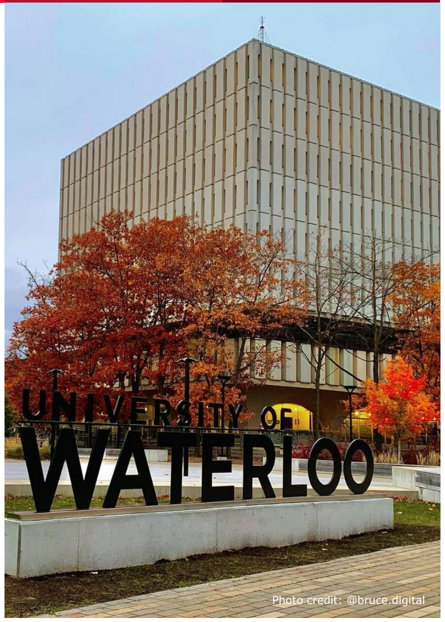
CS846 Project Infrastructure Overview

2025.01.30

Xiang Chen,
David R. Cheriton School of Computer Science

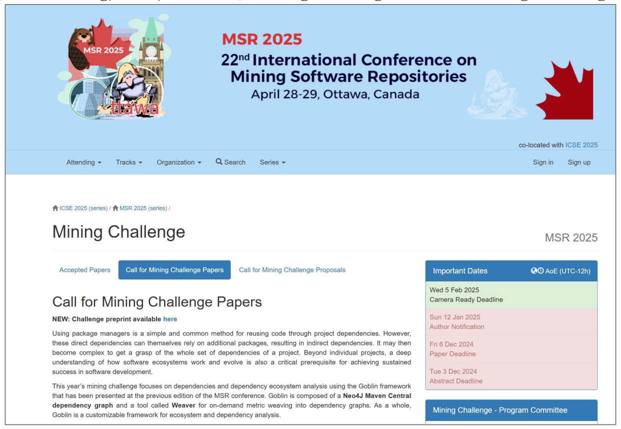
Instructor: Mike Godfrey

UNIVERSITY OF WATERLOO



Taken From MSR 2025 Mining Challenge

https://2025.msrconf.org/track/msr-2025-mining-challenge#Call-for-Mining-Challenge-Papers



MSR 2025
22nd International Conference on Mining Software Repositories
April 28-29, Ottawa, Canada

co-located with ICSE 2025

ICSE 2025 (series) / MSR 2025 (series) /

Mining Challenge

MSR 2025

Call for Mining Challenge Papers

NEW: Challenge preprint available here

Using package managers is a simple and common method for reusing code through project dependencies. However, these direct dependencies can themselves rely on additional packages, resulting in indirect dependencies. It may then become complex to get a grasp of the whole set of dependencies of a project. Beyond individual projects, a deep understanding of how software ecosystems work and evolve is also a critical prerequisite for achieving sustained success in software development.

This year's mining challenge focuses on dependencies and dependency ecosystem analysis using the Goblin framework that has been presented at the previous edition of the MSR conference. Goblin is composed of a **Neo4J Maven Central dependency graph** and a tool called **Weaver** for on-demand metric weaving into dependency graphs. As a whole, Goblin is a customizable framework for ecosystem and dependency analysis.

Important Dates	AoE (UTC-12h)
Wed 5 Feb 2025	Camera Ready Deadline
Sun 12 Jan 2025	Author Notification
Fri 6 Dec 2024	Paper Deadline
Tue 3 Dec 2024	Abstract Deadline

Mining Challenge - Program Committee

What is the MSR Mining Challenge?

- Annual challenge by the Mining Software Repositories (MSR) conference.

What is the MSR Mining Challenge?

- Annual challenge by the Mining Software Repositories (MSR) conference.
- Focuses on analyzing real-world software data.

What is the MSR Mining Challenge?

- Annual challenge by the Mining Software Repositories (MSR) conference.
- Focuses on analyzing real-world software data.
- Provides specific datasets and tools for researchers.

What is the MSR Mining Challenge?

- Annual challenge by the Mining Software Repositories (MSR) conference.
- Focuses on analyzing real-world software data.
- Provides specific datasets and tools for researchers.
- Encourages innovative methods and insights.

What is the MSR Mining Challenge?

- Annual challenge by the Mining Software Repositories (MSR) conference.
- Focuses on analyzing real-world software data.
- Provides specific datasets and tools for researchers.
- Encourages innovative methods and insights.
- Aims to advance software repository mining techniques.

Theme For MSR 2025 Mining Challenge

- **Theme**: Dependency Analysis with **Goblin** framework

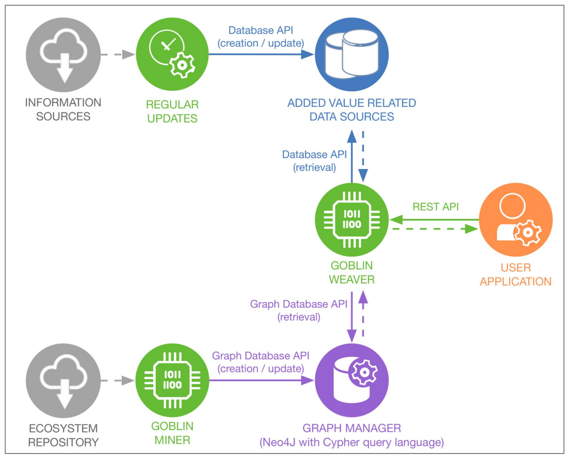
Theme For MSR 2025 Mining Challenge

- **Theme**: Dependency Analysis with **Goblin** framework
- **Goblin Framework Components**
  - Neo4J Maven Central dependency graph
  - **Miner**: generates the dependency graph
  - **Weaver**: Adds custom metrics to dependency graphs, and query the graph

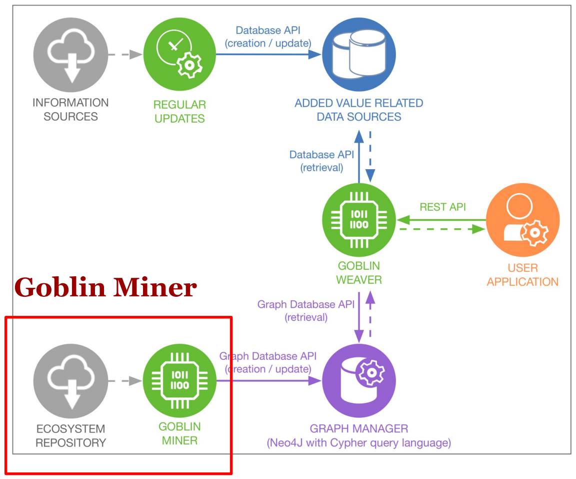
Theme For MSR 2025 Mining Challenge

- **Theme**: Dependency Analysis with **Goblin** framework
- **Goblin Framework Components**
  - Neo4J Maven Central dependency graph
  - **Miner**: generates the dependency graph
  - **Weaver**: Adds custom metrics to dependency graphs, and query the graph
- **Goblin**: A customizable tool for studying software ecosystems and dependencies

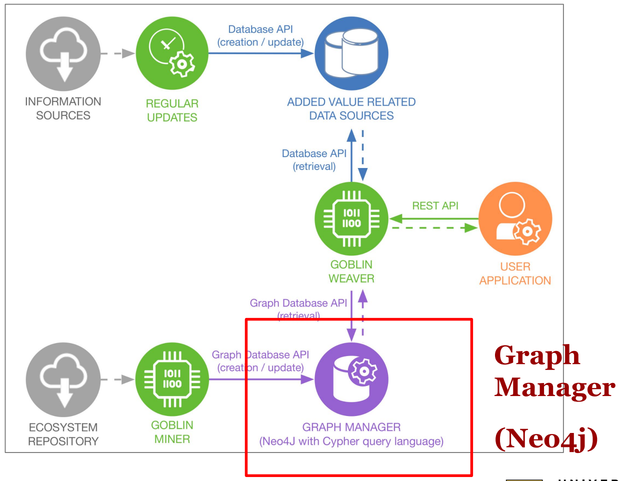
Goblin Framework


INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

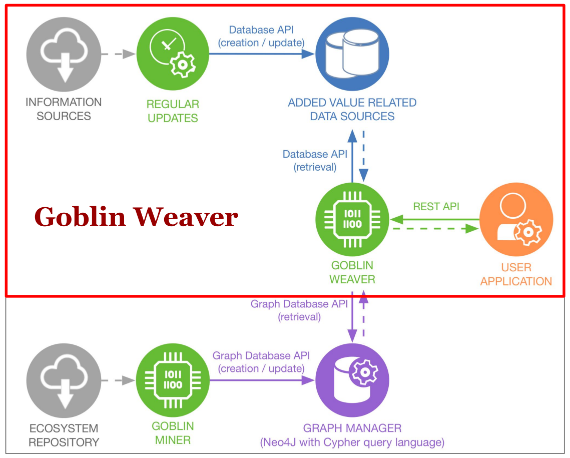
Goblin Framework


INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Goblin Miner
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

Goblin Framework


INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
REST API
GOBLIN WEAVER
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)
Graph Manager
(Neo4j)

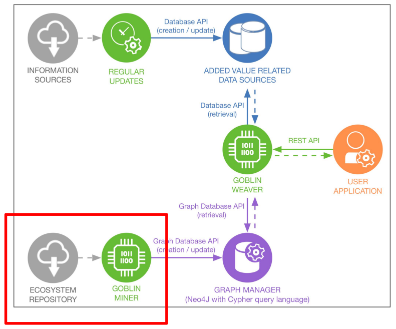
Goblin Framework


INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
Goblin Weaver
REST API
GOBLIN WEAVER
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

Goblin Framework

Goblin Miner

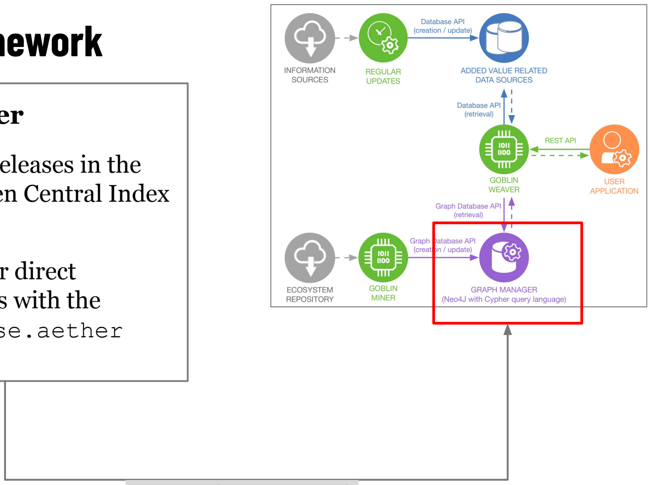
- Retrieve all releases in the Lucene Maven Central Index archive
- Retrieve their direct dependencies with the `org.eclipse.aether` library



Goblin Framework

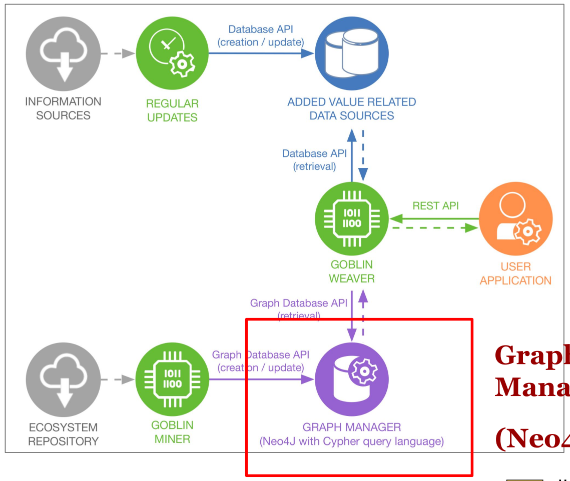
Goblin Miner

- Retrieve all releases in the Lucene Maven Central Index archive
- Retrieve their direct dependencies with the `org.eclipse.aether` library



Dependency graph

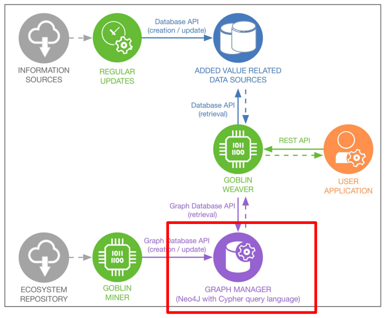
Goblin Framework


INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

Graph Manager

(Neo4j)

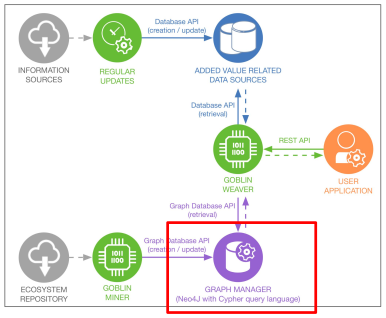
Goblin Framework


INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

Graph Manager (Neo4j)

- Node types
  - Libraries (type `Artifact`)
  - Releases (type `Release`)

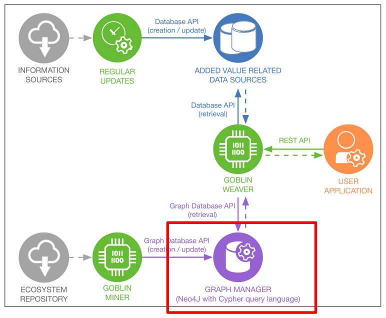
Goblin Framework


INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

Graph Manager (Neo4j)

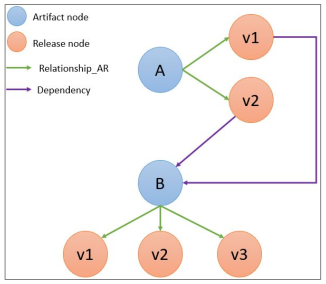
- Node types
  - Libraries (type `Artifact`)
  - Releases (type `Release`)
- Edge types
  - Dependencies (type `dependency`) are from `Release` to `Artifact`
  - Versioning (type `relationship_AR`) are from `Artifact` to `Release`

Goblin Framework


INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

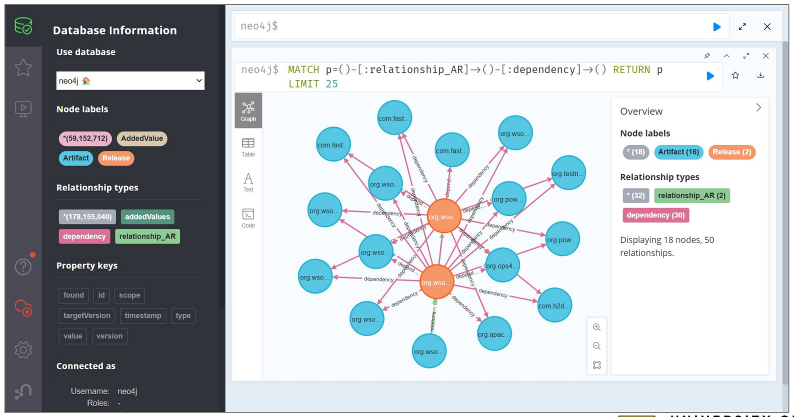
Graph Manager (Neo4j)

- Node types
  - Libraries (type `Artifact`)
  - Releases (type `Release`)
- Edge types
  - Dependencies (type `dependency`) are from `Release` to `Artifact`
  - Versioning (type `relationship_AR`) are from `Artifact` to `Release`



Goblin Framework

Graph Manager (Neo4j)



Database Information

Use database

neo4j

Node labels

*(59,152,712) AddedValue Artifact Release

Relationship types

*(178,155,040) addedValues dependency relationship_AR

Property keys

found id scope targetVersion timestamp type value version

Connected as

Username: neo4j
Roles: -

```
neo4j$
```

```
neo4j$ MATCH p=()-[:relationship_AR]→()-[:dependency]→() RETURN p
       LIMIT 25
```

Overview

Node labels

* (18) Artifact (16) Release (2)

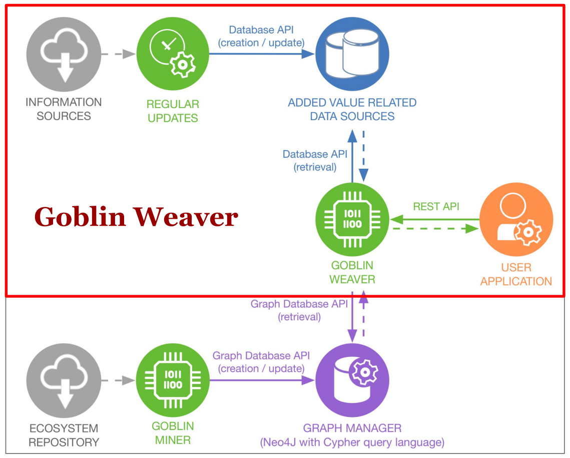
Relationship types

* (32) relationship_AR (2) dependency (30)

Displaying 18 nodes, 50 relationships.

UNIVERSITY OF WATERLOO

Goblin Framework



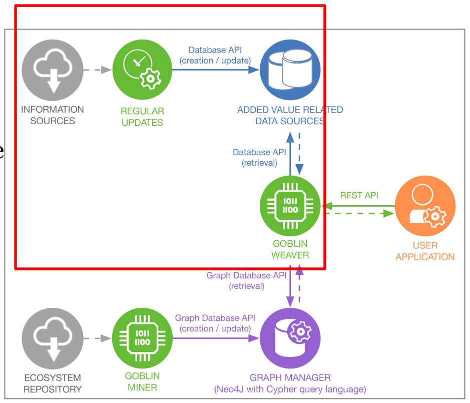
INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
Goblin Weaver
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

UNIVERSITY OF WATERLOO

Goblin Framework

Goblin Weaver

- On-demand enrichment of the dependency graph



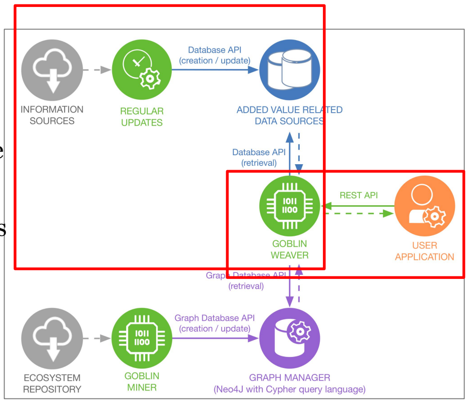
INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

UNIVERSITY OF WATERLOO

Goblin Framework

Goblin Weaver

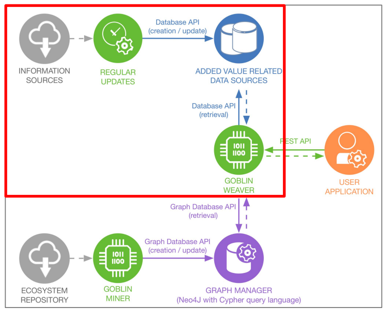
- On-demand enrichment of the dependency graph
- An alternative for direct access to the Neo4j database



INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

UNIVERSITY OF WATERLOO

Goblin Framework

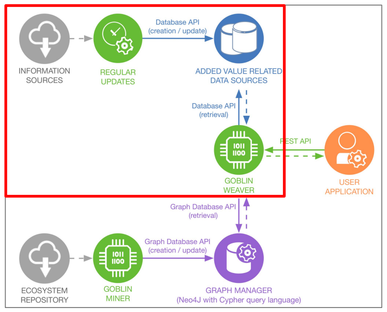

INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

Goblin Weaver

- On-demand enrichment of the dependency graph
- An alternative for direct access to the Neo4j database

UNIVERSITY OF WATERLOO

Goblin Framework


INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

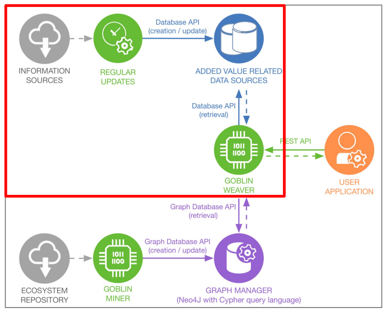
Goblin Weaver

- On-demand enrichment of the dependency graph

❖ <u>Release nodes added values</u>

UNIVERSITY OF WATERLOO

Goblin Framework


INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

Goblin Weaver

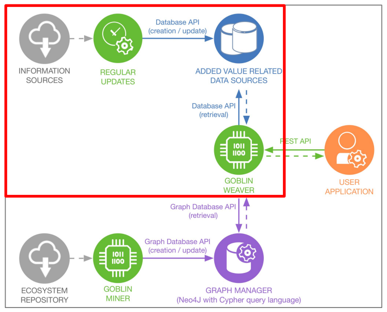
- On-demand enrichment of the dependency graph

❖ <u>Release nodes added values</u>

a. CVE (Common Vulnerabilities and Exposures): use the osv.dev dataset
   - Name, cwe (type of vulnerability) and severity (low, moderate, high, critical)

UNIVERSITY OF WATERLOO

Goblin Framework


INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

Goblin Weaver

- On-demand enrichment of the dependency graph

❖ <u>Release nodes added values</u>

a. CVE (Common Vulnerabilities and Exposures): use the osv.dev dataset
   - Name, cwe (type of vulnerability) and severity (low, moderate, high, critical)

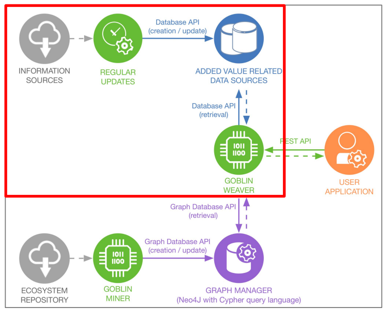
b. FRESHNESS
   - The number of more recent releases available
   - The time elapsed between it and the most recent release

UNIVERSITY OF WATERLOO

Goblin Framework

Goblin Weaver

- On-demand enrichment of the dependency graph


INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

❖ <u>Release nodes added values</u>

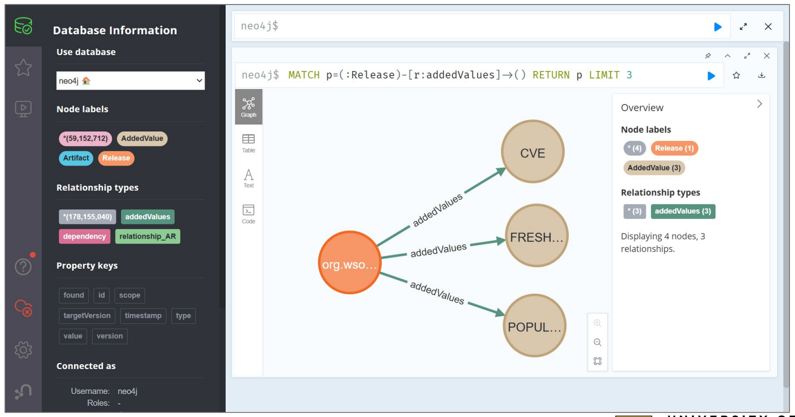
a. CVE (Common Vulnerabilities and Exposures): use the [osv.dev](osv.dev) dataset
   - Name, cwe (type of vulnerability) and severity (low, moderate, high, critical)

b. FRESHNESS
   - The number of more recent releases available
   - The time elapsed between it and the most recent release

c. POPULARITY_1_YEAR
   - Number of dependants over a one year window

Goblin Framework

Goblin Weaver

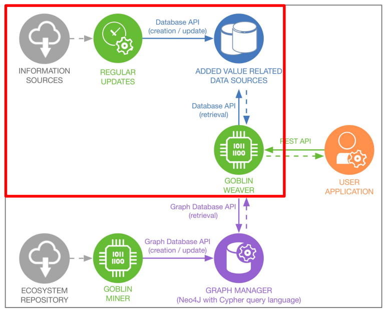


```
neo4j$ MATCH p=(:Release)-[r:addedValues]→() RETURN p LIMIT 3
```

Goblin Framework

Goblin Weaver

- On-demand enrichment of the dependency graph

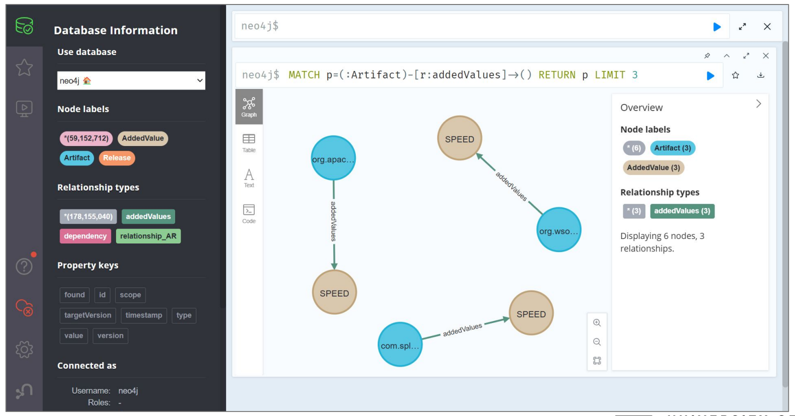

INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

❖ <u>Artifact nodes added values</u>

➢ SPEED
   - Average number of releases per day of a library

Goblin Framework

Goblin Weaver

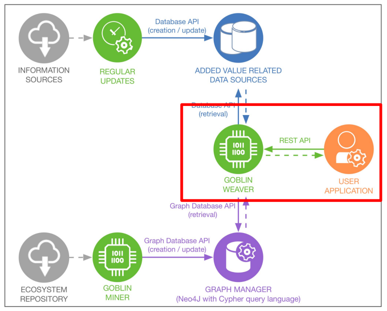


```
neo4j$ MATCH p=(:Artifact)-[r:addedValues]→() RETURN p LIMIT 3
```

Goblin Framework

Goblin Weaver

- An alternative for direct access to the Neo4j database



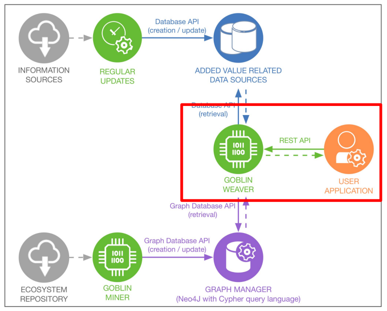
INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
REST API
GOBLIN WEAVER
USER APPLICATION
Graph Database API (retrieval)
Graph Database API (creation / update)
ECOSYSTEM REPOSITORY
GOBLIN MINER
GRAPH MANAGER (Neo4J with Cypher query language)

UNIVERSITY OF WATERLOO

Goblin Framework

Goblin Weaver

- An alternative for direct access to the Neo4j database



INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
REST API
GOBLIN WEAVER
USER APPLICATION
Graph Database API (retrieval)
Graph Database API (creation / update)
ECOSYSTEM REPOSITORY
GOBLIN MINER
GRAPH MANAGER (Neo4J with Cypher query language)

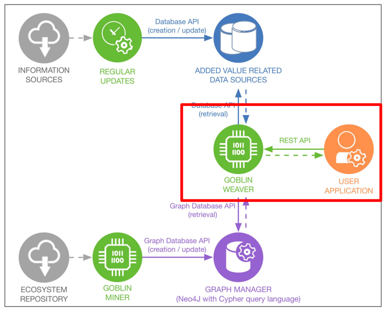
Cypher: Neo4j’s built-in query language

UNIVERSITY OF WATERLOO

Goblin Framework

Goblin Weaver

- An alternative for direct access to the Neo4j database



INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
REST API
GOBLIN WEAVER
USER APPLICATION
Graph Database API (retrieval)
Graph Database API (creation / update)
ECOSYSTEM REPOSITORY
GOBLIN MINER
GRAPH MANAGER (Neo4J with Cypher query language)

Cypher: Neo4j’s built-in query language

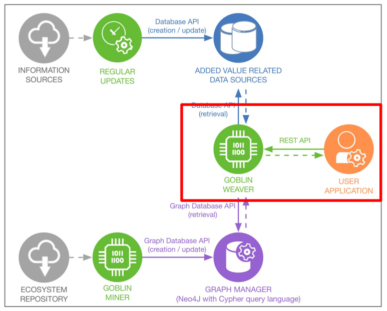
https://neo4j.com/docs/cypher-manual/current/queries/basic/

UNIVERSITY OF WATERLOO

Goblin Framework

Goblin Weaver

- An alternative for direct access to the Neo4j database



INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
REST API
GOBLIN WEAVER
USER APPLICATION
Graph Database API (retrieval)
Graph Database API (creation / update)
ECOSYSTEM REPOSITORY
GOBLIN MINER
GRAPH MANAGER (Neo4J with Cypher query language)

Cypher: Neo4j’s built-in query language

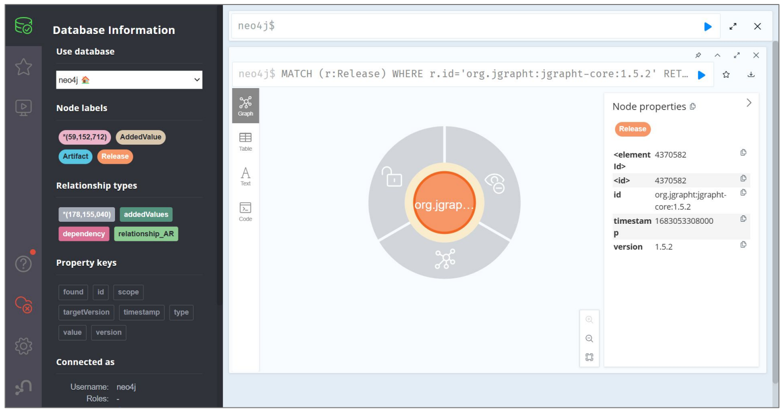
https://neo4j.com/docs/cypher-manual/current/queries/basic/

```
MATCH (r:Release) WHERE r.id='org.jgrapht:jgrapht-core:1.5.2' RETURN r
```

UNIVERSITY OF WATERLOO

Goblin Framework

Cypher: Neo4j's built-in query language



```
neo4j$ MATCH (r:Release) WHERE r.id='org.jgrapht:jgrapht-core:1.5.2' RET…
```

Database Information

Use database: neo4j

Node labels: *(59,152,712) AddedValue Artifact Release

Relationship types: *(178,155,040) addedValues dependency relationship_AR

Property keys: found id scope targetVersion timestamp type value version

Connected as — Username: neo4j Roles: -

Node properties — Release

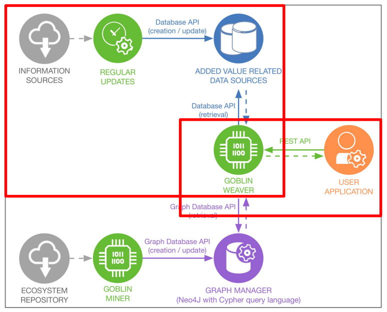
<element Id>	4370582
<id>	4370582
id	org.jgrapht:jgrapht-core:1.5.2
timestamp	1683053308000
version	1.5.2

UNIVERSITY OF WATERLOO

Goblin Framework

Goblin Weaver

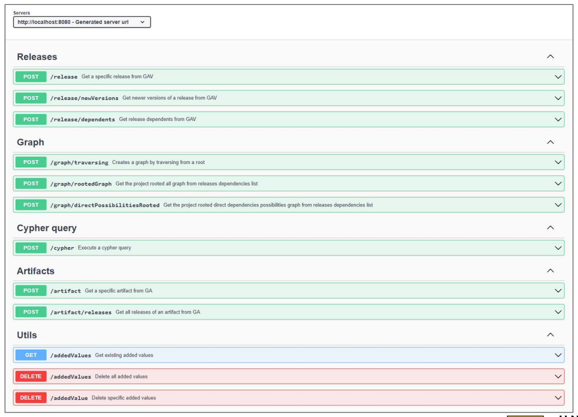
- On-demand enrichment of the dependency graph
- An alternative for direct access to the Neo4j database


INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER
(Neo4J with Cypher query language)

UNIVERSITY OF WATERLOO

Goblin Framework

Goblin Weaver



Servers: http://localhost:8080 - Generated server url

Releases
- POST `/release` Get a specific release from GAV
- POST `/release/newVersions` Get newer versions of a release from GAV
- POST `/release/dependents` Get release dependents from GAV

Graph
- POST `/graph/traversing` Creates a graph by traversing from a root
- POST `/graph/rootedGraph` Get the project rooted all graph from releases dependencies list
- POST `/graph/directPossibilitiesRooted` Get the project rooted direct dependencies possibilities graph from releases dependencies list

Cypher query
- POST `/cypher` Execute a cypher query

Artifacts
- POST `/artifact` Get a specific artifact from GA
- POST `/artifact/releases` Get all releases of an artifact from GA

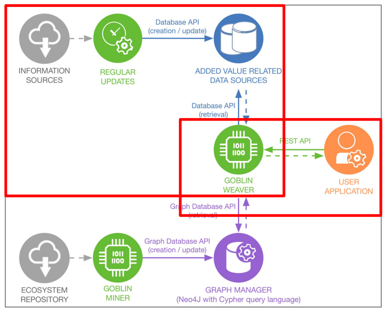
Utils
- GET `/addedValues` Get existing added values
- DELETE `/addedValues` Delete all added values
- DELETE `/addedValue` Delete specific added values

UNIVERSITY OF WATERLOO

Goblin Framework

Goblin Weaver Example

- Which are the latest versions available after `jgrapht-core 1.5.0`?
- Add their CVE, freshness and popularity

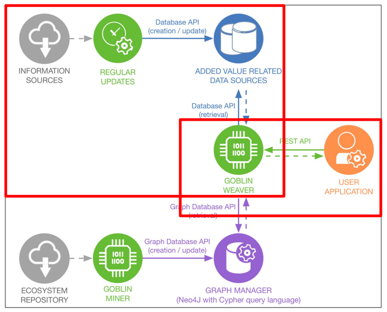

INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER
(Neo4J with Cypher query language)

UNIVERSITY OF WATERLOO

Goblin Framework

Goblin Weaver Example

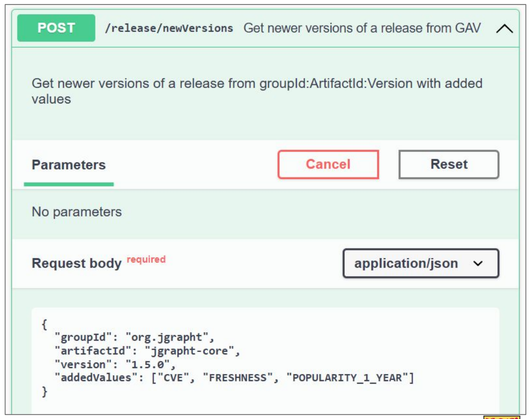
- Which are the latest versions available after `jgrapht-core 1.5.0`?
- Add their CVE, freshness and popularity


INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

```
Method: POST
ROUTE: /release/newVersions
Body:
{
  "groupId": "org.jgrapht",
  "artifactId": "jgrapht-core",
  "version": "1.5.0",
  "addedValues": ["CVE", "FRESHNESS", "POPULARITY_1_YEAR"]
}
```

Goblin Framework

Goblin Weaver Example



`/release/newVersions` Get newer versions of a release from GAV

Get newer versions of a release from groupId:ArtifactId:Version with added values

Parameters

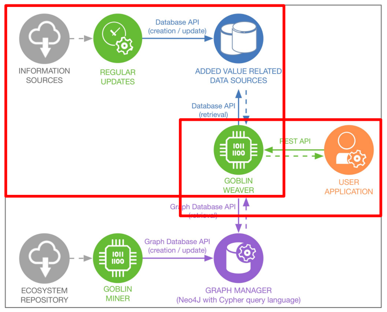
No parameters

Request body required

```
{
  "groupId": "org.jgrapht",
  "artifactId": "jgrapht-core",
  "version": "1.5.0",
  "addedValues": ["CVE", "FRESHNESS", "POPULARITY_1_YEAR"]
}
```

Goblin Framework

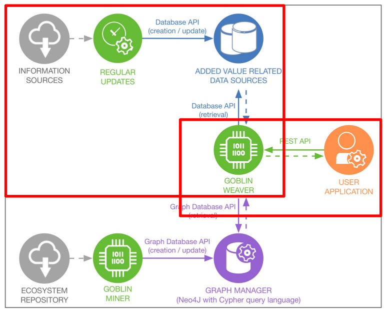
Goblin Weaver Example


INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

```
{
  "nodes": [
    {
      "cve": [],
      "id": "org.jgrapht:jgrapht-core:1.5.1",
      "nodeType": "RELEASE",
      "freshness": {
        "numberMissedRelease": "1",
        "outdatedTimeInMs": "66882028000"
      },
      "version": "1.5.1",
      "popularity_1_year": 105,
      "timestamp": 1616171280000
    },
    {
      "cve": [],
      "id": "org.jgrapht:jgrapht-core:1.5.2",
      "nodeType": "RELEASE",
      "freshness": {
        "numberMissedRelease": "0",
        "outdatedTimeInMs": "0"
      },
      "version": "1.5.2",
      "popularity_1_year": 953,
      "timestamp": 1683053308000
    }
  ]
}
```

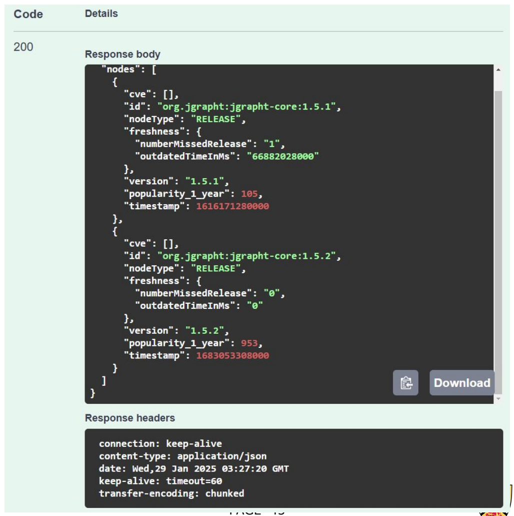
Goblin Framework

Goblin Weaver Example


INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

```
{
  "nodes": [
    {
      "cve": [],
      "id": "org.jgrapht:jgrapht-core:1.5.1",
      "nodeType": "RELEASE",
      "freshness": {
        "numberMissedRelease": "1",
        "outdatedTimeInMs": "66882028000"
      },
      "version": "1.5.1",
      "popularity_1_year": 105,
      "timestamp": 1616171280000
    },
    {
      "cve": [],
      "id": "org.jgrapht:jgrapht-core:1.5.2",
      "nodeType": "RELEASE",
      "freshness": {
        "numberMissedRelease": "0",
        "outdatedTimeInMs": "0"
      },
      "version": "1.5.2",
      "popularity_1_year": 953,
      "timestamp": 1683053308000
    }
  ]
}
```

Goblin Framework



Code | Details

200

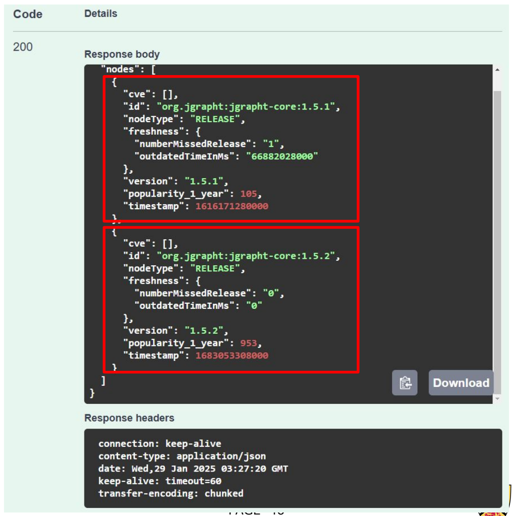
Response body

```
  "nodes": [
    {
      "cve": [],
      "id": "org.jgrapht:jgrapht-core:1.5.1",
      "nodeType": "RELEASE",
      "freshness": {
        "numberMissedRelease": "1",
        "outdatedTimeInMs": "66882028000"
      },
      "version": "1.5.1",
      "popularity_1_year": 105,
      "timestamp": 1616171280000
    },
    {
      "cve": [],
      "id": "org.jgrapht:jgrapht-core:1.5.2",
      "nodeType": "RELEASE",
      "freshness": {
        "numberMissedRelease": "0",
        "outdatedTimeInMs": "0"
      },
      "version": "1.5.2",
      "popularity_1_year": 953,
      "timestamp": 1683053308000
    }
  ]
}
```

Response headers

```
connection: keep-alive
content-type: application/json
date: Wed,29 Jan 2025 03:27:20 GMT
keep-alive: timeout=60
transfer-encoding: chunked
```

Goblin Framework



Code | Details

200

Response body

```
  "nodes": [
    {
      "cve": [],
      "id": "org.jgrapht:jgrapht-core:1.5.1",
      "nodeType": "RELEASE",
      "freshness": {
        "numberMissedRelease": "1",
        "outdatedTimeInMs": "66882028000"
      },
      "version": "1.5.1",
      "popularity_1_year": 105,
      "timestamp": 1616171280000
    },
    {
      "cve": [],
      "id": "org.jgrapht:jgrapht-core:1.5.2",
      "nodeType": "RELEASE",
      "freshness": {
        "numberMissedRelease": "0",
        "outdatedTimeInMs": "0"
      },
      "version": "1.5.2",
      "popularity_1_year": 953,
      "timestamp": 1683053308000
    }
  ]
}
```

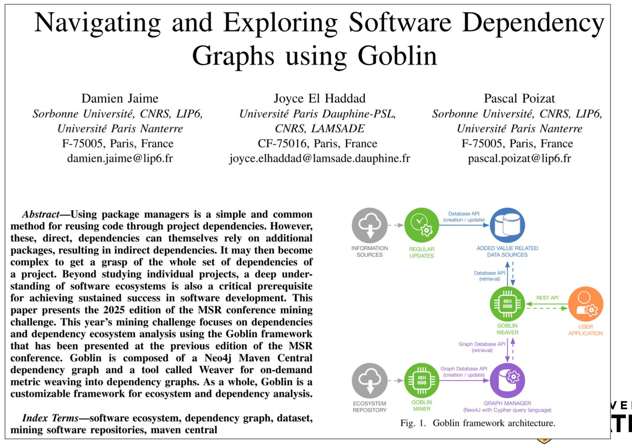
Download

Response headers

```
connection: keep-alive
content-type: application/json
date: Wed,29 Jan 2025 03:27:20 GMT
keep-alive: timeout=60
transfer-encoding: chunked
```

Some Useful Links

1. **Challenge preprint**: https://hal.science/hal-04777703



Navigating and Exploring Software Dependency Graphs using Goblin

Damien Jaime
Sorbonne Université, CNRS, LIP6,
Université Paris Nanterre
F-75005, Paris, France
damien.jaime@lip6.fr

Joyce El Haddad
Université Paris Dauphine-PSL,
CNRS, LAMSADE
CF-75016, Paris, France
joyce.elhaddad@lamsade.dauphine.fr

Pascal Poizat
Sorbonne Université, CNRS, LIP6,
Université Paris Nanterre
F-75005, Paris, France
pascal.poizat@lip6.fr

***Abstract*—Using package managers is a simple and common method for reusing code through project dependencies. However, these, direct, dependencies can themselves rely on additional packages, resulting in indirect dependencies. It may then become complex to get a grasp of the whole set of dependencies of a project. Beyond studying individual projects, a deep understanding of software ecosystems is also a critical prerequisite for achieving sustained success in software development. This paper presents the 2025 edition of the MSR conference mining challenge. This year's mining challenge focuses on dependencies and dependency ecosystem analysis using the Goblin framework that has been presented at the previous edition of the MSR conference. Goblin is composed of a Neo4j Maven Central dependency graph and a tool called Weaver for on-demand metric weaving into dependency graphs. As a whole, Goblin is a customizable framework for ecosystem and dependency analysis.**

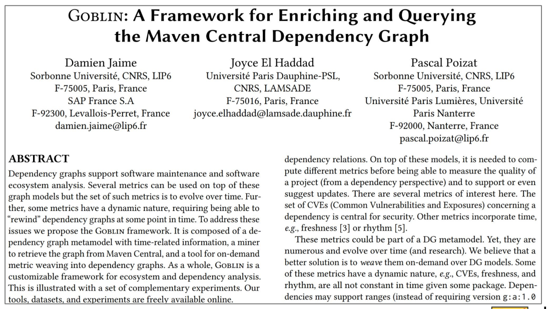
***Index Terms*—software ecosystem, dependency graph, dataset, mining software repositories, maven central**

INFORMATION SOURCES
REGULAR UPDATES
Database API (creation / update)
ADDED VALUE RELATED DATA SOURCES
Database API (retrieval)
GOBLIN WEAVER
REST API
USER APPLICATION
Graph Database API (retrieval)
ECOSYSTEM REPOSITORY
GOBLIN MINER
Graph Database API (creation / update)
GRAPH MANAGER (Neo4J with Cypher query language)

Fig. 1. Goblin framework architecture.

Some Useful Links

2. **Goblin framework**: https://dl.acm.org/doi/abs/10.1145/3643991.3644879?casa_token=bTRwoEEBLooAAAAA:9v4cP65Fwic4JmVCEg0mGGA4XYSlbg6vaMJ-TAbbw84kXgJrt3sFTGrm1UTRGCtSTbFLV7ySOjKy



Goblin: A Framework for Enriching and Querying the Maven Central Dependency Graph

Damien Jaime
Sorbonne Université, CNRS, LIP6
F-75005, Paris, France
SAP France S.A
F-92300, Levallois-Perret, France
damien.jaime@lip6.fr

Joyce El Haddad
Université Paris Dauphine-PSL,
CNRS, LAMSADE
F-75016, Paris, France
joyce.elhaddad@lamsade.dauphine.fr

Pascal Poizat
Sorbonne Université, CNRS, LIP6
F-75005, Paris, France
Université Paris Lumières, Université
Paris Nanterre
F-92000, Nanterre, France
pascal.poizat@lip6.fr

ABSTRACT

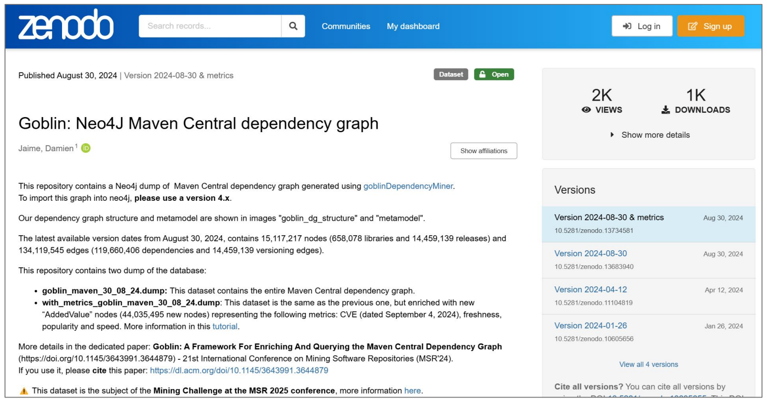
Dependency graphs support software maintenance and software ecosystem analysis. Several metrics can be used on top of these graph models but the set of such metrics is to evolve over time. Further, some metrics have a dynamic nature, requiring being able to "rewind" dependency graphs at some point in time. To address these issues we propose the Goblin framework. It is composed of a dependency graph metamodel with time-related information, a miner to retrieve the graph from Maven Central, and a tool for on-demand metric weaving into dependency graphs. As a whole, Goblin is a customizable framework for ecosystem and dependency analysis. This is illustrated with a set of complementary experiments. Our tools, datasets, and experiments are freely available online.

dependency relations. On top of these models, it is needed to compute different metrics before being able to measure the quality of a project (from a dependency perspective) and to support or even suggest updates. There are several metrics of interest here. The set of CVEs (Common Vulnerabilities and Exposures) concerning a dependency is central for security. Other metrics incorporate time, *e.g.*, freshness [3] or rhythm [5].

These metrics could be part of a DG metamodel. Yet, they are numerous and evolve over time (and research). We believe that a better solution is to *weave* them on-demand over DG models. Some of these metrics have a dynamic nature, *e.g.*, CVEs, freshness, and rhythm, are all not constant in time given some package. Dependencies may support ranges (instead of requiring version `g:a:1.0`

Some Useful Links

3. **Maven Central Neo4j dependency graph datasets:** https://zenodo.org/records/13734581



zenodo | Search records... | Communities | My dashboard | Log in | Sign up

Published August 30, 2024 | Version 2024-08-30 & metrics — Dataset — Open

Goblin: Neo4J Maven Central dependency graph

Jaime, Damien[1]

Show affiliations

This repository contains a Neo4j dump of Maven Central dependency graph generated using goblinDependencyMiner.
To import this graph into neo4j, **please use a version 4.x**.

Our dependency graph structure and metamodel are shown in images "goblin_dg_structure" and "metamodel".

The latest available version dates from August 30, 2024, contains 15,117,217 nodes (658,078 libraries and 14,459,139 releases) and 134,119,545 edges (119,660,406 dependencies and 14,459,139 versioning edges).

This repository contains two dump of the database:

- **goblin_maven_30_08_24.dump:** This dataset contains the entire Maven Central dependency graph.
- **with_metrics_goblin_maven_30_08_24.dump**: This dataset is the same as the previous one, but enriched with new "AddedValue" nodes (44,035,495 new nodes) representing the following metrics: CVE (dated September 4, 2024), freshness, popularity and speed. More information in this tutorial.

More details in the dedicated paper: **Goblin: A Framework For Enriching And Querying the Maven Central Dependency Graph** (https://doi.org/10.1145/3643991.3644879) - 21st International Conference on Mining Software Repositories (MSR'24).
If you use it, please **cite** this paper: https://dl.acm.org/doi/10.1145/3643991.3644879

⚠ This dataset is the subject of the **Mining Challenge at the MSR 2025 conference**, more information here.

2K VIEWS — 1K DOWNLOADS

Show more details

Versions

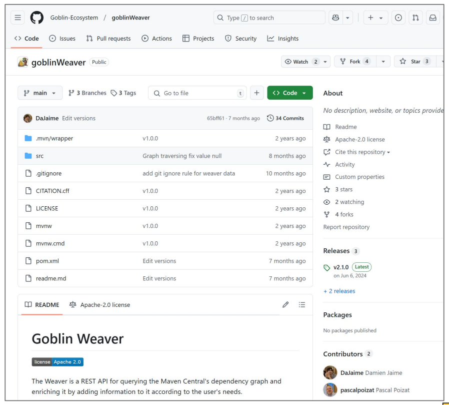
Version 2024-08-30 & metrics 10.5281/zenodo.13734581	Aug 30, 2024
Version 2024-08-30 10.5281/zenodo.13683940	Aug 30, 2024
Version 2024-04-12 10.5281/zenodo.11104819	Apr 12, 2024
Version 2024-01-26 10.5281/zenodo.10605656	Jan 26, 2024

View all 4 versions

Cite all versions? You can cite all versions by

Some Useful Links

4. **Goblin Weaver:** https://github.com/Goblin-Ecosystem/goblinWeaver



Goblin-Ecosystem / goblinWeaver

Code | Issues | Pull requests | Actions | Projects | Security | Insights

goblinWeaver Public — Watch 2 — Fork 4 — Star 3

main — 3 Branches — 3 Tags — Go to file — Code

DaJaime Edit versions — 65bff61 · 7 months ago — 34 Commits

.mvn/wrapper	v1.0.0	2 years ago
src	Graph traversing fix value null	8 months ago
.gitignore	add git ignore rule for weaver data	10 months ago
CITATION.cff	v1.0.0	2 years ago
LICENSE	v1.0.0	2 years ago
mvnw	v1.0.0	2 years ago
mvnw.cmd	v1.0.0	2 years ago
pom.xml	Edit versions	7 months ago
readme.md	Edit versions	7 months ago

README — Apache-2.0 license

Goblin Weaver

license Apache 2.0

The Weaver is a REST API for querying the Maven Central's dependency graph and enriching it by adding information to it according to the user's needs.

About

No description, website, or topics provide

Readme
Apache-2.0 license
Cite this repository
Activity
Custom properties
3 stars
2 watching
4 forks
Report repository

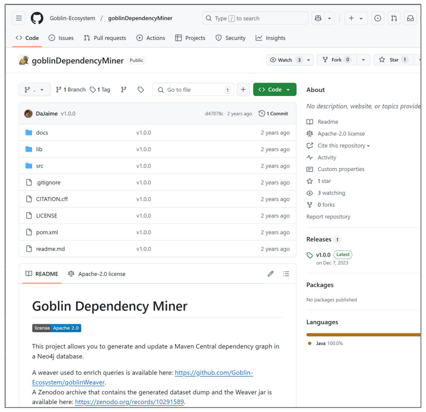
Releases 3
v2.1.0 Latest on Jun 6, 2024
+ 2 releases

Packages
No packages published

Contributors 2
DaJaime Damien Jaime
pascalpoizat Pascal Poizat

Some Useful Links

5. **Goblin Miner:** https://github.com/Goblin-Ecosystem/goblinDependencyMiner



Goblin-Ecosystem / goblinDependencyMiner

Code | Issues | Pull requests | Actions | Projects | Security | Insights

goblinDependencyMiner Public — Watch 3 — Fork 0 — Star 1

1 Branch — 1 Tag — Go to file — Code

DaJaime v1.0.0 — d47078c · 2 years ago — 1 Commit

docs	v1.0.0	2 years ago
lib	v1.0.0	2 years ago
src	v1.0.0	2 years ago
.gitignore	v1.0.0	2 years ago
CITATION.cff	v1.0.0	2 years ago
LICENSE	v1.0.0	2 years ago
pom.xml	v1.0.0	2 years ago
readme.md	v1.0.0	2 years ago

README — Apache-2.0 license

Goblin Dependency Miner

license Apache 2.0

This project allows you to generate and update a Maven Central dependency graph in a Neo4j database.

A weaver used to enrich queries is available here: https://github.com/Goblin-Ecosystem/goblinWeaver.
A Zenodoo archive that contains the generated dataset dump and the Weaver jar is available here: https://zenodo.org/records/10291589.

About

No description, website, or topics provide

Readme
Apache-2.0 license
Cite this repository
Activity
Custom properties
1 star
3 watching
0 forks
Report repository

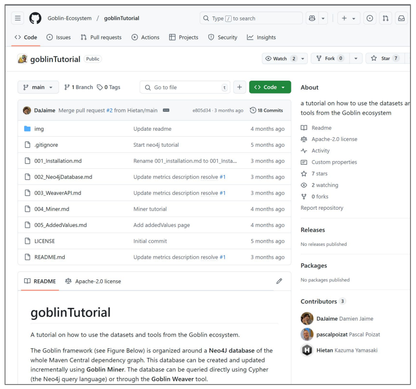
Releases 1
v1.0.0 Latest on Dec 7, 2023

Packages
No packages published

Languages
Java 100.0%

Some Useful Links

6. **Goblin tutorial**: https://github.com/Goblin-Ecosystem/goblinTutorial?tab=readme-ov-file



Goblin-Ecosystem / goblinTutorial

Code | Issues | Pull requests | Actions | Projects | Security | Insights

goblinTutorial Public — Watch 2 — Fork 0 — Star 7

main — 1 Branch — 0 Tags — Go to file — Code

DaJaime Merge pull request #2 from Hietan/main — e805d34 · 3 months ago — 18 Commits

img	Update readme	4 months ago
.gitignore	Start neo4j tutorial	5 months ago
001_Installation.md	Rename 001_installation.md to 001_Insta...	3 months ago
002_Neo4jDatabase.md	Update metrics description resolve #1	3 months ago
003_WeaverAPI.md	Update metrics description resolve #1	3 months ago
004_Miner.md	Miner tutorial	4 months ago
005_AddedValues.md	Add addedValues page	4 months ago
LICENSE	Initial commit	5 months ago
README.md	Update metrics description resolve #1	3 months ago

README — Apache-2.0 license

goblinTutorial

A tutorial on how to use the datasets and tools from the Goblin ecosystem.

The Goblin framework (see Figure Below) is organized around a **Neo4J database** of the whole Maven Central dependency graph. This database can be created and updated incrementally using **Goblin Miner**. The database can be queried directly using Cypher (the Neo4j query language) or through the **Goblin Weaver** tool.

About

a tutorial on how to use the datasets ar tools from the Goblin ecosystem

Readme
Apache-2.0 license
Activity
Custom properties
7 stars
2 watching
0 forks
Report repository

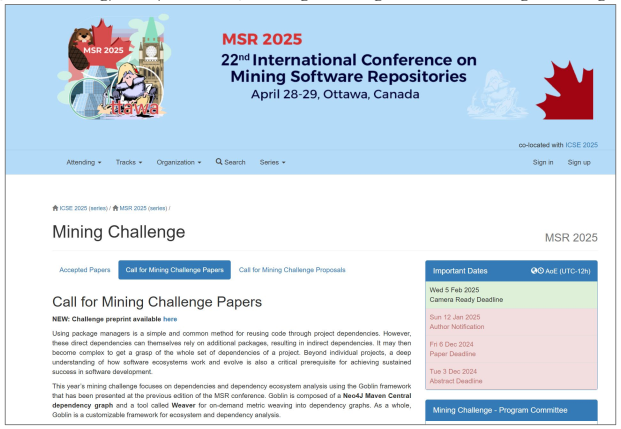
Releases
No releases published

Packages
No packages published

Contributors 3
DaJaime Damien Jaime
pascalpoizat Pascal Poizat
Hietan Kazuma Yamasaki

Taken From MSR 2025 Mining Challenge

https://2025.msrconf.org/track/msr-2025-mining-challenge#Call-for-Mining-Challenge-Papers


MSR 2025
MSR 2025
22nd International Conference on Mining Software Repositories
April 28-29, Ottawa, Canada
co-located with ICSE 2025
Attending
Tracks
Organization
Search
Series
Sign in
Sign up
ICSE 2025 (series) / MSR 2025 (series) /
Accepted Papers
Call for Mining Challenge Papers
Call for Mining Challenge Proposals
Mining Challenge - Program Committee

Mining Challenge

MSR 2025

Call for Mining Challenge Papers

NEW: Challenge preprint available here

Using package managers is a simple and common method for reusing code through project dependencies. However, these direct dependencies can themselves rely on additional packages, resulting in indirect dependencies. It may then become complex to get a grasp of the whole set of dependencies of a project. Beyond individual projects, a deep understanding of how software ecosystems work and evolve is also a critical prerequisite for achieving sustained success in software development.

This year's mining challenge focuses on dependencies and dependency ecosystem analysis using the Goblin framework that has been presented at the previous edition of the MSR conference. Goblin is composed of a **Neo4J Maven Central dependency graph** and a tool called **Weaver** for on-demand metric weaving into dependency graphs. As a whole, Goblin is a customizable framework for ecosystem and dependency analysis.

Important Dates — AoE (UTC-12h)

Wed 5 Feb 2025
Camera Ready Deadline

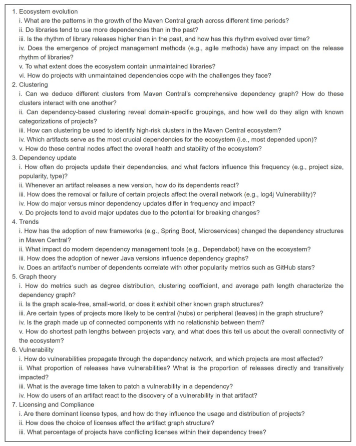
Sun 12 Jan 2025
Author Notification

Fri 6 Dec 2024
Paper Deadline

Tue 3 Dec 2024
Abstract Deadline

Possible Research Questions

https://2025.msrconf.org/track/msr-2025-mining-challenge#Call-for-Mining-Challenge-Papers



1. Ecosystem evolution
   i. What are the patterns in the growth of the Maven Central graph across different time periods?
   ii. Do libraries tend to use more dependencies than in the past?
   iii. Is the rhythm of library releases higher than in the past, and how has this rhythm evolved over time?
   iv. Does the emergence of project management methods (e.g., agile methods) have any impact on the release rhythm of libraries?
   v. To what extent does the ecosystem contain unmaintained libraries?
   vi. How do projects with unmaintained dependencies cope with the challenges they face?
2. Clustering
   i. Can we deduce different clusters from Maven Central's comprehensive dependency graph? How do these clusters interact with one another?
   ii. Can dependency-based clustering reveal domain-specific groupings, and how well do they align with known categorizations of projects?
   iii. How can clustering be used to identify high-risk clusters in the Maven Central ecosystem?
   iv. Which artifacts serve as the most crucial dependencies for the ecosystem (i.e., most depended upon)?
   v. How do these central nodes affect the overall health and stability of the ecosystem?
3. Dependency update
   i. How often do projects update their dependencies, and what factors influence this frequency (e.g., project size, popularity, type)?
   ii. Whenever an artifact releases a new version, how do its dependents react?
   iii. How does the removal or failure of certain projects affect the overall network (e.g., log4j Vulnerability)?
   iv. How do major versus minor dependency updates differ in frequency and impact?
   v. Do projects tend to avoid major updates due to the potential for breaking changes?
4. Trends
   i. How has the adoption of new frameworks (e.g., Spring Boot, Microservices) changed the dependency structures in Maven Central?
   ii. What impact do modern dependency management tools (e.g., Dependabot) have on the ecosystem?
   iii. How does the adoption of newer Java versions influence dependency graphs?
   iv. Does an artifact's number of dependents correlate with other popularity metrics such as GitHub stars?
5. Graph theory
   i. How do metrics such as degree distribution, clustering coefficient, and average path length characterize the dependency graph?
   ii. Is the graph scale-free, small-world, or does it exhibit other known graph structures?
   iii. Are certain types of projects more likely to be central (hubs) or peripheral (leaves) in the graph structure?
   iv. Is the graph made up of connected components with no relationship between them?
   v. How do shortest path lengths between projects vary, and what does this tell us about the overall connectivity of the ecosystem?
6. Vulnerability
   i. How do vulnerabilities propagate through the dependency network, and which projects are most affected?
   ii. What proportion of releases have vulnerabilities? What is the proportion of releases directly and transitively impacted?
   iii. What is the average time taken to patch a vulnerability in a dependency?
   iv. How do users of an artifact react to the discovery of a vulnerability in that artifact?
7. Licensing and Compliance
   i. Are there dominant license types, and how do they influence the usage and distribution of projects?
   ii. How does the choice of licenses affect the artifact graph structure?
   iii. What percentage of projects have conflicting licenses within their dependency trees?