image_is_Worth_16x16_Words:_Transformers_for_Image_ | "true"^^xsd:boolean | "true"^^xsd:boolean | "true"^^xsd:boolean |

FIGURE 5
Results of CQ4 query when executed over the OntoTrack SPARQL Endpoint.

## 5.3 Querying results

The results from the SPARQL queries returned detailed information about the citations (as shown in Figures 4, 5). For each citation, the query linked the citation to both the citing and the cited paper, allowing us to trace these relationships. The results also show where the reasoning process identified specific citation behavior/attributes of each citation entity, represented as data properties, including the indicator values inferred during reasoning.

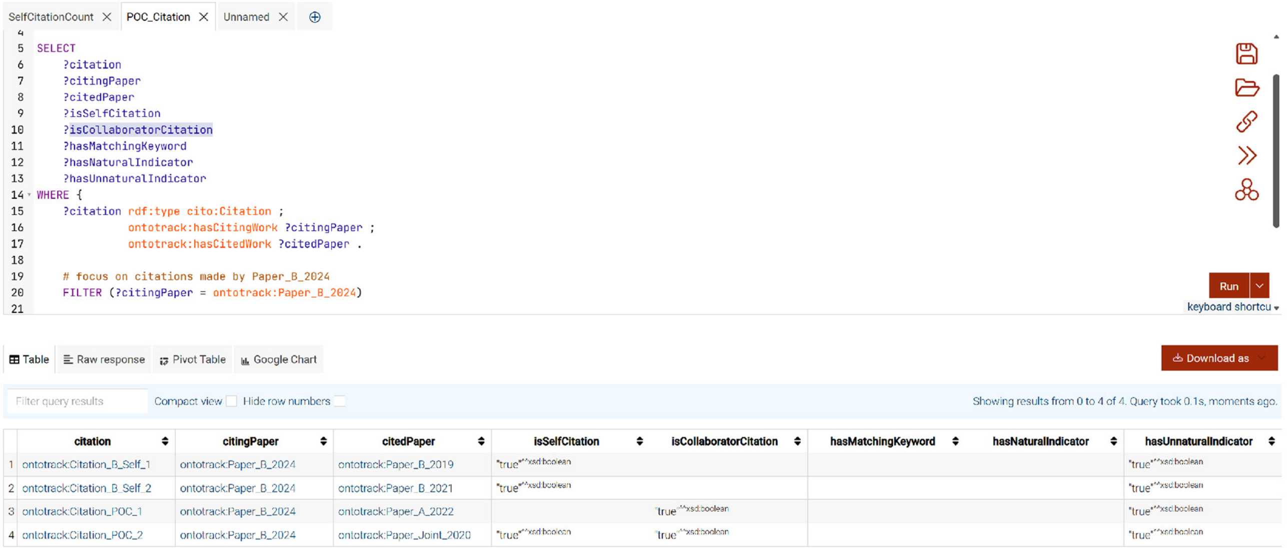
In the example of CQ2 (Table 9), the results show that 163 citation instances within the POC dataset have either matching and/or linked keywords of the citing and cited academic works. These citations were all assigned the *hasNaturalIndicator* flag, reflecting that they exhibit thematic overlap commonly associated with natural citation behavior (Figure 4).

For the example CQ4 (Table 10), the query identified 30 specific citation instances marked as self-citations using the inferred property *ontotrack:isSelfCitation* (Figure 5). These citations were assigned the *hasUnnaturalIndicator* flag, signaling the presence of author-overlap patterns that may warrant closer examination. These results illustrate how the reasoning process detects self-citation patterns as potential indicators of unnatural citation behavior rather than definitive evidence of misconduct.

To verify the accuracy of the inferred data, a random sample of academic papers from the proof-of-concept (POC) dataset was selected and manually cross-checked against the original sources. For the subset of inferred axioms identified as self-citations, each corresponding paper was examined to confirm the presence of an actual self-citation. The verification confirmed that the system accurately inferred these relationships, and repeating this process across all papers in the POC dataset validated the overall reliability of the inferred data.

In addition to the validation against real-world citation data, a small number of deliberately constructed proof-of-concept (POC) citation instances were inserted into the dataset to explicitly demonstrate how the SWRL rules and indicator logic behave under controlled conditions. These synthetic citations were designed to reflect distinct citation scenarios, such as repeated self-citation and citation of a jointly authored publication. The POC instances therefore do not represent empirical evidence of problematic behavior, but serve as illustrative test cases to verify that the helper rules and indicator rules interact as intended. As shown in Figure 6, the reasoning process correctly inferred self-citation indicators for citations where the citing and cited works share the same author, and assigned the corresponding unnatural indicator flags. In contrast, the POC citations that involve collaboration history or shared authorship in prior work demonstrate how multiple indicators can co-occur, reflecting the nuanced and non-deterministic nature of citation analysis in OntoTrack.

This analysis offers a deeper understanding of the citation relationships, providing a comprehensive view of citation practices within the dataset.These examples demonstrate how the reasoner can automatically derive meaningful semantic patterns within the citation network, supporting more sophisticated analyses in subsequent stages.



```
SELECT
    ?citation
    ?citingPaper
    ?citedPaper
    ?isSelfCitation
    ?isCollaboratorCitation
    ?hasMatchingKeyword
    ?hasNaturalIndicator
    ?hasUnnaturalIndicator
WHERE {
    ?citation rdf:type cito:Citation ;
              ontotrack:hasCitingWork ?citingPaper ;
              ontotrack:hasCitedWork ?citedPaper .

    # focus on citations made by Paper_B_2024
    FILTER (?citingPaper = ontotrack:Paper_B_2024)

```

Showing results from 0 to 4 of 4. Query took 0.1s, moments ago.

| | citation | citingPaper | citedPaper | isSelfCitation | isCollaboratorCitation | hasMatchingKeyword | hasNaturalIndicator | hasUnnaturalIndicator |
|---|---|---|---|---|---|---|---|---|
| 1 | ontotrack:Citation_B_Self_1 | ontotrack:Paper_B_2024 | ontotrack:Paper_B_2019 | "true"^^xsd:boolean | | | | "true"^^xsd:boolean |
| 2 | ontotrack:Citation_B_Self_2 | ontotrack:Paper_B_2024 | ontotrack:Paper_B_2021 | "true"^^xsd:boolean | | | | "true"^^xsd:boolean |
| 3 | ontotrack:Citation_POC_1 | ontotrack:Paper_B_2024 | ontotrack:Paper_A_2022 | | "true"^^xsd:boolean | | | "true"^^xsd:boolean |
| 4 | ontotrack:Citation_POC_2 | ontotrack:Paper_B_2024 | ontotrack:Paper_Joint_2020 | "true"^^xsd:boolean | "true"^^xsd:boolean | | | "true"^^xsd:boolean |

FIGURE 6
POC Dummy Citation showing unnatural citation behavior identified by rules.

# 6 Discussion

## 6.1 Performance analysis

Protégé relies on a reasoner to execute SWRL rules that make logical inferences over the ontology. These rules can generate new assertions, which may trigger further inferences, resulting in substantial computational load and memory usage. To support these reasoning tasks, the Java Virtual Machine (JVM) running Protégé was configured with an initial heap size of 512 MB and a maximum heap size of 24 GB (`-Xms512m -Xmx24576m`) via the `run.bat` file. This allocation ensures that sufficient memory is available for ontology processing, even with complex rules or moderately large datasets.

With this configuration, the Pellet reasoner processed a proof-of-concept dataset containing 14,964 triples in 29,073ms, using up to approximately 2.3GB of heap memory. For larger datasets, such as those being collected as part of the NFDI4DS[2] project (and Open Research Knowledge Graph (ORKG[3]) as a collaboration Project), the system continues to provide accurate reasoning results when applied to a sample dataset. Although processing time increases due to the larger computational load. This setup provides a stable environment for ontology development and reasoning while making efficient use of available system resources. Currently (as part of future work), we are working on implementing the OntoTrack system on the DBLP data[4] (i.e., N-Triple version of DBLP publications data), which is 2.7 GB in size (in zip format) and reaches 32.7 GB when unzipped. For such a large dataset we are planning to use the divide-and-conquer approach (initially), by discipline-wise dividing the datasets and analyzing results.

## 6.2 Challenges and limitations

Identifying unnatural citations in academic works poses several challenges and limitations, which are critical to our analysis.

### 6.2.1 Keywords and areas of expertise

A major challenge that we faced while building up the dataset for our case study and experimentation was the limited availability of keywords for the papers analyzed. Many academic works do not include a keywords section, making it difficult to conduct a detailed analysis based on this attribute. Keywords play a key role in understanding the context of the research and act as a simple measure to compare the relevance between two papers. When this data is missing, it limits the reasoning and data-matching ability of the OntoTrack solution to reach accurate conclusions. In our research, we found that for most datasets, keyword data was often blank and there was no information on the relationship between keywords. This limitation extends to the area of expertise as well, where insufficient data can limit our ability to assess the relevance and overlap of research topics accurately.

### 6.2.2 Classifying self-citation behavior

In some situations, a high occurrence of self-citations can be a required behavior and not with the intent to push citation counts. In long-term research projects, researchers frequently refer to previous findings and results to build upon them. Consequently, citations in these contexts might appear unnatural when, in reality, they are legitimate references to foundational work that supports building new research. This scenario makes it challenging to differentiate between genuinely relevant citations and those that might artificially increase citation counts.

[2] https://www.nfdi4datascience.de/

[3] https://orkg.org/

[4] https://zenodo.org/records/7638511

### 6.2.3 Citation-cartel identification

Researchers often cite publications within their professional network or institution. This tendency can be attributed to the availability heuristic, where individuals rely on information that is readily accessible and familiar. Informal knowledge sharing within these networks also contributes to this pattern. As a result, citations may appear clustered within certain networks, raising questions about their naturalness. However, this clustering may be a natural result of collaborative research and shared interests rather than an indication of citation manipulation.

These limitations highlight the complexity of identifying unnatural citations. Legitimate academic practices and natural dynamics of research communities can overlap with unnatural citation behavior and make the detection of unnatural citations difficult. Addressing these challenges requires the development of more sophisticated methods and datasets to improve the accuracy of our analyses to reduce the chance of a false classification of legitimate academic practices as unnatural citations.

# 7 Conclusion and future work

## 7.1 Conclusion

The scientific profile of any researcher depends on various factors and citation count is one of these. The trend of citation manipulation (i.e., unnatural citations) is increasing to improve one's recognition artificially. In this study, we have presented our solution (named OntoTrack) for identifying unnatural citation patterns in the scientific literature. Building and extending on the groundwork of existing ontologies for citation tracking, we showed how OntoTrack utilizes the reasoning capabilities of ontologies and Linked Open Data (LOD) technologies to identify various forms of citation behavior. Specifically, the OntoTrack Solution integrates key attributes such as academic work, author information, and keywords overlap to model and analyse citation relationships. We presented the detailed architecture (with its various phases) of our OntoTrack Solution. We also described the OntoTrack ontology and the implementation of rules (by using SWRL language) in this ontology. As a key contribution, we described in detail how these rules are used in reasoning and inferencing implicit relationships and classifying citations based on predefined criteria, such as citations with no shared keywords or excessive self-citation and flagging them as unnatural based on the analyzed data. As part of the OntoTrack Solution, we also described OntoTrack ontology, a proof-of-concept dataset, rules implementation, how these entities are loaded into the graph database and finally establishing the SPARQL Endpoint for smart query answering purposes. As part of the use case for performance and evaluation purposes, we defined a comprehensive list of Competency Questions (CQs) and executed them over the SPARQL Endpoint. We explained the structured approach to detect citation manipulation and showed the results as a classification between natural and unnatural citations. We also showed the results by implementing the use case scenario, which shows that OntoTrack uses a data model to establish relationships between entities, such as thematic overlap or author collaboration and applies logical rules to detect patterns that indicate unnatural citations.

## 7.2 Future work

While SWRL provides a powerful mechanism for extending OWL ontologies with rule-based reasoning, its limitations in handling negation, aggregation, and closed-world assumptions restrict the complexity of analyses that can be performed. In future work, these gaps could be addressed through complementary reasoning techniques. For instance, SPARQL CONSTRUCT queries could be used to dynamically generate inferred triples based on complex logical patterns or value comparisons that SWRL cannot express. Similarly, the Shapes Constraint Language (SHACL) enables constraint validation and rule-based inference, supporting conditional logic and cardinality checks that extend beyond OWL's native expressivity. Integrating these tools alongside SWRL would create a more flexible and scalable reasoning environment, capable of detecting more nuanced forms of citation manipulation and supporting richer analytical use cases in ontology-based citation analysis. Furthermore, as part of the NFDI project, we plan to apply the OntoTrack ontology to a large-scale dataset comprising several million records. Future work will focus on systematic verification strategies, including random validation, sample-based dataset verification, and rule-based consistency checks to ensure the reliability of inferred results at scale. In addition, we aim to explore the integration of the ontology-based reasoning framework with machine learning and large language model (LLM) approaches to enhance the system's ability to detect complex citation behaviors and uncover deeper patterns within scholarly communication networks.

# Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/MarkEMD95/OntoTrack.

# Author contributions

MD: Conceptualization, Methodology, Writing – original draft, Data curation, Software, Visualization. MAA: Conceptualization, Methodology, Writing – review & editing, Supervision. RF: Conceptualization, Investigation, Software, Writing – review & editing. SS: Conceptualization, Writing – review & editing, Funding acquisition.

# Funding

The author(s) declared that financial support was received for this work and/or its publication. This work received funding from the Federal Ministry of Research, Technology and Space (BMFTR) and the State of Berlin under grant number 16DII141 (Weizenbaum Institute) and by the German Research Foundation (DFG) under project number 460234259 (NFDI4DataScience).

## Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declared that generative AI was used in the creation of this manuscript. Generative AI tools were used to improve language and grammar and create LaTeX table structures.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Ashikuzzaman, M. (2018). *What Is Citation?* LIS Edu Network. Available online at: https://www.lisedunetwork.com/citation-analysis (Accessed January 14, 2026).

Aslam, M., and Aljohani, N. (2017). Spedia: a central hub for the linked open data of scientific publications. *Int. J. Semant. Web Inf. Syst.* 13, 128–146. doi: 10.4018/IJSWIS.2017010108

Aslam, M. A., and Aljohani, N. R. (2016). "Spedia: a semantics based repository of scientific publications data," in *Web-Age Information Management*, eds. B. Cui, N. Zhang, J. Xu, X. Lian, and D. Liu (Cham: Springer International Publishing), 479–490. doi: 10.1007/978-3-319-39937-9_37

Avros, R., Keshet, S., Kitai, D. T., Vexler, E., and Volkovich, Z. (2023). Detecting pseudo-manipulated citations in scientific literature through perturbations of the citation graph. *Mathematics* 11:3820. doi: 10.3390/math11183820

Chandrasekaran, B., Josephson, J. R., and Benjamins, V. R. (1999). What are ontologies, and why do we need them? *IEEE Intell. Syst. Applic.* 14, 20–26. doi: 10.1109/5254.747902

Connected Papers (2024). *Connected Papers*. Available online at: https://www.connectedpapers.com/ (Accessed January 14, 2026).

COPE Council (2019). *COPE Discussion Document: Citation Manipulation*. Committee on Publication Ethics. doi: 10.24318/cope.2019.3.1

DeBellis, M. (2020). *Drools vs. Pellet for SWRL Rules*. Available online at: https://www.michaeldebellis.com/post/drools-vs-pellet-for-swrl-rules (Accessed January 14, 2026).

Färber, M., Lamprecht, D., Krause, J., Aung, L., and Haase, P. (2023). "Semopenalex: the scientific landscape in 26 billion RDF triples," in *International Semantic Web Conference* (Cham: Springer Nature Switzerland), 94–112. doi: 10.1007/978-3-031-47243-5_6

Garfield, E. (1955). Citation indexes for science: a new dimension in documentation through association of ideas. *Science* 122, 108–111. doi: 10.1126/science.122.3159.108

Kojaku, S., Livan, G., and Masuda, N. (2021). Detecting anomalous citation groups in journal networks. *Sci. Rep.* 11:14524. doi: 10.1038/s41598-021-93572-3

Liu, J., Xia, F., Feng, X., Ren, J., and Liu, H. (2022). Deep graph learning for anomalous citation detection. *IEEE Trans. Neural Netw. Learn. Syst.* 33, 2543–2557. doi: 10.1109/TNNLS.2022.3145092

Nicolaisen, J. (2007). Citation analysis. *Ann. Rev. Inf. Sci. Technol.* 41, 609–641. doi: 10.1002/aris.2007.1440410120

O'Connor, M. J., and Das, A. (2006). "The swrltab: an extensible environment for working with swrl rules in protégé-owl," in *Proceedings of the 2nd International Conference Rules Rule Markup Language Semantic Web*, 1–2.

Peroni, S. (2024). *OpenCitations Ontology*. Available online at: https://opencitations.github.io/ontology/current/ontology.html (Accessed January 14, 2026).

Rossi, R., and Ahmed, N. (2015). "The network data repository with interactive graph analytics and visualization," in *Proceedings of the AAAI Conference on Artificial Intelligence* (New York, NY, USA), 4292–4293. doi: 10.1609/aaai.v29i1.9277

Small, H. (1999). Visualizing science by citation mapping. *J. Am. Soc. Inf. Sci.* 50, 799–813. doi: 10.1002/(SICI)1097-4571(1999)50:9<799::AID-ASI9>3.0.CO;2-G

Smith, L. C. (1981). *Citation Analysis*. Graduate School of Library and Information Science; University of Illinois.

SPAR Ontologies (2018). *Citation Typing Ontology (CiTO)*. Available online at: https://sparontologies.github.io/cito/current/cito.html (Accessed January 14, 2026).

SPAR Ontologies (2019). *Functional Annotation of Biomedical Ontologies (FABIO)*. Available online at: https://sparontologies.github.io/fabio/current/fabio.html (Accessed January 14, 2026).

SPAR Ontologies (2024). *Citation Counting and Context Characterization Ontology (C40)*. Available online at: https://sparontologies.github.io/c4o/current/c4o.html (Accessed January 14, 2026).

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., and Su, Z. (2008). "Arnetminer: extraction and mining of academic social networks," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA: Association for Computing Machinery), 990–998. doi: 10.1145/1401890.1402008

Uschold, M., and Gruninger, M. (1996). Ontologies: principles, methods and applications. *Knowl. Eng. Rev.* 11, 93–136. doi: 10.1017/S0269888900007797

Vickery, B. C. (1997). Ontologies. *J. Inf. Sci.* 23, 277–286. doi: 10.1177/016555159702300402

Wren, J. D., and Georgescu, C. (2020). Detecting potential reference list manipulation within a citation network. *bioRxiv*. doi: 10.1101/2020.08.12.248369

Zhao, D., and Strotmann, A. (2015). *Analysis and Visualization of Citation Networks*. San Rafael, CA: Morgan and Claypool Publishers. doi: 10.1007/978-3-031-02291-3