# Documentation Ricgraph - Research in context graph

Rik D.T. Janssen, Utrecht University, ORCID 0000-0001-9510-0802

# 1 Tutorial Ricgraph - Research in context graph

#### 1.1 Ricgraph introduction

Ricgraph, also known as Research in context graph, enables the exploration of researchers, teams, their results, collaborations, skills, projects, and the relations between these items.

Ricgraph can store many types of items into a single graph. These items can be obtained from various systems and from multiple organizations. Ric-graph facilitates reasoning about these items because it infers new relations between items, relations that are not present in any of the separate source systems. It is flexible and extensible, and can be adapted to new application areas.

In this tutorial, we explain the possibilities of Ricgraph. For a more detailed explanation, read the full documentation of Ricgraph on the documentation website https://docs.ricgraph.eu.

#### 1.2 Motivation

Ricgraph is software that is about relations between items. These items can be collected from various source systems and from multiple organizations. We explain how Ricgraph works by applying it to the application area *research information*. We show the insights that can be obtained by combining information from various source systems, insight arising from new relations that are not present in each separate source system.

*Research information* is about anything related to research: research results, the persons in a research team, their collaborations, their skills, projects in which they have participated, as well as the relations between these entities. Examples of *research results* are publications, data sets, and software.

The following sections show three use cases that use different types of information (called *items*): researchers, skills, publications, etc. Most often, these types of information are not stored in one system, so the use cases may be difficult or time-consuming to answer. However, by using Ricgraph, these use cases (and many others) are easy to answer.

#### 1.2.1 Use case for a journalist

As a journalist, I want to find researchers with a certain skill S and their publications, so that I can interview them for a newspaper article. Example skills can be: *climate change* or *stem cells*. The items surrounded by the red line are the solution to this use case.



Figure 1.1: Ricgraph use case for a journalist.

#### 1.2.2 Use case for a librarian

As a librarian, I want to enrich my local research information system with research results from person A that are in other systems (in orange, *RIS*<sub>2</sub>) but not in ours (in green, *RIS*<sub>1</sub>), so that we have a more complete view of research

at our university. The items surrounded by the red line are the solution to this use case.



Figure 1.2: Ricgraph use case for a librarian.

#### 1.2.3 Use case for a researcher

As a researcher A, I want to find researchers from other universities that have co-authored publications written by the co-authors of my own publications, so that I can read their publications to find out if we share common research interests. The items surrounded by the red line are the solution to this use case.



Figure 1.3: Ricgraph use case for a researcher.

#### 1.3 Main contributions of Ricgraph

- Ricgraph can store many types of items in a single graph.
- Ricgraph harvests multiple source systems into a single graph.
- Ricgraph Explorer is the exploration tool for Ricgraph.
- Ricgraph facilitates reasoning about items because it infers new relations between items.
- Ricgraph can be tailored for an application area.
- Ricgraph has a REST API to programmatically get items from Ricgraph.

## 1.4 Installation guide

This section describes how to install both a graph database backend and Ricgraph on a Linux machine. For this, you will need to be able to change to user *root*. Read more about installing Ricgraph without the need to be able to change to user *root*, or about installing Ricgraph as a server on Linux.

- 1. Check the requirements.
- 2. Get the most recent Ricgraph Makefile. Type as regular user (i.e., be sure you are not user *root*):

cd

wget https://raw.githubusercontent.com/UtrechtUniversity/ricgraph/main/Makefile

3. Install Neo4j Community Edition. Only for this step you need to be user *root*. Type:

sudo bash
make install\_enable\_neo4j\_community
exit

On success, the Makefile will print installed successfully.

4. Download and install Ricgraph in your home directory. Type as regular user (i.e., be sure you are not user *root*):

make install\_ricgraph\_singleuser\_neo4j\_community

On success, the Makefile will print installed successfully.

- 5. For more detail, read Fast and recommended way to install Ricgraph for a single user. This is also a good place to start if you encounter errors.
- 6. Done.

## **1.5 Harvesting source systems**

To be able to use Ricgraph, you will first need to harvest information. For some source systems, you need an authentication key. For some others, this is not necessary. To harvest two source systems in Ricgraph without the need for an authentication key, type:

cd \$HOME/ricgraph\_venv make run\_bash\_script

This will harvest the data repository Yoda and the Research Software Directory. It will print a lot of output, and it will take a few minutes. When ready, it will print *Done*.

To read more about harvesting, see Ricgraph harvest scripts.

## **1.6 Browsing using Ricgraph Explorer**

Ricgraph provides an exploration tool, so users do not need to learn a graph query language. This tool is called *Ricgraph Explorer*. As it is a Python Flask application, it can be customized as needed. New queries (buttons) can be added, or the user interface can be modified to fit a certain use case, user group, or application area.

Ricgraph Explorer has several pre-build queries tailored to this application area, each with its own button, for example:

- find a person, a (sub-)organization, a skill;
- when a person has been found, find its identities, skills, research results.

Ricgraph Explorer offers faceted navigation. That means, if a query results in a table with e.g. *journal articles, data sets*, and *software*, you can narrow down on one or more of these categories by checking or unchecking their corresponding checkbox. An alternative view uses tabbed navigation.

This section describes some possibilities of Ricgraph Explorer. For a more extensive description, read Ricgraph Explorer.

Start Ricgraph Explorer to browse the information harvested:

cd \$HOME/ricgraph\_venv
make run\_ricgraph\_explorer

The Makefile will tell you to go to your web browser, and go to http://127.0.0.1: 3030.

#### 1.6.1 A usage flow through Ricgraph Explorer



ters, data sets, other contributions, books, reviews, and software (cf. row with orange rectangle, this is an example of the tabbed navigation). The "comment" column contains the titles of the journal articles. By clicking on an entry in the "value" column, in this case a DOI value, the user will go to this neighbor. Ricgraph Explorer will show a page with persons who have contributed to that journal article.

Figure 1.4: Screenshots of a usage flow through Ricgraph Explorer. Some field values have been blurred for privacy reasons.

The figure above shows screenshots of web pages of Ricgraph Explorer for answering the research question "What are the research results of person A".

The screenshot at the top left is the home page. After clicking "search for a person", Ricgraph Explorer shows a search page (top right). A user types a name, and the person options page is shown (bottom left). After clicking "show research results related to this person", the results page is shown (bottom right). In that page, the rows in the second table are (in this case) the journal article neighbors of the item in the first table (the person the user searched). This person also has other types of research results: book chap-

#### **1.6.2 Example research questions**

In the figure in the previous section, after a click on a value in the "value" column in the bottom right result page, the user will get the persons who have contributed to that research result, as in figure (b) below.



Figure 1.5: Example research questions for Ricgraph.

Clicking "find persons that share any result types with this person" in the bottom left person option page (in the figure in the previous section) corresponds to figure (c), and clicking "show personal information related to this person" corresponds to figure (e).

The research question "What are the research results of person A" from the previous section corresponds to figure (a).

#### 1.7 More information

- Read the full documentation of Ricgraph on https://docs.ricgraph.eu.
- For a gentle introduction in Ricgraph, read the reference publication: Rik D.T. Janssen (2024). Ricgraph: A flexible and extensible graph to explore research in context from various systems. *SoftwareX*, 26(101736). https: //doi.org/10.1016/j.softx.2024.101736.
- Read more about publications, presentations, newsletters, projects with students, use, and mentions of Ricgraph.

#### **1.8 Contact**

Ricgraph has been created and is being maintained by Rik D.T. Janssen from Utrecht University in the Netherlands. You can find contact details at his Utrecht University employee page. He also has an ORCID profile on ORCID 0000-0001-9510-0802. You can contact him for presentations, demos and workshops.

He is also very interested in working together on projects involving Ricgraph. Ricgraph is a flexible platform that brings together information from multiple systems into a single graph. It allows users to analyze this information and explore how it relates to other types of information. We could work together on use cases applying Ricgraph to research information, such as exploring collaborations or analyzing how people or organizations contribute to research results. We could also explore entirely different domains. Any application that involves representing and analyzing interconnected information as nodes and relations in a graph, regardless of the field, is of interest.

# Table of contents

1	Tut	torial Ricgraph - Research in context graph	2
	1.1	Ricgraph introduction	2
	1.2	Motivation	2
	1.3	Main contributions of Ricgraph	3
	1.4	Installation guide	3
	1.5	Harvesting source systems	4
	1.6	Browsing using Ricgraph Explorer	4
	1.7	More information	6
	1.8	Contact	6
2	Do	cumentation Ricgraph - Research in context graph	13
	2.1	What is Ricgraph?	13
	2.2	What to find on the Documentation website for Ricgraph?	13
	2.3	Contact	14
3	Ric	granh - Research in context granh	15
3	<b>Ric</b> 3 1	graph - Research in context graph What is Riegraph?	<b>15</b>
3	<b>Ric</b> 3.1	<b>:graph - Research in context graph</b> What is Ricgraph?	<b>15</b> 15 15
3	<b>Ric</b> 3.1 3.2	:graph - Research in context graph         What is Ricgraph?         Motivation         Motivation         Main contributions of Ricgraph	<b>15</b> 15 15 16
3	<b>Ric</b> 3.1 3.2 3.3	:graph - Research in context graph         What is Ricgraph?         Motivation         Main contributions of Ricgraph         Why Ricgraph?	<b>15</b> 15 15 16
3	Ric 3.1 3.2 3.3 3.4 3.5	Sgraph - Research in context graph         What is Ricgraph?         Motivation         Main contributions of Ricgraph         Why Ricgraph?         Examples	<b>15</b> 15 16 17
3	<b>Ric</b> 3.1 3.2 3.3 3.4 3.5 3.6	Sigraph - Research in context graph         What is Ricgraph?         Motivation         Main contributions of Ricgraph         Why Ricgraph?         Examples         Ricgraph in bullet points	<b>15</b> 15 16 17 17
3	Ric 3.1 3.2 3.3 3.4 3.5 3.6 3.7	cgraph - Research in context graph         What is Ricgraph?         Motivation         Main contributions of Ricgraph         Why Ricgraph?         Examples         Ricgraph in bullet points         Next steps	<b>15</b> 15 16 17 17 18 19
3	Ric, 3.1 3.2 3.3 3.4 3.5 3.6 3.7 3.8	cgraph - Research in context graph         What is Ricgraph?         Motivation         Main contributions of Ricgraph         Why Ricgraph?         Examples         Ricgraph in bullet points         Next steps         Contact	<ol> <li>15</li> <li>15</li> <li>16</li> <li>17</li> <li>17</li> <li>18</li> <li>19</li> <li>20</li> </ol>
3	Ric, 3.1 3.2 3.3 3.4 3.5 3.6 3.7 3.8	graph - Research in context graph         What is Ricgraph?         Motivation         Main contributions of Ricgraph         Why Ricgraph?         Examples         Ricgraph in bullet points         Next steps         Contact	<ol> <li>15</li> <li>15</li> <li>16</li> <li>17</li> <li>17</li> <li>18</li> <li>19</li> <li>20</li> </ol>
3	Ric, 3.1 3.2 3.3 3.4 3.5 3.6 3.7 3.8 Ins	graph - Research in context graph         What is Ricgraph?         Motivation         Main contributions of Ricgraph         Why Ricgraph?         Examples         Ricgraph in bullet points         Next steps         Contact	<b>15</b> 15 15 16 17 17 18 19 20 <b>21</b>
3	Ric, 3.1 3.2 3.3 3.4 3.5 3.6 3.7 3.8 Ins 4.1	graph - Research in context graph         What is Ricgraph?         Motivation         Main contributions of Ricgraph         Why Ricgraph?         Why Ricgraph?         Ricgraph in bullet points         Next steps         Contact	<b>15</b> 15 15 17 17 18 19 20 <b>21</b> 21
3	Ric 3.1 3.2 3.3 3.4 3.5 3.6 3.7 3.8 Ins 4.1 4.2	graph - Research in context graph         What is Ricgraph?         Motivation         Main contributions of Ricgraph         Why Ricgraph?         Examples         Ricgraph in bullet points         Next steps         Contact         Stall and configure Ricgraph         Fast and recommended way to install Ricgraph for a single user         Requirements for Ricgraph	<ol> <li>15</li> <li>15</li> <li>16</li> <li>17</li> <li>17</li> <li>18</li> <li>19</li> <li>20</li> <li>21</li> <li>22</li> <li>22</li> </ol>

	4.4 Ricgraph initialization file	24
	4.5 Ricgraph on Windows	25
	4.6 Steps to take to install Ricgraph for a single user by hand $\ldots$	25
5	Ricgraph harvest scripts	27
	5.1 Introduction to harvest scripts	27
	5.2 Organization abbreviation	28
	5.3 Scripts that harvest multiple sources $\ldots$	28
	5.4 Scripts that harvest a single source $\ldots$	28
	5.5 Order of running the harvest scripts	31
	5.6 How to make your own harvesting scripts	31
6	Ricgraph Explorer	32
	6.1 How to start Ricgraph Explorer	32
	6.2 Working with Ricgraph Explorer	33
	6.3 Home page	34
	6.4 Search page	35
	6.5 Person options page	36
	6.6 Organization options page	37
	6.7 Results page	38
	6.8 Browsing Ricgraph	38
	6.9 Showing a privacy statement or privacy measures document	38
	6.10 Video Ricgraph Explorer	38
7	Ricgraph publications, presentations, use, and mentions	40
	7.1 Ricgraph publications	40
	7.2 Ricgraph presentations	41
	7.3 Ricgraph newsletters $\ldots$	42
	7.4 Ricgraph projects with students $\ldots$	42
	7.5 Ricgraph use $\ldots$	42
	7.6 Ricgraph mentions	42
	7.7 Arjan Sieverink presents Ricgraph at Pure International Conference 2023	43
8	Ricgraph as a server on Linux	44
	8.1 Fast and recommended way to install Ricgraph as a server $\ldots$	44
	8.2 Run Ricgraph scripts from the command line or as a cronjob $\ldots$	45
	8.3 Use a service unit file to run Ricgraph Explorer and the Ricgraph REST API	46

	<ul> <li>8.4 Use Apache, WSGI, and ASGI to make Ricgraph Explorer and the Ricgraph REST API accessible from outside your virtual machine</li> <li>8.5 Use Nginx, WSGI, and ASGI to make Ricgraph Explorer and the Ricgraph REST API accessible from outside your virtual machine</li> <li>8.6 How to install Ricgraph and Ricgraph Explorer on SURF Research Cloud</li> <li>8.7 Steps to take to install Ricgraph as a server by hand</li> </ul>	47 48 50 51
g	Ricgraph in a container9.1 Notes on the Ricgraph Podman container9.2 Install Podman9.3 Install and run the Ricgraph Podman container9.4 Advanced use of the Ricgraph Podman container	<b>54</b> 54 54 55 55 56
1	<ul> <li>0 Ricgraph with Neo4j graph database backend</li> <li>10.1 Neo4j Community Edition</li> <li>10.2 Neo4j Desktop</li> <li>10.3 Dumping, restoring, and emptying the Ricgraph database</li> <li>10.4 How to reset the Neo4j Community Edition password</li> <li>10.5 How to solve an AttributeError: Neo4jDriver object has no attribute executequery</li> </ul>	<b>57</b> 57 58 61 64 64
1	<b>1 Ricgraph with Memgraph graph database backend</b> 11.1 Install and start Memgraph	<b>66</b> . 66
1	2 Ricgraph REST API         12.1 Installation of the Ricgraph REST API         12.2 Use of the Ricgraph REST API         12.3 Ricgraph REST API endpoint documentation         12.4 Generate the Ricgraph REST API documentation page	67 . 67 . 67 . 68 . 68
1	3 Ricgraph - Research in context graph REST API         13.1 What to find on this page?         13.2 How to use         13.3 Technicalities         13.4 Base URL         13.5 Endpoints         13.6 GET /person/search         13.7 GET /person/all_information         13.8 GET /person/share_research_results         13.9 GET /person/collaborating_organizations         13.10GET /person/enrich	69 . 69 . 69 . 69 . 70 . 71 . 71 . 71 . 72 . 72
		•

	13.11 GET /organization/search	• 7	73
	13.12GET /organization/all_information	• 7	73
	13.13GET /organization/information_persons_results	. 7	74
	13.14GET /organization/enrich	• 7	75
	13.15GET /competence/search	. 7	76
	13.16GET /competence/all_information	. 7	76
	13.17GET /broad_search	. 7	76
	13.18GET /advanced_search	- 7	77
	13.19GET /get_all_personroot_nodes	• 7	77
	13.20GET /get_all_neighbor_nodes	• 7	77
	13.21GET /get_ricgraph_list	. 7	78
14	4 Implementation details	\$	80
1	1/. 1. Person identifiers		30
	14.2 Person-root node in Ricgraph	. 8	30
	14.3 Properties of nodes in Ricgraph	. 8	30
1	5 Ricgraph miscellaneous scripts	8	82
	15.1 Construct a Ricgraph from a csv file (construct_ricgraph_from_csv)	. 8	82
	15.2 Import nodes and edges from a csv file, raw version (ricgraph_import_raw_from_csv)	. 8	83
	15.3 Export nodes and edges to a csv file, raw version (ricgraph_export_raw_to_csv)	. 8	34
	15.4 Count the number of organizations that contributed to a category (count_organizations_contributed_to_category)	. 8	34
	15.5 Export nodes to a file (export_person_identifiers and export_person_node_properties)	. 8	85
	15.6 Delete personal data from Ricgraph (delete_personal_data)	. 8	85
	15.7 Rename (sub-)organizations in Ricgraph (rename_organizations)	. 8	36
	15.8 Script to enrich persons (enrich_orcids_scopusids)	. 8	87
	15.9 Script to find person identifiers pointing to different persons (find_double_pids)	. 8	87
	15.10Create a table of contents of the Ricgraph documentation (create_toc_documentation)	. 8	87
	15.11Create an index of the Ricgraph documentation (create_index_documentation)	. 8	87
	15.12Create the Ricgraph REST API documentation (convert_openapi_to_mddoc)	. 8	87
16	5 Ricgraph script writing	1	88
	16.1. How to make your own harvesting scripts	ş	28
	16.2 General program structure of a Python script using Ricgraph	. c	28
	16.2 Structure of a Python script that is harvesting data	. c	88
	16./ Function calls for inserting nodes	. c ş	Ro
	16.5 Function call for unifying personal identifiers		30
	- 101 J Lancaon can for annying percental activities a second s	· ·	- 7

16.6 Function calls to create, read (find), update and delete (CRUD) nodes	89
16.7 Function calls to get neighbors of nodes	89
17 Ricgraph comparison	91
17.1 General overview	92
17.2 Extendability	93
18 Ricgraph known bugs	95
18.1 Error while emptying Ricgraph with Neo4j	• • 95
19 Table of contents Ricgraph documentation	96
20 Index Ricgraph documentation	100

**Full documentation Ricgraph** 

# 2 Documentation Ricgraph - Research in context graph

## 2.1 What is Ricgraph?

Ricgraph, also known as Research in context graph, enables the exploration of researchers, teams, their results, collaborations, skills, projects, and the relations between these items.

Ricgraph can store many types of items into a single graph. These items can be obtained from various systems and from multiple organizations. Ric-graph facilitates reasoning about these items because it infers new relations between items, relations that are not present in any of the separate source systems. It is flexible and extensible, and can be adapted to new application areas.

Currently, there are people working on the following extensions for Ric-graph:

- BackToPure is a tool designed to enhance an organization's Research Information System Pure by enriching its content. BackToPure can identify items (such as publications, data sets or software) that exist in other external sources but are missing from the organization's Pure, and then insert (enrich) those items into Pure. The result is a more complete overview of research at that organization. Status: experimental stage (beta).
- A chatbot that allows you to "talk" to Ricgraph. You can formulate questions in plain English, such as "Please give me the research results of the Geosciences faculty of Utrecht University?" or "With what organizations does that faculty collaborate?". It uses a local Large Language Model. Status: planning stage (pre-alpha).
- A project that uses AI and Large Language Models to cluster and visualize large amounts of research information. It assigns *topics* to publications, data sets, and software. By selecting a number of topics, research results

are grouped, and possibly experts on these topics can be found. Status: planning stage (pre-alpha).

• A tool that helps Pure administrators to clean up "external organizations" in Pure. Status: planning stage (pre-alpha).

# 2.2 What to find on the Documentation website for Ricgraph?

This documentation website offers quite a bit of information about Ricgraph. In the yellow sidebar you can find the various options. Basically, you can read the Ricgraph tutorial, or browse through the full documentation:

- Ricgraph tutorial:
  - Read the tutorial on this webpage.
  - Read the tutorial as pdf.
- Full documentation:
  - Start reading at the Ricgraph README.
  - Choose something from the documentation tree in the yellow sidebar.
  - Use the search box.
  - Of course, every webpage has links that allow you to go to other sections or webpages.
  - Read the full documentation as pdf. This pdf is the concatenation of all Ricgraph documentation pages in one pdf.

This website has been generated from the Ricgraph GitHub documentation.

#### 2.3 Contact

Ricgraph has been created and is being maintained by Rik D.T. Janssen from Utrecht University in the Netherlands. You can find contact details at his Utrecht University employee page. He also has an ORCID profile on ORCID 0000-0001-9510-0802. You can contact him for presentations, demos and workshops.

He is also very interested in working together on projects involving Ricgraph. Ricgraph is a flexible platform that brings together information from multiple systems into a single graph. It allows users to analyze this information and explore how it relates to other types of information. We could work together on use cases applying Ricgraph to research information, such as exploring collaborations or analyzing how people or organizations contribute to research results. We could also explore entirely different domains. Any application that involves representing and analyzing interconnected information as nodes and relations in a graph, regardless of the field, is of interest.

# 3 Ricgraph - Research in context graph

#### 3.1 What is Ricgraph?

Ricgraph, also known as Research in context graph, enables the exploration of researchers, teams, their results, collaborations, skills, projects, and the relations between these items.

Ricgraph can store many types of items into a single graph. These items can be obtained from various systems and from multiple organizations. Ric-graph facilitates reasoning about these items because it infers new relations between items, relations that are not present in any of the separate source systems. It is flexible and extensible, and can be adapted to new application areas.

Currently, there are people working on the following extensions for Ric-graph:

- BackToPure is a tool designed to enhance an organization's Research Information System Pure by enriching its content. BackToPure can identify items (such as publications, data sets or software) that exist in other external sources but are missing from the organization's Pure, and then insert (enrich) those items into Pure. The result is a more complete overview of research at that organization. Status: experimental stage (beta).
- A chatbot that allows you to "talk" to Ricgraph. You can formulate questions in plain English, such as "Please give me the research results of the Geosciences faculty of Utrecht University?" or "With what organizations does that faculty collaborate?". It uses a local Large Language Model. Status: planning stage (pre-alpha).
- A project that uses AI and Large Language Models to cluster and visualize large amounts of research information. It assigns *topics* to publications, data sets, and software. By selecting a number of topics, research results

are grouped, and possibly experts on these topics can be found. Status: planning stage (pre-alpha).

• A tool that helps Pure administrators to clean up "external organizations" in Pure. Status: planning stage (pre-alpha).

Throughout this documentation, we illustrate how Ricgraph works by applying it to the application area research information.

#### 3.2 Motivation

Ricgraph is software that is about relations between items. These items can be collected from various source systems and from multiple organizations. We explain how Ricgraph works by applying it to the application area *research information*. We show the insights that can be obtained by combining information from various source systems, insight arising from new relations that are not present in each separate source system.

*Research information* is about anything related to research: research results, the persons in a research team, their collaborations, their skills, projects in which they have participated, as well as the relations between these entities. Examples of *research results* are publications, data sets, and software.

The following sections show three use cases that use different types of information (called *items*): researchers, skills, publications, etc. Most often, these types of information are not stored in one system, so the use cases may be difficult or time-consuming to answer. However, by using Ricgraph, these use cases (and many others) are easy to answer, as will be explained throughout this documentation.

Although this documentation illustrates Ricgraph in the application area research information, the principle "relations between items from various source systems" is general, so Ricgraph can be used in other application areas.

#### 3.2.1 Use case for a journalist

As a journalist, I want to find researchers with a certain skill S and their publications, so that I can interview them for a newspaper article. Example skills can be: *climate change* or *stem cells*. The items surrounded by the red line are the solution to this use case.



Figure 3.1: Ricgraph use case for a journalist.

#### 3.2.2 Use case for a librarian

As a librarian, I want to enrich my local research information system with research results from person A that are in other systems (in orange, *RIS*<sub>2</sub>) but not in ours (in green, *RIS*<sub>1</sub>), so that we have a more complete view of research at our university. The items surrounded by the red line are the solution to this use case.



Figure 3.2: Ricgraph use case for a librarian.

#### 3.2.3 Use case for a researcher

As a researcher A, I want to find researchers from other universities that have co-authored publications written by the co-authors of my own publications, so that I can read their publications to find out if we share common research interests. The items surrounded by the red line are the solution to this use case.



Figure 3.3: Ricgraph use case for a researcher.

## 3.3 Main contributions of Ricgraph

- Ricgraph can store many types of items in a single graph.
- Ricgraph harvests multiple source systems into a single graph.
- Ricgraph Explorer is the exploration tool for Ricgraph.

- Ricgraph facilitates reasoning about items because it infers new relations between items.
- Ricgraph can be tailored for an application area.

## 3.4 Why Ricgraph?

Ricgraph can answer questions like:

- Which researcher has contributed to which publication, dataset, software package, project, etc.?
- Given e.g. a dataset, software package, or project, who has contributed to it?
- What identifiers does a researcher have (e.g. ORCID, ISNI, organization employee ID, email address)?
- What skills does a researcher have?
- Show a network of researchers who have worked together?
- Which organizations have worked together?

Also, more elaborate information can be found using Ricgraph and Ricgraph Explorer, the exploration tool for Ricgraph:

- You can find information about persons or their results in a (sub-)organization (unit, department, faculty, university). For example, you can find out what data sets or software are produced in your faculty. Or the skills of all persons in your department. Of course this is only possible in case you have harvested them.
- You can find out with whom a person shares research output types. For example, you can find out with whom someone shares software or data sets.
- You can get tables showing how you can enrich a source system based on other systems you have harvested. For example, suppose you have harvested both Research Information System Pure and OpenAlex, using this feature you can find out which publications in OpenAlex are not in Pure. You might want to add those to Pure.
- You can get a table that shows the overlap in harvests from different source systems. For example, after a query to show all ORCID nodes, the

table summarizes the number of ORCID nodes which were only found in one source, and which were found in multiple sources. Another table gives a detailed overview how many nodes originate from which different source systems. Then, you can drill down by clicking on a number in one of these two tables to find the nodes corresponding to that number.

If you would like to get this information programmatically, you can use the Ricgraph REST API.

With Ricgraph, you can get metadata from objects from any source system you'd like. You run the harvest script for that system, and data will be imported in Ricgraph and will be combined automatically with data which is already there. Ricgraph provides harvest scripts for the systems mentioned above. Scripts for other sources can be written easily.

# 3.5 Examples

See the figures below for example graphs that show how Ricgraph works. In this section, we use the following symbols and colors.



The following figure shows one person with several research outputs. It shows one person *A* using a *person-root* node, a node which "represents" a

person as it is called in Ricgraph. This person has contributed to three articles, two data sets and one software package. Two articles and one data set are from the Research Information System Pure (their color is green), one data set is from the data repository Yoda (in orange), one article is from OpenAlex (in purple), and the software package is from the Research Software Directory (in blue).



Figure 3.4: One person with several research outputs.

The following figure shows several persons with several research outputs (the symbols) and how these are related (i.e. which person contributed to which research output). It also shows from which source system these research outputs have originated (using different colors).



Figure 3.5: Several persons with several research outputs.

The following figure shows one person having several identifiers and several research outputs. This person has two different ORCIDs, one ISNI, one SCO-PUS\_AUTHOR\_ID, and two FULL\_NAMEs (which differ in spelling). These

identifiers have also been obtained from different source systems, as their color indicates.



Figure 3.6: One person with several identifiers and research outputs.

The following figure shows how three persons have contributed to one research output.



Figure 3.7: Three persons with one research output.

## 3.6 Ricgraph in bullet points

- The philosophy of Ricgraph is that it stores metadata, not the objects the metadata refer to. To access an object, a node has a link to that object in the system it was obtained from.
- We have chosen a graph as a data structure, since it is a logical and efficient method to access objects which are close to objects they have a relation to. For example, starting with a person, its research outputs are only one step away by following one edge, and other contributors to that research output are again one step (edge) away.

- Ricgraph can be used to store, manipulate and read metadata of any object that has a relation to another object, as long as every object can be "represented" by at least a *name* and a *value*. In Ricgraph, one node represents one object, and an edge represents the relation between two objects.
- Ricgraph and Ricgraph Explorer are written in Python. You can use two different graph database backends: Neo4j and Memgraph.
- The objective of Ricgraph is to get metadata from objects from a source system in a process called "harvesting". That means that e.g. persons and publications can be harvested from one system, data sets from another system, and software from a third system. Everything found will be combined into one graph.
- Ricgraph can harvest from many sources, and you can write your own harvesting scripts. Example scripts are included to harvest from the OpenAlex, the Research Information System Pure, the data repository Yoda, the Research Software Directory, and for the Utrecht University staff pages.
- Ricgraph is an ID resolver. It can, given an identifier of a person, easily find other identifiers of that person. When new identifiers are found when harvesting from new systems, they will be added automatically.
- Ricgraph can enrich a source system based on information that is present in one source system, but not in another source system. See the librarian use case above.

## 3.7 Next steps

#### 3.7.1 Further information about Ricgraph

- Explore the full documentation of Ricgraph on https://docs.ricgraph.eu.
- For a gentle introduction in Ricgraph, read the reference publication: Rik D.T. Janssen (2024). Ricgraph: A flexible and extensible graph to explore research in context from various systems. *SoftwareX*, 26(101736). https://doi.org/10.1016/j.softx.2024.101736.
- For a further introduction, you might want to read these two presentations:

- A general presentation about Ricgraph, presenting Ricgraph in a visual manner: https://doi.org/10.5281/zenodo.12634234.
- A presentation explaining how to enrich the Research Information System Pure (and other source systems) using Ricgraph and Back-ToPure: https://doi.org/10.5281/zenodo.12634658. This presentation explains that after harvesting several Pure systems (from different institutions) and other source systems, such as OpenAlex, one can enrich its own Pure A by using information in other source systems, not present in one's own Pure A.
- Read more about publications, presentations, newsletters (to subscribe, go to Ricgraph Contact), projects with students, use, and mentions of Ricgraph.
- The Ricgraph main website is www.ricgraph.eu. Alternative sites are www.ricgraph.nl and www.ricgraph.com. These are generated using the ricgraph-documentation GitHub repository.

# 3.7.2 Steps to take if you would like to install Ricgraph and harvest data

- Install and configure Ricgraph.
- Start harvesting data, see Ricgraph harvest scripts, e.g. by doing a harvest for Utrecht University data sets and software. You will observe that the information from two sources is neatly combined into one graph.
- Unfortunately, there is a bug, see known bugs. This bug may occur if you start a harvest script, and as first step in the script you want to empty Ricgraph. In that case, a Python error might occur while emptying Ricgraph. Follow the link to read more and find out how to repair that.

# 3.7.3 Steps to take if you would like to use Ricgraph

- Use Ricgraph Explorer, the exploration tool for Ricgraph.
- $\cdot\,$  Use the Ricgraph REST API, the REST API for Ricgraph.
- You can also write your own harvest scripts, or modify any part of the Ricgraph code.

#### 3.8 Contact

Ricgraph has been created and is being maintained by Rik D.T. Janssen from Utrecht University in the Netherlands. You can find contact details at his Utrecht University employee page. He also has an ORCID profile on ORCID 0000-0001-9510-0802. You can contact him for presentations, demos and workshops.

He is also very interested in working together on projects involving Ricgraph. Ricgraph is a flexible platform that brings together information from multiple systems into a single graph. It allows users to analyze this information and explore how it relates to other types of information. We could work together on use cases applying Ricgraph to research information, such as exploring collaborations or analyzing how people or organizations contribute to research results. We could also explore entirely different domains. Any application that involves representing and analyzing interconnected information as nodes and relations in a graph, regardless of the field, is of interest.

About three to four times a year we send out a newsletter. Please click here to subscribe to it. Your email program will open with a subscribe email text. Please complete it by filling in your name and email address. We also have a newsletter archive.

If you like Ricgraph, please give it a GitHub Star by clicking on the top right *Star* button on the Ricgraph GitHub page https://github.com/ UtrechtUniversity/ricgraph. If you have any suggestions or improvements, please let me know by creating a GitHub *Issue* at the top left of that page.

This README.md file is part of the documentation for version 2.12 of Ricgraph – Research in context graph.

# 4 Install and configure Ricgraph

This page describes how to install Ricgraph for a single user on Linux. If you would like to use Ricgraph in a multi-user environment on Linux, you will need to install Ricgraph differently. In case you have no idea what would be the best for your situation, please install Ricgraph for a single user on Linux, as described on this page. Or go for Ricgraph in a container.

To install and run Ricgraph for a single user, read Fast and recommended way to install Ricgraph for a single user.

On this page, you can find:

- · Fast and recommended way to install Ricgraph for a single user
- Requirements for Ricgraph
- Ricgraph Makefile
- Ricgraph initialization file
- Ricgraph on Windows
- Steps to take to install Ricgraph for a single user by hand

Return to main README.md file.

# 4.1 Fast and recommended way to install Ricgraph for a single user

#### 4.1.1 You can change to user *root*

To follow this procedure, you need to be able to change to user *root*. This is the recommended method to install Ricgraph for a single user, since it will install everything automatically (by using the Ricgraph Makefile).

1. Check the requirements.

2. Get the most recent Ricgraph Makefile. Type as regular user (i.e., be sure you are not user *root*):

cd

wget https://raw.githubusercontent.com/UtrechtUniversity/ricgraph/main/Makefile

Read more at Ricgraph Makefile.

3. Install Neo4j Community Edition. Only for this step you need to be user *root*. Type:

sudo bash
make install\_enable\_neo4j\_community
exit

On success, the Makefile will print installed successfully.

4. Download and install Ricgraph in your home directory. Type as regular user (i.e., be sure you are not user *root*):

make install\_ricgraph\_singleuser\_neo4j\_community

On success, the Makefile will print installed successfully.

5. Harvest two source systems in Ricgraph:

cd \$HOME/ricgraph\_venv make run\_bash\_script

This will harvest two source systems, the data repository Yoda and the Research Software Directory. It will print a lot of output, and it will take a few minutes. When ready, it will print *Done*.

To read more about harvesting data, see Ricgraph harvest scripts. To read more about writing harvesting scripts, see Ricgraph script writing.

6. Start Ricgraph Explorer to browse the information harvested:

```
cd $HOME/ricgraph_venv
make run_ricgraph_explorer
```

The Makefile will tell you to go to your web browser, and go to http://127. 0.0.1:3030. Read more at Ricgraph Explorer.

If everything succeeded, you are done installing Ricgraph for a single user. If not, sections Steps to take to install Ricgraph for a single user by hand or Install and start Neo4j Community Edition graph database backend may help in finding solutions.

#### 4.1.2 You cannot change to user *root*

If you are not able to change to user *root*, change step 3 and 4 in the previous section. This is a less recommended method to install Ricgraph for a single user, since you have to do a number of things manually.

3. Install Neo4j Desktop. Type:

```
make install_neo4j_desktop
```

On success, the Makefile will print *installed successfully*. You will need to do a number of post install steps. Any time you want to use Neo4j Desktop, you will need to start it by hand, read Start Neo4j Desktop.

4. Download and install Ricgraph in your home directory. Read more at the sections below. Type:

```
{\tt make install\_ricgraph\_singleuser\_neo4j\_desktop}
```

On success, the Makefile will print installed successfully.

If everything succeeded, you are done installing Ricgraph for a single user. If not, sections Steps to take to install Ricgraph for a single user by hand or Install and start Neo4j Community Edition graph database backend may help in finding solutions.

# 4.2 Requirements for Ricgraph

- To install the graph database backend and Ricgraph, you will need access to a Linux machine. All instructions in this documentation are written for a Linux machine, except when otherwise noted. This doesn't mean that you cannot install the graph database backend and Ricgraph *directly* on a Mac or Windows machine, it means that the author has no experience with this, except for Read here if you would like to install Ricgraph on Windows.
- If you have a Mac or Windows machine, you can create a Linux virtual machine (VM) on it using e.g. VirtualBox. This also holds if you have a Linux machine and would like to install the graph database backend and Ricgraph separated from your usual working environment.
- A virtual machine is like a separate computer in a box that is running on your own computer. You can install anything in it, and it will not interfere with your usual working environment on your machine.
- So, to install the graph database backend and Ricgraph, you can use your own Linux machine, your own Linux virtual machine, a Linux VM provided by your organization, or a Linux VM provided by a Cloud provider such as SURF Research Cloud.

Own Linux machine:

- You already have Linux, skip to the next bullet in this list.

#### Own Linux virtual machine:

- On your own Mac, Windows, or Linux machine, install VirtualBox. Create a virtual machine in VirtualBox. Install a Linux distribution "in" the virtual machine you have just created. Then, install the graph database backend and Ricgraph in the Linux virtual machine you have just created, as described on this page or on Install and configure Ricgraph as a server.
- There are many tutorials on installing VirtualBox on internet. For example, How to Install VirtualBox on Ubuntu (Beginner's Tutorial).

- Also, you will need to install the VirtualBox GuestAdditions. They will enable e.g. shared folders with the host and automatic adjustment of guest display resolution. You install them inside a virtual machine after the guest operating system has been installed. Read, for example, How to Install & Use VirtualBox Guest Additions on Ubuntu.
- Very probably you will *not* need the VirtualBox Extension Pack. You need to be user *root* (Linux) or *Administrator* (Windows) on your computer to be able to do this. Don't do it for now.
- Almost any Linux distribution will work, the author uses both Open-SUSE Leap and Ubuntu. Others will also work.
- For the configuration in VirtualBox, a VM of size 25GB with 8GB memory will work. This depends on the (size of the) sources you plan to harvest and the capabilities of your computer. The more, the better. The author uses a VirtualBox VM of size 35GB with 10GB memory and 3 vCPUs on an 11th gen Intel i7 mobile processor.

Linux VM provided by your organization:

- Ask your organization.

Linux VM on SURF Research Cloud:

- Read How to install Ricgraph and Ricgraph Explorer on SURF Research Cloud.
- You will need at least Python 3.9 or newer. Ricgraph has been developed with Python 3.11. Check your Python version by typing:

#### python3 --version

Look at the first two numbers. If they are lower than 3.9, you will need to install a newer Python version. How you do it, is dependent on your Linux version. E.g., if you have Ubuntu 20.04, you can install Python 3.11 as follows:

- Login as user root.
- Type the following commands:

add-apt-repository ppa:deadsnakes/ppa apt install python3.11

- Exit from user *root*. For other Linux distributions there will be similar commands.
- You need the linux command make. Try typing

make

If you get an error, install it using your package manager.

# 4.3 Ricgraph Makefile

A Ricgraph installation involves a number of steps. Ricgraph uses *make* and a Makefile to make installation of (parts of) Ricgraph easier. A Makefile automates a number of these steps. A make command is executed by typing:

make [target]

or

make [make command line parameter]=[value] [target]

To use the Ricgraph Makefile, first go to your home directory on Linux and then download the most recent version from the GitHub repository, by typing:

```
cd
```

 $wget \ https://raw.githubusercontent.com/UtrechtUniversity/ricgraph/main/Makefile$ 

In the example above, the *[target]* specifies what has to be done. Assuming that you are in your home directory, you can execute one of these commands to find the possible targets:

make make help make allhelp

You can add command line parameters to the make command, e.g. to get the Ricgraph *cutting edge* version, or to specify an installation path. In that case, the make command may look like

```
make ricgraph_version=cuttingedge [target]
```

or

```
make ricgraph_server_install_dir=/opt/ricgraph_venv [target]
```

Look in file *Makefile* for possibilities. Any variable defined in the Makefile can be used as make command line parameter. For an example, see the Podman Containerfile in file *Containerfile*.

Most often, you do not need to install the make command, but if you get a "command not found" error message, you need to install it using your Linux package manager.

If you read the documentation below or on page Ricgraph as a server on Linux, you will notice that some sections start with mentioning a Makefile command. That means, that if you execute that command, the steps in that section will be done automatically. Sometimes, you will have to do some post-install steps, e.g. because you have to choose a password for the graph database.

## 4.4 Ricgraph initialization file

Ricgraph requires an initialization file. A sample file is included as *ricgraph.ini-sample*. You need to copy this file to *ricgraph.ini* and modify it to include settings for your graph database backend, and API keys and/or email addresses for other systems you plan to use. If you have used the Ricgraph Makefile, the copying and the settings for the graph database backend will

have been set, but you still need to fill in API keys and/or email addresses for other systems you plan to use.

#### 4.4.1 Settings for the graph database backend

Ricgraph has a [*GraphDB*] section where you have to specify the graph database backend that you will be using. First, you will need to set the parameter *graphdb* to the graph database backend name (you can choose between *neo4j* and *memgraph*). Further down that section, you will have to fill in six parameters for hostname, port number, username, etc. The comments in the initialization file explain how to do that.

#### 4.4.2 Extending Ricgraph with new properties in the nodes

Optionally, you can extend Ricgraph by adding new properties of nodes.

## 4.4.3 RICGRAPH\_NODEADD\_MODE

There is a parameter *RICGRAPH\_NODEADD\_MODE* in the initialization file that influences how nodes are added to Ricgraph. Suppose we harvest a source system and that results in the following table:

ISNI	ORCID
ISNI-1	0000-0001-1111-1111
ISNI-2	0000-0001-1111-2222
ISNI-3	0000-0001-1111-2222
ISNI-4	0000-0001-1111-3333

*ISNI-2* and *ISNI-3* have the same ORCID. This may be correct, e.g. if the person with the *ORCID* has multiple *ISNI* records. But it also may be incorrect, e.g. if *ISNI-2* and *ISNI-3* do not refer to the same person, possibly caused by a typing mistake in a source system. There is no way for Ricgraph to know which of these two options it is.

RICGRAPH\_NODEADD\_MODE can be either *strict* or *lenient*:

- strict (default setting): only add nodes to Ricgraph which conform to the model described in the Implementation details. In the example above, ORCID 0000-0001-1111-2222 will not be inserted.
- *lenient*: add every node. In the example above, *ORCID 0000-0001-1111-* 2222 will be inserted.

This will have the following consequences:

- *strict*: since *ORCID 0000-0001-1111-2222* will not be inserted, a research output from a person with that *ORCID* may not be inserted in Ricgraph. Or the research output will be inserted, but it might not be linked to the person with this *ORCID*.
- *lenient*: as has been described Implementation details, *person-root* "represents" a person. Person identifiers (such as *ORCID*) and research outputs are connected to the *person-root* node of a person. That means that the *person-root* node is connected to everything a person has contributed to.

In the example above, ORCID 0000-0001-1111-2222 is inserted. That means that the *person-roots* of the two persons with *ISNI-2* or *ISNI-3* are "merged", and that all research outputs of *ISNI-2* and *ISNI-3* will be connected to one *person-root* node. After this has been done, there is no way to know which research output belongs to *ISNI-2* or *ISNI-3*.

As said, that is fine if *ISNI*-2 and *ISNI*-3 refer to the same person (having two ISNIs), but not fine if they refer to two different persons.

*Lenient* is advisable if the sources you harvest from do not contain errors. However, with source systems that contain a lot of information this is not likely, therefore the default is *strict*.

#### 4.5 Ricgraph on Windows

The easiest way to go is to Install and use Ricgraph in a container. This is relatively quick but it offers limited possibilities.

If you would like to go for a "full" install of Ricgraph on Windows using either Install and configure Ricgraph for a single user or Install and configure Ricgraph as a server, you are very probably the first person to do so, as far as known. The creator of Ricgraph has no experience in developing software on Windows. So please let me know which steps you have taken, so I can add them to this documentation. If you are a Windows user, I would recommend to create a Linux virtual machine using e.g. VirtualBox as explained in section Requirements, and install Ricgraph in that virtual machine as described above.

# 4.6 Steps to take to install Ricgraph for a single user by hand

Skip this section if you have done the Fast and recommended way to install Ricgraph for a single user and there were no errors.

- 1. Install your graph database backend.
- 2. Download Ricgraph.
- 3. Use a Python virtual environment and install Python requirements.
- 4. Create and update the Ricgraph initialization file. This is also the place where you specify which graph database backend you use.
- 5. Start harvesting data, see Ricgraph harvest scripts, or writing scripts, see Ricgraph script writing.
- 6. Start browsing using Ricgraph Explorer.

## 4.6.1 Install your graph database backend

Install your graph database backend (choose one of these):

- Install and start Neo4j Community Edition (recommended, only possible if you are able to change to user *root*).
- Install Neo4j Desktop Optional: Install the Bloom configuration.
- Install and start Memgraph.

#### 4.6.2 Download Ricgraph

You can choose two types of downloads for Ricgraph:

- The latest released version. Go to the Release page of Ricgraph, choose the most recent version, download either the *zip* or *tar.gz* version.
- The "cutting edge" version. Go to the GitHub page of Ricgraph, click the green button "Code", choose tab "Local", choose "Download zip".

# 4.6.3 Use a Python virtual environment and install Python requirements

To be able to use Ricgraph, you will need a Python virtual environment. Virtual environments are a kind of lightweight Python environments, each with their own independent set of Python packages installed in their site directories. A virtual environment is created on top of an existing Python installation. There are two ways of doing this:

- Using Python's venv module;
- Using a Python Integrated development environment (IDE).

#### 4.6.3.1 Using Python's venv module

- Using Python's venv module. Read Create a Python virtual environment and install Ricgraph in it. This documentation has been written for a multi-user installation of Ricgraph. To use it for a single users install (as you are doing since you are on this page):
  - Suppose you are a user with login *alice*.
  - Suppose your home directory is /home/alice (check this by typing cd followed by pwd).
  - For every occurrence of */opt* in Create a Python virtual environment and install Ricgraph in it, read */home/alice*, and ignore any references to "login as user *root*" and chown.
  - Follow the other instructions as written.

#### 4.6.3.2 Using a Python Integrated development environment (IDE)

- Using a Python Integrated development environment (IDE), such as Py-Charm. An IDE will automatically generate a virtual environment, and any time you use the IDE, it will "transfer" you to that virtual environment. It will also help to execute and debug your scripts.
  - If PyCharm does not automatically generate a virtual environment, you need to go to File -> Settings -> Project: [your project name]
     -> Python Interpreter, and check if there is a valid interpreter in the right column next to "Python Interpreter". If not, add one, using "Add Interpreter", and choose for example "Add Local Interpreter". A venv will be generated.
  - Next, unzip or tar xf the downloaded file for Ricgraph (see previous section).
  - Install the Python requirements. Depending on the Python IDE, single or double-click on file *requirements.txt*. Probably, a button or text appears that asks you to install requirements. Click on it.
     If this does not work, type in the IDE (PyCharm) Terminal:

pip3.11 install -r requirements.txt

You may want to change 3.11 in *pip3.11* for the Python version you use.

# 5 Ricgraph harvest scripts

This page describes scripts for harvesting sources and inserting the results in Ricgraph. They can be found in directory *harvest* and *harvest\_multiple\_sources*. Read more about scripts to import and export items from Ricgraph, about scripts to enhance (finding, enriching, etc.) information in Ricgraph, or about writing your own scripts.

On this page, you can find:

- Introduction to harvest scripts
- Organization abbreviation
- Scripts that harvest multiple sources
- Scripts that harvest a single source
  - Harvest of OpenAlex (harvest\_openalex\_to\_ricgraph)
  - Harvest of Pure (harvest\_pure\_to\_ricgraph)
  - Harvest of data sets from Yoda-DataCite (harvest\_yoda\_datacite\_to\_ricgraph)
  - Harvest of Utrecht University staff pages (harvest\_uustaffpages\_to\_ricgraph)
  - Harvest of software from the Research Software Directory (harvest\_rsd\_to\_ricgraph)
- Order of running the harvest scripts
- How to make your own harvesting scripts

All code is documented and hints to use it can be found in the source files.

Return to main README.md file.

#### 5.1 Introduction to harvest scripts

One of the most useful features of Ricgraph is that it is possible to harvest sources that are important to a user or organization. By doing this, one is able to create a system that perfectly suits a certain information need of that person or organization. In creating harvest scripts, it is possible to harvest only that information that is relevant for a certain purpose.

E.g., one can harvest generally available sources such as OpenAlex. It is also possible to harvest sources that are specific for a certain organization. For example, one of the harvest scripts harvests the Utrecht University staff pages. These pages cannot be harvested by other organizations due to the privileges required. Also, it is possible to harvest a source that is internal to an organization.

Ricgraph can be installed on any internal or external accessible system according to your needs, so the data in Ricgraph is only accessible for persons of a certain organization, or for anyone.

Ricgraph provides a number of scripts for batch harvesting multiple sources with one script. These are in directory *harvest\_multiple\_sources*. Read more in the section that describes the scripts that harvest multiple sources. These scripts are based on the Ricgraph harvest scripts to harvest a single source. They are in directory *harvest*. Read more in the section that describes the scripts that harvest a single source.

Each of these scripts can be adapted to your needs, see their code. It is best to run harvest scripts in a specific order.

#### 5.2 Organization abbreviation

Ricgraph uses the term *organization abbreviation*. This is a string of a few letters that can be passed to some harvest scripts to determine for which organization such a script will harvest data. Examples are *UU* or *UMCU*. You will find this in keys in the Ricgraph initialization file. Examples are *organization\_name\_UU* or *organization\_name\_UMCU*. The general format is *key\_XXXX*, with XXXX the organization abbreviation.

If your organization abbreviation is not in the Ricgraph initialization file, feel free to add one. You can use any (short) string and pass it to a harvest script. You only need to insert keys (and values) for the organization(s) you are planning to harvest.

## 5.3 Scripts that harvest multiple sources

These bash scripts are in directory harvest\_multiple\_sources.

There are two general scripts to harvest multiple sources:

- *multiple\_harvest\_organization\_rsd\_yoda.sh*: harvests the Research Software Directory and Yoda.
- *multiple\_harvest\_organization.sh:* harvest Pure and OpenAlex. Then, it calls script *multiple\_harvest\_organization\_rsd\_yoda.sh.*

Usage:

./[script name].sh [options]

#### Options:

- -o, --organization [organization]
  - The organization to harvest. Specify the organization abbreviation.
- -e, --empty\_ricgraph [yes|no]

Whether to empty Ricgraph before harvesting the

Usage:

first organization. If absent, Ricgraph will not be emptied. harvest\_openalex\_to\_ricgraph.py [options]

-c, --python\_cmd [python interpreter]

The python interpreter to use. If absent, and a python virtual environment is used, that interpreter is used.

-p, --python\_path [python path]

The value for PYTHONPATH, the path to python libraries. If absent, the current directory is used.

-h, --help

Show this help text.

There are three wrapper scripts:

- *multiple\_harvest\_demo.sh*: calls script *multiple\_harvest\_organization\_rsd\_yoda.sh* with organization *UU*. These sources can be used to demonstrate Ricgraph, since these sources do not need a REST API key.
- *multiple\_harvest\_uu.sh*: calls *multiple\_harvest\_organization.sh* for organization *UU*, and also harvests the UU staff pages.
- *multiple\_harvest\_rik.sh*: a script that harvests the favorite sources of the author of Ricgraph.

You can use the Ricgraph Makefile to run these harvest scripts, e.g. to run *multiple\_harvest\_demo.sh*, execute command:

 ${\tt make run\_bash\_script bash\_script=harvest\_multiple\_sources/multiple\_harvest\_demo.sh}$ 

#### 5.4 Scripts that harvest a single source

These Python scripts are in directory harvest.

#### 5.4.1 Harvest of OpenAlex (harvest\_openalex\_to\_ricgraph)

To harvest OpenAlex, use the script *harvest\_openalex\_to\_ricgraph.py*.

usage:

Options:

--empty\_ricgraph <yes|no>

'yes': Ricgraph will be emptied before harvesting.

'no': Ricgraph will not be emptied before harvesting.

If this option is not present, the script will prompt the user what to do.

--organization <organization abbreviation>

Harvest data from organization <organization abbreviation>.

The organization abbreviations are specified in the Ricgraph ini file.

If this option is not present, the script will prompt the user what to do.

This script harvests OpenAlex Works, and by harvesting these Works, it also harvests OpenAlex Authors. This script needs the parameters *organiza-tion\_name\_XXXX*, *organization\_ror\_XXXX* and *openalex\_polite\_pool\_email* to be set in the Ricgraph initialization file. *XXXX* is your organization abbreviation.

There is a lot of data in OpenAlex, so your harvest may take a long time. You may reduce this by adjusting parameters at the start of the script. Look in the section "Parameters for harvesting persons and research outputs from OpenAlex": *OPENALEX\_RESOUT\_YEARS* and *OPENALEX\_MAX\_RECS\_TO\_HARVEST*.

#### 5.4.2 Harvest of Pure (harvest\_pure\_to\_ricgraph)

To harvest Pure, use the script *harvest\_pure\_to\_ricgraph.py*.

Usage: harvest\_pure\_to\_ricgraph.py [options]

Options:

--empty\_ricgraph <yes|no>

'yes': Ricgraph will be emptied before harvesting.
'no': Ricgraph will not be emptied before harvesting.

If this option is not present, the script will prompt the user what to do.

--organization <organization abbreviation>

Harvest data from organization <organization abbreviation>.

The organization abbreviations are specified in the Ricgraph ini file.

If this option is not present, the script will prompt the user what to do.

--harvest\_projects <yes|no>

'yes': projects will be harvested.

'no' (or any other answer): projects will not be harvested.

If this option is not present, projects will not be harvested, the script will not prompt the user.

With this script, you can harvest persons, organizations and research outputs. This script needs two parameters in the Ricgraph initialization file: the url to Pure in *pure\_url\_XXXX*, and the Pure API key in *pure\_api\_key\_XXXX*. XXXX is your organization abbreviation.

#### 5.4.2.1 Limit the amount of data to harvest from Pure

There is a lot of data in Pure, so your harvest may take a long time. You may reduce this by adjusting parameters at the start of the script. Look in the sections "Parameters for harvesting persons/organizations/research outputs from Pure". E.g., for research outputs you can adjust the years to harvest with the parameter *PURE\_RESOUT\_YEARS* and the maximum number of records to harvest with *PURE\_RESOUT\_MAX\_RECS\_TO\_HARVEST*.

#### 5.4.2.2 Pure READ and Pure CRUD API

Pure has two APIs, a READ and a CRUD API. The Pure READ API ("old" API) is only for reading data from Pure. The Pure CRUD API ("new" API) can be used to create, read, update and delete data in Pure (hence the name: CRUD). You do not need to specify which API you want to use, the script will be able to determine it for you (just include the API key in the initialization file). The

author recommends to use the READ API. You can use both of them to harvest data from Pure, but each of them has its own advantages and disadvantages:

- The Pure READ API has a number of filters, which allow to reduce data requested from Pure on the Pure server, thereby preventing this data to be sent to the computer which is running the Pure harvest script, and for the harvest script, to process all this data. E.g., the READ API has a filter for persons, so only active persons in Pure will be sent to the harvesting computer, thereby reducing the number of persons to process in the harvest script from all persons in Pure to only active persons in Pure. Another filter is the start and end publication year for research outputs. This makes it possible for the Pure harvest script to only process research outputs from a certain year, instead of all research outputs in Pure. This prevents potential memory problems.
- The Pure CRUD API allows the Pure administrator to specify which Pure fields are allowed to be sent from the Pure server to the computer that is running the Pure harvest script. This allows for only sending data that is requested. However, since this API is in development, a lot of the filters present in the READ API do not exist (yet), especially the filters mentioned in the previous bullet. That means, if you run the Pure harvest script, you might encounter memory problems while harvesting research outputs due to the number of research outputs in Pure, unless you set *PURE\_RESOUT\_MAX\_RECS\_TO\_HARVEST* in the Pure harvest script to some suitable value.

#### 5.4.2.3 Pure harvesting of projects

You can also harvest projects from Pure, if your organization uses them. You will need to use the Pure READ API, harvesting projects with the PURE CRUD API has not been implemented yet.

Projects will be connected to related persons, related organizations (the chair/department/faculty the person is from), related research outputs, and related projects (if any).

Note that it can happen that you may find a project that is connected to a *person-root* which only has a *PURE\_UUID\_PERS* and nothing else, or that you may find a project that is connected to a research output which only has a *PURE\_UUID\_RESOUT* and nothing else. This is probably caused by the Pure harvest script that (in its standard configuration) only harvests active persons, those are persons who are working at your organization at the time of harvest (so you will not have any information about persons that have worked on the project, e.g. PhDs or postdocs, since they may already have left your organization), and only harvests research outputs from 2020 onward, and your project may have research outputs from before 2020.

#### 5.4.3 Harvest of data sets from Yoda-DataCite (harvest\_yoda\_datacite\_to\_ricgraph)

To harvest data sets from the data repository Yoda (via DataCite), use the script harvest\_yoda\_datacite\_to\_ricgraph.py.

```
Usage:
```

harvest\_yoda\_datacite\_to\_ricgraph.py [options]

#### Options:

- --empty\_ricgraph <yes|no>
  - 'yes': Ricgraph will be emptied before harvesting.
  - 'no': Ricgraph will not be emptied before harvesting.
  - If this option is not present, the script will prompt the user what to do.

#### --organization <organization abbreviation>

Harvest data from organization <organization abbreviation>. The organization abbreviations are specified in the Ricgraph ini file.

If this option is not present, the script will prompt the user what to do.

This script can be used out of the box since it doesn't need an API key.

#### 5.4.4 Harvest of Utrecht University staff pages (harvest\_uustaffpages\_to\_ricgraph)

To harvest the Utrecht University staff pages, use the script harvest\_uustaffpages\_to\_ricgraph.py.

Usage: harvest\_uustaffpages\_to\_ricgraph.py [options]

Options:

--empty\_ricgraph <yes|no>

'yes': Ricgraph will be emptied before harvesting.

'no': Ricgraph will not be emptied before harvesting.

If this option is not present, the script will prompt the user what to do.

This script needs the parameter *uustaff\_url* to be set in the Ricgraph initialization file.

#### 5.4.5 Harvest of software from the Research Software Directory (harvest\_rsd\_to\_ricgraph)

To harvest software packages from the Research Software Directory, use the script *harvest\_rsd\_to\_ricgraph.py*.

Usage: harvest\_rsd\_to\_ricgraph.py [options]

#### Options:

--empty\_ricgraph <yes|no>

'yes': Ricgraph will be emptied before harvesting.

'no': Ricgraph will not be emptied before harvesting.

If this option is not present, the script will prompt the user

what to do.

--organization <organization abbreviation>

Harvest data from organization <organization abbreviation>. The organization abbreviations are specified in the Ricgraph ini file.

If this option is not present, the script will prompt the user what to do.

This script needs one parameter in the Ricgraph initialization file: *rsd\_organization\_XXXX. XXXX* is your organization abbreviation. It can be used out of the box since it doesn't need an API key.

# 5.5 Order of running the harvest scripts

The order of running the harvesting scripts does matter. The author harvests only records for Utrecht University and uses this order:

- harvest\_pure\_to\_ricgraph.py (since it has a lot of data which is mostly correct);
- harvest\_yoda\_datacite\_to\_ricgraph.py (not too much data, so harvest is fast, but it contains several data entry errors);
- 3. *harvest\_rsd\_to\_ricgraph.py* (not too much data);
- 4. *harvest\_openalex\_to\_ricgraph.py* (a lot of data from a number of sources);
- 5. harvest\_uustaffpages\_to\_ricgraph.py.

Best practice is to start with that source that has the most data. If you don't have an API key for Pure, one can best start with the harvest script for Ope-nAlex.

## 5.6 How to make your own harvesting scripts

For making your own harvesting scripts, refer to the code used in the harvest scripts described above, and read How to make your own harvesting scripts.

# 6 Ricgraph Explorer

Ricgraph provides an exploration tool, so users do not need to learn a graph query language. This tool is called *Ricgraph Explorer*. As it is a Python Flask application, it can be customized as needed. New queries (buttons) can be added, or the user interface can be modified to fit a certain use case, user group, or application area. The code can be found in directory *ric-graph\_explorer*.

In this documentation and on this page, we use the application area *research information*. Ricgraph Explorer has several pre-build queries tailored to this application area, each with its own button, for example:

- find a person, a (sub-)organization, a skill;
- when a person has been found, find its identities, skills, research results.

Ricgraph Explorer offers faceted navigation. That means, if a query results in a table with e.g. *journal articles, data sets*, and *software*, you can narrow down on one or more of these categories by checking or unchecking their corresponding checkbox. An alternative view uses tabbed navigation.

This page describes what you can do with Ricgraph Explorer. It does this by showing the flow through the web application by listing the buttons available. The text below lists these buttons. For some of these buttons, a more extensive description is given. After clicking a few buttons, and entering values in the fields provided, the user will get a Results page. Since there are many possible result pages, we only show one result page as example.

On this page, you can learn more about:

- Read about how to start Ricgraph Explorer.
- Working with Ricgraph Explorer.
- The Home page of Ricgraph Explorer.
- The Search page of Ricgraph Explorer.
- The Person options page of Ricgraph Explorer.

- The Organization options page of Ricgraph Explorer.
- The Results page of Ricgraph Explorer.
- Read about browsing Ricgraph.
- Learn how to show a privacy statement or privacy measures document.
- Video Ricgraph Explorer

Return to main README.md file.

## 6.1 How to start Ricgraph Explorer

Depending on how Ricgraph has been installed, there are various ways to start Ricgraph Explorer.

- If you have installed Ricgraph for a single user, you might need to start your graph database backend first:
  - If you use Neo4j Community Edition: this will be already running.
  - If you use Neo4j Desktop: read Start Neo4j Desktop.
  - If you use Memgraph: read Install and start Memgraph.
- Depending on how you have installed Ricgraph, Ricgraph Explorer may already be running. To check, in your web browser, go to http://127.0.0.1: 3030.
- If you get an error message, you will need to start Ricgraph Explorer:
  - Use the Ricgraph Makefile and execute command

make run\_ricgraph\_explorer

In your web browser, go to http://127.0.0.1:3030.

Run the ricgraph\_explorer.py script in directory ricgraph\_explorer. In your web browser, go to http://127.0.0.1:3030.

- If you have a domain name to access Ricgraph, for example www.ricgraphexample.com:
  - Open a web browser and go to https://www.ricgraph-example.com.

#### 6.2 Working with Ricgraph Explorer

#### 6.2.1 A usage flow through Ricgraph Explorer



Figure 6.1: Screenshots of a usage flow through Ricgraph Explorer. Some field values have been blurred for privacy reasons.

The figure above shows screenshots of web pages of Ricgraph Explorer for answering the research question "What are the research results of person A". The screenshot at the top left is the home page. After clicking "search for a person", Ricgraph Explorer shows a search page (top right). A user types a name, and the person options page is shown (bottom left). After clicking "show research results related to this person", the results page is shown (bottom right). In that page, the rows in the second table are (in this case) the journal article neighbors of the item in the first table (the person the user searched). This person also has other types of research results: book chapters, data sets, other contributions, books, reviews, and software (cf. row with orange rectangle, this is an example of the tabbed navigation). The "comment" column contains the titles of the journal articles. By clicking on an entry in the "value" column, in this case a DOI value, the user will go to this neighbor. Ricgraph Explorer will show a page with persons who have contributed to that journal article.

#### 6.2.2 Example research questions

In the figure in the previous section, after a click on a value in the "value" column in the bottom right result page, the user will get the persons who have contributed to that research result, as in figure (b) below.



Figure 6.2: Example research questions for Ricgraph.

Clicking "find persons that share any result types with this person" in the bottom left person option page (in the figure in the previous section) corresponds to figure (c), and clicking "show personal information related to this person" corresponds to figure (e).

The research question "What are the research results of person A" from the previous section corresponds to figure (a).

## 6.3 Home page

The figure below shows part of the home page.

Ricgraph Explorer      Migraph - Research in context graph, Ricgraph - Research in context graph, enables trelations between these items.  Ricgraph can store many types of items into a single graph (ne below for the sources of the items in this instance of Ricgraph items, relations that are not present in any of the separate so You can use Ricgraph Explorer to explore Ricgraph. There are	Advanced search Broad search BEST API doc the exploration of researchers, teams, their results, collaborations, skills, projects, and the twork). These items can be obtained from various systems and from multiple organizations (see Biggraph facilitates reasoning about these items because it infers new relations between urce systems. It is flexible and extensible, and can be adapted to new application areas. various methods to start exploring:
search for a person	
search for a (sub-)organization	
search for a skill, expertise area or research area	
search for anything (broad search)	
advanced search	•
More information	
<ul> <li>or a gentle introduction in Ricgraph, please read the reference research in context from various systems. SoftwareX, 26(1017 and source code can be found in the GitHub repository <u>https:</u></li> </ul>	:e publication: Nik D.1. Janssen (2024). Riggraph: A flexible and extensible graph to explore 36). <u>https://doi.org/10.1016/j.softx.2024.101736</u> . Extensive documentation, publications, videos //github.com/UtrechUniversity/riggraph.
Ctatictice	

Figure 6.3: Home page of Ricgraph Explorer.

The home page lets the user choose between various methods to explore Ricgraph:

- Button search for a person.
  - in the search page that is shown after you have clicked this button, to search, type a full name or substring of a name.
  - if there is more than one result, select one person.
  - this person is shown on the person options page for further exploration.
- Button search for a (sub-)organization.
  - in the search page that is shown after you have clicked this button, to search, type a full organization name or substring of an organization name.

- if there is more than one result, select one organization.
- this organization is shown on the organization options page for further exploration.
- Button search for a skill, expertise area or research area [only available if you have nodes of category competence in Ricgraph]:
  - in the search page that is shown after you have clicked this button, to search, type a skill, expertise area or research area, or substring of one of these.
  - if there is more than one result, select one.
  - the results are shown on the results page.
- Button search for anything (broad search) or button advanced search.
  - in the search page that is shown after you have clicked this button, type something to search, the advanced search is a case-sensitive exact match search on one or more of the Ricgraph fields *name*, *cat-egory*, or *value*.
  - if there is more than one result, select one.
  - depending on the type of result:
    - \* if the result is a person, the result is shown on the person options page for further exploration.
    - \* if the result is an organization, the result is shown on the organization options page for further exploration.
    - \* for all other results, the results are shown on the results page.

In the yellow top bar, there are more buttons:

- Button Home. This button returns the user to the home page.
- Button Advanced search. This leads to the advanced search page, see section search page.
- Button *Broad search*. This leads to the broad search page, see section search page.
- Button *REST API doc.* This leads to the documentation and try out page of the Ricgraph REST API.

# 6.4 Search page

There are two search pages:

- A broad search page, having one input field. This search is a case-insensitive search.
- An advanced search page, having three input fields. This search is a casesensitive exact match search on one or more of the Ricgraph fields *name*, *category*, or *value*.

The figures below show these search pages.

>	꿇 Ricgraph Explorer	Home	Advanced search	Broad search	
Type yo	ur search string:				
Please s	specify how you like to view your results:				
O per	son_view: only show relevant columns, results ails_view: show all columns, research results a	are presented in a tabbea re presented in a table witl	f format 🕕 h <i>facets</i> 😗		
_	-				
	search				

Figure 6.4: Broad search page of Ricgraph Explorer.

	꿇 Ricgraph Explorer	Home	Advanced search	Broad search	
Search	for a value in Ricgraph field <i>name</i> :				
Search	for a value in Ricgraph field category.				
Search	for a value in Ricgraph field value:				
Search These	i for a value in Ricgraph field <i>value</i> : fields are case-sensitive and use exact match sea	rch. If you enter values in	more than one field, these	fields are combined using	g AND.
Search These Please	I for a value in Ricgraph field <i>value</i> : fields are case-sensitive and use exact match sea specify how you like to view your results:	rch. If you enter values in	more than one field, these	fields are combined using	g AND.
Search These Please <i>pe</i>	r for a value in Ricgraph field value: fields are case-sensitive and use exact match sea specify how you like to view your results: <i>rson_view.</i> only show relevant columns, results a	rch. If you enter values in re presented in a <i>tabbed</i> f	more than one field, these prmat 🜖	fields are combined using	g AND.
Search These Please Ope Ode	r for a value in Ricgraph field value: fields are case-sensitive and use exact match sea specify how you like to view your results: rson_view: only show relevant columns, results are tails, view. show all columns, research results are	rch. If you enter values in re presented in a <i>tabbed</i> f presented in a table with	more than one field, these ormat <mark>()</mark> facets <b>(</b> )	fields are combined using	g AND.

Figure 6.5: Advanced search page of Ricgraph Explorer.

There are two methods for viewing the results:

- *person\_view*: only show relevant columns, results are presented in a tabbed format. Tables have fewer columns to reduce information overload.
- *details\_view*: show all columns, research results are presented in a table with facets. This view shows all columns .

You only get on this page if the result of your search is a person. The figure

#### 6.5 Person options page

below shows the person options page.



Figure 6.6: Person options page of Ricgraph Explorer.

You can use one of these exploration options:

- Button show personal information related to this person.
- Button show organizations related to this person.
- Button show research results related to this person.
- Button show everything except identities related to this person.
- Button show all information related to this person.
- Button find persons that share any research result types with this person.
- Button find persons that share a specific research result type with this person.
- You will need to enter a research result type from a drop-down list.
- Button find organizations that this person collaborates with.

This button gives an overview of organizations that this person collaborates with. A person *X* from organization *A* collaborates with a person *Y* from organization *B* if *X* and *Y* have both contributed to the same research result.

• Button find information harvested from other source systems, not present in this source system.

You will need to enter a source system from a drop-down list. Next, this button gives an overview of information that can be added to the source system entered, based on the information harvested from other source systems. The process of improving or enhancing information in a source system is called "enriching" that source system. This is only possible if you have harvested more than one source system.

• Button find the overlap in source systems for the neighbor nodes of this node.

In case more than one source systems have been harvested, and the information in Ricgraph for the neighbors of this node have originated from more than one source system, clicking this button will show you from which ones.

For each of these buttons, the results are shown on the results page.

### 6.6 Organization options page

You only get on this page if the result of your search is an organization. The figure below shows the organization options page.

Ricgraph Explorer found item:							
name 🛦 🔻	category ⊾ ▼	value 🔺 🔻	comment∡▼	year 🛛 🔻	url_main	url_other	
ORGANIZATION_NAME	organization	Utrecht University			url_main link		
Vhat would you like to see from this organization? show persons related to this organization							
show all information r	elated to this organization	1					
Advanced information	related to this orga	nization					
Advanced information More information abo	related to this orga out persons or their m rom all persons in this org	nization esults in this organiza ganization	ation.				
Advanced information More information abo find research results f find skills from a find any information fro	related to this orga out persons or their r rom all persons in this org Il persons in this organizz om persons or their result	nization esults in this organiza ganization tion	ation.				
Advanced information More information abb find research results f find skills from a find any information fro By using the fields below, Search for persons or resi	related to this orga but persons or their rr rom all persons in this organizz in persons or their result you can choose what you you can choose what you	nization esults in this organiza ganization sin this organization would like to see from th	ation.	n this organizat	ion. You can use (	one or both fields.	
Advanced information More information abb find research results f find skills from a find any information fro By using the fields below, Search for persons or results Search for persons or results	related to this orga but persons or their r rrom all persons in this organizz in persons in this organizz an persons or their result and the source of their result you can choose what you you can choose what you uts using field <i>name</i>	nization esults in this organization sin this organization would like to see from th	ation. ne persons or their results	n this organizat	ion. You can use e	one or both fields.	

Figure 6.7: Organization options page of Ricgraph Explorer.

You can use one of these buttons:

- Button show persons related to this organization.
- Button show all information related to this organization.
- Button find research results from all persons in this organization.
- Button find skills from all persons in this organization.
- Button find any information from persons or their results in this organization.
- Button find specific information.

You will need to enter a value for Ricgraph fields *name* or *category* from a drop-down list. Next, this button gives an overview of the persons or their results in this organization.

For all of these buttons the results are shown on the results page.

#### 6.7 Results page

The results page will look different depending on what results are shown. The figure below shows an example of a part of the skills in an organization.

Ricgrap	ျ Explorer	Home	Advan	ced search	Bro	ad search	
I for finding infor	nation about the	persons or their results of	this organizatio	n:			
Ű	category <b>▲</b> ▼	value 🔺	comment <b>▲</b> ▼	year▲▼	url_m	ain	url_other
NAME	organization	Utrecht University			<u>url m</u>	ain link	
category A T	value	zation:		comment . T	This tal	url main	first 250 row
competence	NVivo					url main link	
competence	Qualitative Analy	Qualitative Analysis				url main link	
competence	Qualitative Meth	Qualitative Methods and Research Design				url main link	
competence	Interdisciplinary	Education and Research				url main link	
competence	Writing					url main link	
competence	Training, Support	and Advice				url main link	
competence	Presenting and H	osting Events				url main link	
competence	Facilitate Worksh	ops				url main link	
competence	Qualitative Resea	irch				url main link	
competence	Diagnostics of De	velopment and Parenting Prob	lems			url main link	
competence	Supervising Stud	ents in the Context of Tutoring				url main link	
	Ricgraph for finding information NAME "competence" ltt Response for the second competence competence competence competence competence competence competence competence competence competence competence competence competence competence competence	Ricgraph Explorer for finding information about the i category + organization  category + value + competence* information  competence competenc	Home           Information about the persons or their results of the persons of the per	Ricgraph Explorer         Home         Advantage           for finding information about the persons or their results of this organization         commentage         commentage           NAME         organization         Utrecht University         commentage           "competence" items of this organization:         commentage         commentage           "competence" items of this organization:         competence         Make +           competence         NAVe         Dualitative Analysis           competence         Dualitative Methods and Research Design         competence           competence         Making Support and Advice         competence           competence         Training. Support and Advice         competence           competence         Excertific Research         competence           competence         Excentres of Development and Persearch Tenoring <td>Interfacional about the persons or their results of this organization:         Advanced search           for finding information about the persons or their results of this organization:         comment at         year at           NAME         organization         Utretht University         comment at         year at           "competence" items of this organization:         comment at         year at         comment at         year at           "competence" items of this organization:         competence         NAVe         comment at         comment at           competence         NAVe         competence         Qualitative Analogis         comment at         comment at           competence         Qualitative Methods and Research         competence         competence</td> <td>Ricgraph Explorer         Home         Advanced search         Browne           for finding information about the persons or their results of this organization:         Image: Commentant Search         Image: Commentant Search<td>Riggraph Explorer         Home         Advanced search         Broad search           for finding information about the persons or their results of this organization:         ufuminit         ufuminit           NAME         organization         ufuminit         ufuminit           NAME         organization         ufuminit         ufuminit           'competence'         this organization:         This table shows the persons of this organization:         This table shows the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons the persons of this organization:         This table shows the persons of this organization:           competence         Qualitative Acadesis         Competence         Lut main link persons of the perso</td></td>	Interfacional about the persons or their results of this organization:         Advanced search           for finding information about the persons or their results of this organization:         comment at         year at           NAME         organization         Utretht University         comment at         year at           "competence" items of this organization:         comment at         year at         comment at         year at           "competence" items of this organization:         competence         NAVe         comment at         comment at           competence         NAVe         competence         Qualitative Analogis         comment at         comment at           competence         Qualitative Methods and Research         competence         competence	Ricgraph Explorer         Home         Advanced search         Browne           for finding information about the persons or their results of this organization:         Image: Commentant Search         Image: Commentant Search <td>Riggraph Explorer         Home         Advanced search         Broad search           for finding information about the persons or their results of this organization:         ufuminit         ufuminit           NAME         organization         ufuminit         ufuminit           NAME         organization         ufuminit         ufuminit           'competence'         this organization:         This table shows the persons of this organization:         This table shows the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons the persons of this organization:         This table shows the persons of this organization:           competence         Qualitative Acadesis         Competence         Lut main link persons of the perso</td>	Riggraph Explorer         Home         Advanced search         Broad search           for finding information about the persons or their results of this organization:         ufuminit         ufuminit           NAME         organization         ufuminit         ufuminit           NAME         organization         ufuminit         ufuminit           'competence'         this organization:         This table shows the persons of this organization:         This table shows the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons of this organization:         This table shows the persons the persons the persons of this organization:         This table shows the persons of this organization:           competence         Qualitative Acadesis         Competence         Lut main link persons of the perso

Figure 6.8: Results page of Ricgraph Explorer.

#### 6.8 Browsing Ricgraph

You can browse the graph that Ricgraph has obtained by harvesting source systems by clicking on a value in the *value* column. In the figure in section **Results page**, the search was started with organization *Utrecht University* (in the first table). The *SKILL* neighbors of *Utrecht University* are listed in the second table. By clicking on e.g. *NVivo*, you will traverse to the *NVivo* node, and you will see the neighbors of this node. This can be repeated as desired. It is also possible to start a new search using the buttons *Home*, *Advanced search*, or *Broad search* in the yellow title bar.

# 6.9 Showing a privacy statement or privacy measures document

Ricgraph Explorer offers the possibility to show a privacy statement or privacy measures document (or both) if desired for your Ricgraph production environment. Placeholders have been provided in the files *pri-vacy\_statement\_placeholder.pdf* and *privacy\_measures\_placeholder.pdf* in directory *ricgraph\_explorer/static*.

How to use a privacy statement in Ricgraph Explorer?

- Create a privacy statement that fits your situation.
- Use the filename privacy\_statement.[extension of your text editor].
- Convert this file to a pdf file and use filename privacy\_statement.pdf.
- Put it in directory ricgraph\_explorer/static.
- In Ricgraph Explorer, in the footer you will find a link to this file.

How to use a privacy measures document in Ricgraph Explorer?

- Create a privacy measures document that fits your situation.
- Use the filename privacy\_measures.[extension of your text editor].
- Convert this file to a pdf file and use filename privacy\_measures.pdf.
- Put it in directory *ricgraph\_explorer/static*.
- In Ricgraph Explorer, in the footer you will find a link to this file.

#### 6.10 Video Ricgraph Explorer

The video Ricgraph Explorer (2m20s) (click to view or download) gives an impression of how a user-centric interface may look like. This video uses Ricgraph Explorer from March 2023 (the current version has much more possibilities). In this video, we look up a specific person with his last name. As can be observed, there are three *FULL\_NAME* nodes for this person, each with a different spelling, from four different sources we have harvested. If we click on one of them, we observe that we have found a lot of information about this person. The first table shows the node used for the search. The second table displays the IDs of the person found, connected to the *person-root* node, and the third table shows all other nodes connected to the *person-root* node. This includes research outputs like publications, data sets and software, as well as the (sub)organization this person works. We can sort columns, and we can use faceted navigation (i.e. filter on *name* or *category* nodes).

This is documentation for version 2.12 of Ricgraph – Research in context graph. It is a pdf of docs.ricgraph.eu. For more information about Ricgraph, Ricgraph Explorer, and the Ricgraph REST API, please read the reference publication doi.org/10.1016/j.softx.2024.101736, visit the website www.ricgraph.eu, or go to the GitHub repository github.com/UtrechtUniversity/ricgraph.

# 7 Ricgraph publications, presentations, use, and mentions

This page contains the following sections:	When		Authors	Title and link to pdf	Weblink
<ul> <li>Ricgraph publications</li> <li>Ricgraph presentations</li> <li>Ricgraph newsletters</li> <li>Ricgraph projects with students</li> <li>Ricgraph use</li> <li>Ricgraph mentions</li> <li>Arjan Sieverink presents Ricgraph at Pure International Conference 2023</li> </ul>	October 2023	26,	Arjan Sieverink	Ricgraph: Showcasing research in context using Pure and other sources, Ac- cepted abstract for Pure Inter- national Con- ference 2023,	DOI 10.5281/zen- odo.10057998
Please let us know if you publish, present, use, or mention Ricgraph, then we can add that to this page.	January 31, 1	2024	Rik D.T. Janssen	Dubrovník, Croatia Ricgraph: A Flexible and	DOI 10.2139/ssrn.4712466
Return to main README.md file.				Graph to Ex- plore Research in Context from Various Sys- tems, Submit- ted to SoftwareX	
7.1 Ricgraph publications	April 22, 20	24	Rik D.T. Janssen	[preprint] Ricgraph: A flexible and extensible	DOI 10.1016/j.softx.2024.101736
If you publish about Ricgraph, please let us know so that we can add it to this page.	ŧ0			graph to ex- plore research in context from various systems, SoftwareX, 26(101736)	

### 7.2 Ricgraph presentations

7.2 Ricgraph presentations	When		Who		Where	Title and link to pdf	Weblink
	March 2023	30,	Rik Jansser	D.T.	Presentation for FAIR Re- search IT innovation lunch at Utrecht University, Utrecht, NL	link to pdf Ricgraph – Research in context graph (pre- sentation was a con- densed version of the pre- sentation	
	April 2023	26,	Rik Jansser	D.T.	Presentation for PID graph pilot meeting at SURF, Utrecht, NL	of May 10, 2023) Ricgraph – Research in context graph (pre- sentation was a con- densed version of the pre- sentation of May 10,	
	May 2023	10,	Rik Jansser	D.T. 1	Presentation for PID graph workshop at SURF,	2023) Ricgraph - Research in context graph	
If you give a presentation about Ricgraph, please let us know so that we can add it to this page.	October 2023	26,	Arjan erink	Siev-	Utrecht, NL Presentation for Pure In- ternational Confer- ence 2023,	Ricgraph: Showcasing research in context using Pure	DOI 10.5281/zen- odo.10057998

This is documentation for version 2.12 of Ricgraph – Research in context graph. It is a pdf of docs.ricgraph.eu. For please read the reference publication doi.org/10.1016/j.softx.2024.101736, visit the website www.ricgraph.eu, or go сp 'r y gr Croatia sources

(presenta-

tion)

#### 7.3 Ricgraph newsletters

These are the newsletters that have been sent out to our newsletter subscribers. To subscribe to the newsletter email list, go to Ricgraph Contact.

Date	Title
April 2024	Ricgraph newsletter April 2024. Journal article - Ricgraph: A flexi- ble and extensible graph to explore research in context
July 2024	Ricgraph newsletter July 2024

When	Description		
March - August, 2023 June 2023	NWO PID graph pilot Project report NWO NWOpen-API		
July 2023 July 2023 February - June, 2025	Elsevier Data Monitor Elsevier Grant Award API University of Applied Sciences Utrecht innovation project		

#### 7.4 Ricgraph projects with students

If you use Ricgraph in a project with students, please let us know so that we can add it to this page.

When	Education Insti- tute	Who	Description
February - June, 2025	University of Applied Sci- ences, Utrecht	Six 3rd year stu- dents	Using AI and LLM techniques to cluster and visualize large amounts of research infor- mation.

#### 7.6 Ricgraph mentions

#### 7.5 Ricgraph use

If you use Ricgraph, please let us know so that we can add it to this page.

If you mention Ricgraph, please let us know so that we can add it to this page.

This is documentation for version 2.12 of Ricgraph – Research in context graph. It is a pdf of docs.ricgraph.eu. For more information about Ricgraph, Ricgraph Explorer, and the Ricgraph REST API, please read the reference publication doi.org/10.1016/j.softx.2024.101736, visit the website www.ricgraph.eu, or go to the GitHub repository github.com/UtrechtUniversity/ricgraph.

When	Where	Title and link to pdf	Weblink
January 25, 2023	SURF communi- ties, blog	Ricgraph – Research in context graph	link
May 2, 2023	FAIR Research IT newsletter	Op ontdekking gaan is avon- tuurlijk, maar vaak tiidrovend	
May 12, 2023	SURF communi- ties, blog	PID Graphs, the road to data integration and discovery	link
October 26, 2023	RDA Open Sci- ence Graphs For FAIR Data Interest Group	Ricgraph - Research in context graph	link
February 1, 2024	RDA Open Sci- ence Graphs For FAIR Data Interest Group	Ricgraph: A Flexible and Extensible Graph to Ex- plore Research in Context from Various Systems	link
March, 2024	Elsevier Re- search Intel- ligence on LinkedIn	Unveil Ricgraph, a groundbreak- ing tool reshap- ing research data into a dynamic graph	link
March 20, 2024	Elsevier on Twitter/X	Introducing Ricgraph: rev- olutionizing research data into dynamic graphs	link
April 24, 2024	RDA Open Sci-	Ricgraph: A	link

### 7.7 Arjan Sieverink presents Ricgraph at Pure International Conference 2023

In the video of the presentation at the Pure International Conference 2023, Dubrovnik, Croatia (19m27s) (click to view or download), Arjan Sieverink presents the abstract *Ricgraph: Showcasing research in context using Pure and other sources* (see the Ricgraph publications section).

This is documentation for **PMS** fon 2.12**GT app** for the Ricgraph REST API, please read the reference **proj** control of the Ricgraph REST API, visit the website www.ricgraph.eu, or go to the GitHub repository github.com/UtrechtUniversity/ricgraph.

Interest Group graph to explore research

# 8 Ricgraph as a server on Linux

This page describes how to install and run Ricgraph in a multi-user environment on Linux. *Multi-user environment* means that you install Ricgraph on a Linux (virtual) machine, and that various persons can log on to that machine, each with his own user id and password. Each person will be able to use Ricgraph by using a web link in their web browser. For other Ricgraph install options start reading at Install and configure Ricgraph for a single user.

The reason that a Linux multi-user environment for Ricgraph is different from installing and using Ricgraph on your own user id, is that you will need to run the graph database backend and Ricgraph Explorer as a system user instead of running it using your own user id. In case you run Ricgraph with your own user id, you will be the only user able to use it. In case other persons on that same machine would like to use Ricgraph, they have to install it for themselves. By installing Ricgraph as a server, as described on this page, Ricgraph will be started automatically when your machine boots, and it can be used by any user on that machine.

To install and run Ricgraph in a multi-user environment, read Fast and recommended way to install Ricgraph as a server.

On this page, you can find:

- Fast and recommended way to install Ricgraph as a server
- Run Ricgraph scripts from the command line or as a cronjob
- Use a service unit file to run Ricgraph Explorer and the Ricgraph REST API
- Use Apache, WSGI, and ASGI to make Ricgraph Explorer and the Ricgraph REST API accessible from outside your virtual machine
- Use Nginx, WSGI, and ASGI to make Ricgraph Explorer and the Ricgraph REST API accessible from outside your virtual machine
- How to install Ricgraph and Ricgraph Explorer on SURF Research Cloud
- Steps to take to install Ricgraph as a server by hand

Return to main README.md file.

# 8.1 Fast and recommended way to install Ricgraph as a server

To follow this procedure, you need to be able to change to user root.

- 1. Check the requirements.
- 2. Login as user root.
  - sudo bash
- 3. Get the most recent Ricgraph Makefile. Type:
  - cd

44

wget https://raw.githubusercontent.com/UtrechtUniversity/ricgraph/main/Makefile

Read more at Ricgraph Makefile.

4. Install Neo4j Community Edition. Type:

make install\_enable\_neo4j\_community

On success, the Makefile will print installed successfully.

5. Download and install Ricgraph in system directory /opt.

make install\_ricgraph\_server

On success, the Makefile will print installed successfully.

6. Harvest two source systems in Ricgraph:

cd /opt/ricgraph\_venv
make run\_bash\_script

This will harvest two source systems, the data repository Yoda and the Research Software Directory. It will print a lot of output, and it will take a few minutes. When ready, it will print *Done*.

To read more about harvesting data, see Ricgraph harvest scripts. To read more about writing harvesting scripts, see Ricgraph script writing.

7. Start Ricgraph Explorer to browse the information harvested:

cd /opt/ricgraph\_venv
make run\_ricgraph\_explorer

The Makefile will tell you to go to your web browser, and go to http://127. 0.0.1:3030. Read more at Ricgraph Explorer. For the Ricgraph REST API, read more on the Ricgraph REST API page.

8. Optional: use a service unit file to run Ricgraph Explorer and the Ricgraph REST API. Type:

make install\_enable\_ricgraphexplorer\_restapi

On success, the Makefile will print *installed successfully*. Read more at Use a service unit file to run Ricgraph Explorer and the Ricgraph REST API.

9. Optional and possibly dangerous: use Apache or Nginx webserver, WSGI, and ASGI to make Ricgraph Explorer and the Ricgraph REST API accessible from outside your virtual machine. Read more at Use Apache.... or at Use Nginx.... On success, the Makefile will print *installed successfully*.

10. Exit from user *root*.

If everything succeeded, you are done installing Ricgraph as a server. If not, sections Steps to take to install Ricgraph as a server by hand or Install and start Neo4j Community Edition graph database backend may help in finding solutions.

# 8.2 Run Ricgraph scripts from the command line or as a cronjob

## 8.2.1 Using the Makefile

The Ricgraph Makefile can also be used to execute a Python script or a bash script. Such a script can be used to harvest the sources specific to your organization. The Makefile provides a command make run\_python\_script, to run any Ricgraph Python script. You will need to add the name of the script to run using the Makefile command line parameter *python\_script*, e.g.

make run\_python\_script python\_script=[path]/[python script]

There is a similar command for running bash scripts:

make run\_bash\_script bash\_script=[path]/[bash script]

Both make commands execute the script and the output will appear both on your screen and in a file. The Makefile will tell you the name of this log file.

Examples of commands you can use are:

• harvest from the Research Software Directory:

make run\_python\_script python\_script=harvest/harvest\_rsd\_to\_ricgraph.py

• harvest two sources without needing any keys or configuration:

make run\_bash\_script

• run Ricgraph Explorer:

#### 8.2.2 In case you have installed Ricgraph as a server

After following the steps in Create a Python virtual environment and install Ricgraph in it, it is possible to run Ricgraph from the command line or as a cronjob. To be able to run these scripts you need to be user *ricqraph* and group *ricgraph*. You can do this by using user *ricgraph* in your crontab file (e.g. in */etc/crontab*), or by using the command

#### sudo su - ricgraph

If you are finished with these commands, exit from user *ricgraph*.

Examples of commands you can use are:

- harvest from the Research Software Directory:
- cd /opt/ricgraph\_venv/harvest; ../bin/python harvest\_rsd\_to\_ricgraph.py
- harvest two sources without needing any keys or configuration:
- run Ricgraph Explorer:

#### 8.2.3 In case you have installed Ricgraph for a single user

After following the steps in Create a Python virtual environment and install Ricgraph in it, it is possible to run Ricgraph from the command line. You do not need to be user *ricgraph* and group *ricgraph*. The following assumes your Python virtual environment is in your Linux home directory \$HOME.

Examples of commands you can use are:

- harvest from the Research Software Directory:
- cd \$HOME/ricgraph\_venv/harvest; ../bin/python harvest\_rsd\_to\_ricgraph.py
- harvest all your favorite sources:

- cd \$HOME/ricgraph\_venv/harvest\_multiple\_sources; ./multiple\_harvest\_demo.sh
- run Ricgraph Explorer:

cd \$HOME/ricgraph\_venv/ricgraph\_explorer; ../bin/python ricgraph\_explorer.py

# 8.3 Use a service unit file to run Ricgraph Explorer and the **Ricgraph REST API**

Using a service unit file to run Ricgraph Explorer is very useful if you would like to set up a virtual machine that you want to use as a demo server, or if you would like to use the Ricgraph REST API. After the steps in this section, Ricgraph Explorer and the Ricgraph REST API are run automatically at the start of the virtual machine, so you can immediately start giving the demo.

For comparison, if you had installed the graph database backend and Ricgraph for a single user, as described in the documentation describing the incd /opt/ricgraph\_venv/harvest\_multiple\_sources; ./multiple\_harvest\_demo.shstallation and configuration of Ricgraph for a single user, after the start of the virtual machine, you would need to start the graph database backend, the virtual environment, and *ricgraph* explorer.py by hand.

cd /opt/ricgraph\_venv/ricgraph\_explorer; ../bin/python ricgraph\_explorer.py To use a service unit file, you can either use the Ricgraph Makefile and execute command:

make install\_enable\_ricgraphexplorer\_restapi

or follow the steps below.

Using a service unit file will not expose Ricgraph Explorer, the Ricgraph REST API, and Ricgraph data to the outside world. All data will only be accessible in the virtual machine.

- · Follow the steps in Create a Python virtual environment and install Ricgraph in it.
- Login as user root.

• Install the *Ricgraph Explorer* service unit file: copy file *ricgraph\_server\_config/ricgraph\_explorer\_gunicorn.service* to /etc/systemd/system, type:

cp /opt/ricgraph\_venv/ricgraph\_server\_config/ricgraph\_explorer\_gunicorn

#### Make it run by typing:

systemctl enable ricgraph\_explorer\_gunicorn.service
systemctl start ricgraph\_explorer\_gunicorn.service

Check the log for any errors, use one of:

systemctl -l status ricgraph\_explorer\_gunicorn.service
journalctl -u ricgraph\_explorer\_gunicorn.service

- Exit from user root.
- Now you can use Ricgraph Explorer by typing http://localhost:3030 in your web browser (i.e., the web browser of the virtual machine). You can use the Ricgraph REST API by using the path http://localhost:3030/api followed by a REST API endpoint.

## 8.4 Use Apache, WSGI, and ASGI to make Ricgraph Explorer and the Ricgraph REST API accessible from outside your virtual machine

#### 8.4.1 Introduction Apache webserver

**Ricgraph Explorer** is written in Flask, a framework for Python to build web interfaces. Flask contains a development web server, and if you start Ricgraph Explorer by typing *ricgraph\_explorer.py*, it will be started using that development web server. As this development web server is sufficient for development and demoing, it is certainly *not* sufficient for exposing Ricgraph data to the outside world (that is, to users outside your own virtual machine). The same holds for the Ricgraph REST API. For this, you will need a web server and a WSGI environment. For the REST API, you will need an ASGI environment. This section describes how to do that with Apache and *gunicorn*. Note that the example configuration file for Apache exposes Ricgraph Explorer to the outside world on a http (unenserviced) to fine the former of authentication. Certainly, this is not the way to do it. At least you should expose Ricgraph Explorer and the REST API using a https (encrypted) connection, possibly with additional authentication.

Therefore, the configuration file provided is an example for further development. There is no example code for a https connection, nor for authentication, nor for automatically obtaining and renewing SSL certificates, because these are specific to a certain situation (such as your external IP address, hostname, web server, domain name, SSL certificate provider, authentication source, etc.). So only expose Ricgraph Explorer, the Ricgraph REST API, and the data in Ricgraph to the outside world if you have considered these subjects, and have made an informed decision what is best for your situation.

To prevent accidental exposure of Ricgraph Explorer, the REST API, and the data in Ricgraph to the outside world, you will have to modify the Apache configuration file. You need to make a small modification to make it work. How to do this is described in the comments at the start of the configuration file.

Note that it is also possible to use Nginx as a webserver. If you are using SURF Research Cloud, you will need to use Nginx.

To use Apache c.s., you can either use the Ricgraph Makefile and execute command:

make prepare\_webserver\_apache

or follow the steps below.

### 8.4.2 Installation Apache

Note that different Linux editions use different paths. In the steps below, path names from OpenSUSE Leap are used. Please adapt them to you own

Linux edition:

- OpenSUSE Leap: apache2 and /etc/apache2/vhosts.d
- Ubuntu: apache2 and /etc/apache/sites-available
- Fedora: httpd and /etc/httpd/conf.d

Using Apache, WSGI, and ASGI will expose Ricgraph Explorer, the Ricgraph REST API, and Ricgraph data to the outside world.

- Follow the steps in Create a Python virtual environment and install Ricgraph in it.
- Login as user root.
- Make sure Apache has been installed.
- *Gunicorn* has already been installed when you installed the Python requirements.
- Enable two Apache modules (they have already been installed when you installed Apache):

a2enmod mod\_proxy a2enmod mod\_proxy\_http

• Install the Apache *Ricgraph Explorer* configuration file: copy file *ricgraph\_server\_config/ricgraph\_explorer.conf-apache* to /etc/apache2/vhosts.d, type:

cp /opt/ricgraph\_venv/ricgraph\_server\_config/ricgraph\_explorer.conf-apache /etc/apache chmod 600 /etc/apache2/vhosts.d/ricgraph\_explorer.conf-apache

# 8.4.3 Post-install steps Apache

- Login as user root.
- Move the Apache Ricgraph Explorer configuration file to its final location:

mv /etc/apache2/sites-available/ricgraph\_explorer.conf-apache /etc/apache2/sites ln -s /etc/apache2/sites-enabled/ricgraph\_explorer.conf /etc/apache2/sites-avail

Change *ricgraph\_explorer.conf* in such a way it fits your situation. Make the modification to *ricgraph\_explorer.conf* as described in the comments at the start of *ricgraph\_explorer.conf*. Test the result.

• Make it run by typing:

systemctl enable apache2.service
systemctl start apache2.service

Check the log for any errors, use one of:

systemctl -l status apache2.service
journalctl -u apache2.service

- Exit from user root.
- Now you can use Ricgraph Explorer from inside your virtual machine by typing http://localhost in your web browser in the virtual machine, or from outside your virtual machine by going to http://[your IP address] or http://[your hostname].
- You can use the Ricgraph REST API from inside your virtual machine by using the path http://localhost:3030/api followed by a REST API endpoint, or from outside your virtual machine by using the path http://[your IP address/api] or http://[your hostname]/api, both fol-/elowed by a REST API endpoint.

8.5 Use Nginx, WSGI, and ASGI to make Ricgraph Explorer and the Ricgraph REST API accessible from outside your virtual machine

8.5.1 Introduction Nginx webserver

mv /etc/apache2/vhosts.d/ricgraph\_explorer.conf-apache /etc/apache2/vhosts.d/ricgraph\_explorer.conf
Please read the introduction of section Use Apache, WSGI, and ASGI to make
Ricgraph Explorer and the Ricgraph REST API accessible from outside your

This is documentation for version 2.12 of Ricgraph – Research in context graph. It is a pdf of docs.ricgraph.eu. For more information about Ricgraph, Ricgraph Explorer, and the Ricgraph REST API, please read the reference publication doi.org/10.1016/j.softx.2024.101736, visit the website www.ricgraph.eu, or go to the GitHub repository github.com/UtrechtUniversity/ricgraph.

virtual machine. The same explanation and words of caution as for using Apache as a webserver hold for Nginx as a webserver.

To prevent accidental exposure of Ricgraph Explorer, the REST API, and the data in Ricgraph to the outside world, you will have to modify the Nginx configuration file. You need to make a small modification to make it work. How to do this is described in the comments at the start of the configuration file.

Note that it is also possible to use Apache as a webserver.

To use Nginx c.s., you can either use the Ricgraph Makefile and execute command:

make prepare\_webserver\_nginx

or follow the steps below.

#### 8.5.2 Installation Nginx

Note that different Linux editions use different paths. In the steps below, path names from OpenSUSE Leap are used. Please adapt them to you own Linux edition:

- OpenSUSE Leap: /etc/nginx/vhosts.d
- Ubuntu: /etc/nginx/sites-available
- Fedora: /etc/nginx/conf.d

Using Nginx, WSGI, and ASGI will expose Ricgraph Explorer, the Ricgraph REST API, and Ricgraph data to the outside world.

- Follow the steps in Create a Python virtual environment and install Ricgraph in it.
- Login as user root.
- Make sure Nginx has been installed.
- *Gunicorn* has already been installed when you installed the Python requirements.

• Install the Nginx *Ricgraph Explorer* configuration file: copy file *ricgraph\_server\_config/ricgraph\_explorer.conf-nginx* to /etc/nginx/vhosts.d, type:

cp /opt/ricgraph\_venv/ricgraph\_server\_config/ricgraph\_explorer.conf-nginx /etc/n chmod 600 /etc/nginx/vhosts.d/ricgraph\_explorer.conf-nginx

#### 8.5.3 Post-install steps Nginx

- Login as user root.
- Move the Nginx *Ricgraph Explorer* configuration file to its final location:

mv /etc/nginx/vhosts.d/ricgraph\_explorer.conf-nginx /etc/nginx/vhosts.d/ricgraph

However, for Ubuntu do:

mv /etc/nginx/sites-available/ricgraph\_explorer.conf-nginx /etc/nginx/sites-avai ln -s /etc/nginx/sites-enabled/ricgraph\_explorer.conf /etc/nginx/sites-available

Change *ricgraph\_explorer.conf* in such a way it fits your situation. Make the modification to *ricgraph\_explorer.conf* as described in the comments at the start of *ricgraph\_explorer.conf*. Test the result.

• Make it run by typing:

systemctl enable nginx.service
systemctl start nginx.service

Check the log for any errors, use one of:

systemctl -l status nginx.service
journalctl -u nginx.service

- Exit from user root.
- Now you can use Ricgraph Explorer from inside your virtual machine by typing http://localhost in your web browser in the virtual machine, or from outside your virtual machine by going to http://[your IP address] or http://[your hostname].

• You can use the Ricgraph REST API from inside your virtual machine by using the path http://localhost:3030/api followed by a REST API endpoint, or from outside your virtual machine by using the path http://[your IP address]/api or http://[your hostname]/api, both followed by a REST API endpoint.

## 8.6 How to install Ricgraph and Ricgraph Explorer on SURF Research Cloud

SURF Research Cloud is a portal where you can easily build a virtual research environment. You can use preconfigured workspaces, or you can add them yourself. A virtual research environment or workspace is a virtual machine that you can use to install Ricgraph. Please follow these steps if you would like to install Ricgraph and Ricgraph Explorer on SURF Research Cloud.

#### 8.6.1 Preliminaries

- Make sure you have access to SURF Research Cloud and that you have a wallet available. A *wallet* is a budget. A wallet has *credits*, and these credits are used to pay for the SURF computing resources. The more resources you use, the more you have to pay. For *resources*, think of disk space, the number of CPUs, the amount of memory, and the time the virtual machine is running.
- If you do not have access to SURF Research Cloud or you do not have a wallet, please contact the SURF Research Cloud contact person at your organization. These persons may be at the Research Data Management Support desk, service desk, or help desk of your organization, or they might be persons like research engineers, data stewards, data managers, or data consultants.

#### 8.6.2 Create a SURF Research Cloud workspace

To create a SURF Research Cloud workspace, follow the following steps:

- Go to the SURF Research Cloud portal and log in.
- Optional: Allocate storage. This step is only required if you expect to install a lot of programs on the virtual research environment and expect to create or use a lot of data. In the case of Ricgraph: > 100M nodes and edges. This is for advanced use only, since this storage will be attached to /data in the virtual research environment, and not to /var/lib, where the Neo4j Community Edition graph database lives.
  - Click on "Create new storage".
  - Select the collaborative organization that you want to use for running Ricgraph. If you have only one, it will be preselected.
  - Select your wallet. If you have only one, it will be preselected.
  - Select the cloud provider. We use "SURF HPC Cloud volume".
  - Choose the size of your storage. In the video below we use "100GB".
     The larger, the more credits it will cost.
  - Enter a name and a description.
  - After a few moments your storage will be created and available.
- Create a workspace (that is, a virtual machine to run Ricgraph in):
  - Click on "Create new workspace".
  - Select the collaborative organization that you want to use for running Ricgraph (as above). If you have only one, it will be preselected.
  - Select your wallet (as above). If you have only one, it will be preselected.
  - Now select a "catalogue item", that is, a pre-installed virtual machine. Choose "Ubuntu Desktop".
  - Select the cloud provider. We use "SURF HPC Cloud".
  - Select which version of Ubuntu you want to use. Choose "Ubuntu 22.04 Desktop".
  - Select a configuration. In the video below we use "1 Core 8 GB RAM". The larger, the more credits it will cost.
  - By default, the workspace has ~95GB of storage on the system and home partition.
  - Optionally you can add more storage, above is explained how to allocate it. If you have done this, select this additional storage.
  - Rename your workspace.
  - After some minutes your workspace will be created and available. It

will be started up automatically.

- Note that your workspace has a *will be removed* date. You might want to set it to a suitable date.

• Done.

You might want to watch the video how to install Ricgraph and Ricgraph Explorer on SURF Research Cloud (2m14s) (click to view or download). Note that in the video, we use an old version of Ubuntu. Please use Ubuntu 22.04 as described above.

#### 8.6.3 Install Ricgraph in a SURF Research Cloud workspace

The next steps in your workspace are to install the graph database backend and Ricgraph. You can install Ricgraph for a single user or Ricgraph as a server. Note that if you would like to use a webserver, you will need to use Nginx.

#### 8.6.4 Pause and resume a SURF Research Cloud workspace

On the SURF Research Cloud portal, you can *pause* and *resume* your workspace. *Pausing* means that the workspace will not run, and of course then it will not be accessible. If you have paused your workspace, it does not cost credits (money). If you *resume* your workspace, you can use it again.

#### 8.6.5 Access a SURF Research Cloud workspace

On the workspace window, you will find the name of the workspace. It will be a *https* link that ends with *.src.surf-hosted.nl*. SURF Research Cloud uses *guacamole*, which provides you with a desktop window in your browser. There are two ways to access your workspace and authenticate:

• Use port 443, then you will login on your workspace using SURF conext. In this case, the link will look like: https://[name of your workspace].src.surf-hosted.nl.

• Use port 3389, then you will login on your workspace using a one time password. In this case, the link will look like: https://[name of your workspace].src.surf-hosted.nl:3389.

#### 8.7 Steps to take to install Ricgraph as a server by hand

Skip this section if you have done the Fast and recommended way to install Ricgraph as a server and there were no errors.

- 1. Install your graph database backend.
- 2. Create a ricgraph user and group.
- 3. Create a Python virtual environment and install Ricgraph in it.
- 4. Create and update the Ricgraph initialization file. This is also the place where you specify which graph database backend you use.
- 5. Start harvesting data, see Ricgraph harvest scripts, or writing scripts, see Ricgraph script writing.
- 6. Start browsing using Ricgraph Explorer.

#### 8.7.1 Install your graph database backend

Install your graph database backend (choose one of these):

- Install and start Neo4j Community Edition.
- Install and start Memgraph.

#### 8.7.2 Create a ricgraph user and group

Follow these steps:

- Login as user root.
- Create group and user *ricgraph*. First check if they exist:

grep ricgraph /etc/group
grep ricgraph /etc/passwd

If you get output, they already exist, and you don't need to do this step. If you get no output, you will need to create the group and user:

```
groupadd --system ricgraph
```

useradd --system --comment "Ricgraph user" --no-create-home --gid ricgraph ricgraph tar xf /opt/ricgraph-X.YY.tar.gz

• Exit from user root.

8.7.3 Create a Python virtual environment and install Ricgraph in it

Follow these steps:

- Suppose you are a user with login *alice* and you are in Linux group users.
- Login as user root.
- For Debian/Ubuntu: type:
- apt install python3-venv
- Go to directory */opt*, type:

cd /opt

• Create a Python virtual environment: in */opt*, type:

python3 -m venv ricgraph\_venv

• Change the owner and group to your own user *alice* and group *users*, in /opt, type:

chown -R alice:users /opt/ricgraph\_venv

- The path */opt/ricgraph venv* is hardwired in the configuration ricgraph server config/ricgraph explorer qunicorn.service files and ricgraph server config/ricgraph explorer.conf-apache. This is done for security reasons. If you change the path, also change it in these files.
- Exit from user *root*. Do the following steps as your own user.

- Download the latest release of Ricgraph from the Ricgraph downloads page to directory /opt/ricgraph venv. Get the tar.gz version.
- Install Ricgraph: go to */opt/ricgraph venv*, type:

(X.YY is the version number you downloaded). You will get a directory /opt/ricgraph venv/ricgraph-X.YY.

 Merge the Ricgraph you have extracted with tar with the virtual environment, and do some cleanup: in /opt/ricgraph\_venv, type:

mv ricgraph-X.YY/\* /opt/ricgraph\_venv rm -r /opt/ricgraph\_venv/ricgraph-X.YY rm /opt/ricgraph\_venv/ricgraph-X.YY.tar.gz

- Activate the Python virtual environment: in */opt/ricgraph\_venv*, type: source bin/activate
- Install the standard Python requirements: in */opt/ricgraph\_venv*, type:

pip install setuptools pip wheel

• Install the Python requirements for Ricgraph: in /opt/ricgraph venv, type:

pip install -r requirements.txt

If you get an error message

ERROR: Could not find a version that satisfies the requirement neo4j>=5.8

then your Python version is too old. Please read How to solve an AttributeError: Neo4jDriver object has no attribute executequery.

- · Create a Ricgraph initialization file, read Ricgraph initialization file. This is also the place where you specify which graph database backend you use. You can find these settings in section GraphDB.
  - For Neo4j, enter the new password for Neo4j from section Install and start Neo4j Community Edition at the parameter graphdb password.

• Deactivate the Python virtual environment: type

deactivate

- Login as user root.
- Change the owner and group to ricgraph of directory */opt/ricgraph\_venv*. In */opt*, type
- chown -R ricgraph:ricgraph /opt/ricgraph\_venv
- Exit from user *root*.

# 9 Ricgraph in a container

This page describes how to install and run Ricgraph in a Podman container. A container is like a small box that holds an application and everything it needs to run, such as files and settings. It makes sure the application works the same on any computer. Podman is a tool for creating and running containers, and it is safe because it doesn't need special permissions to work. This makes it a good choice for running programs quickly and reliably. For other Ricgraph install options start reading at Install and configure Ricgraph for a single user.

To install and run Ricgraph in a Podman container, follow these steps:

- Read the notes on the Ricgraph Podman container.
- Install Podman.
- Install and run the Ricgraph Podman container.
- Optional: Advanced use of the Ricgraph Podman container.

Return to main README.md file.

#### 9.1 Notes on the Ricgraph Podman container

Warning: Do not use the Ricgraph Podman container in a production environment, since it does not provide a web server (such as Apache). If you would like to harvest a lot of items, you are better off using Ricgraph for a single user or Ricgraph as a server. The Ricgraph Podman container is ideal for instructional or personal use.

The Ricgraph Podman container has been designed in such a way that it is easy to use. This also means that some "good container design practices" have not been followed, in particular:

- the harvested items are stored in the Ricgraph Podman container (usually, one would use data storage in the filesystem on the host);
- Neo4j Community Edition and Ricgraph are in the same Ricgraph Podman container (usually, one would use separate containers).

So, do not use the Ricgraph Podman container in a production environment.

#### 9.2 Install Podman

Only for this step you will need *Administrator* or *root* access on your machine. Choose one of the following:

- Linux:
  - Install command line version Podman.
  - Install GUI version Podman Desktop.
- MacOS:
  - Install command line version Podman.
  - Install GUI version Podman Desktop.
- $\cdot$  Windows:
  - Install command line version Podman.
  - Install GUI version Podman Desktop.

The following information might help you to install Podman on Windows. Note that you will need *Administrator* privileges.

- For both the command line version and the Desktop GUI version of Podman, you will need *Windows Subsystem for Linux (WSL) 2.* Install it as follows:

- \* Open a PowerShell (Admin) or Terminal (Admin).
- \* Type:

dism.exe /online /enable-feature /featurename:VirtualMachinePlatfopowerShenwhitew. If you use the Podman Windows command line version, wsl --install --no-distribution

Next, for the command line version of Podman:

- Get Podman from the GitHub releases page https://github.com/ containers/podman/releases, choose the latest version with file name extension .msi.
- Install it by clicking on the downloaded file, follow the instructions.
- Exit your Administrator privileges.
- Open Windows PowerShell or Terminal and use the commands in the sections below.

Or, for the Desktop GUI version of Podman:

- Follow the instructions from Install GUI version Podman Desktop.
- You might get a Windows Firewall message, asking "Do you want to allow public, private, and domain networks to access this app?". Choose Private networks from the list and click "Allow".
- After installation of the Desktop GUI version of Podman, vou will need to make a Virtual Machine to run Podman. Podman Desktop will ask you to do this. Follow the instructions.
- I got a weird error during installation, but rebooting my Windows machine resolved it. Try it, if you get a similar message.
- Note that Podman Desktop GUI also installs Podman command line version.
- Exit your Administrator privileges.
- Open Windows PowerShell or Terminal and use the commands in the sections below.
- Or use Podman Desktop GUI to pull the Ricgraph image (for location see below) and to start it.

# 9.3 Install and run the Ricgraph Podman container

For the following commands, you will need a Terminal or Command line or first start Windows Subsystem for Linux (WSL) by typing wsl in the Windows Terminal or PowerShell window.

The Ricgraph container is hosted on the GitHub Container repository https: //ghcr.io/utrechtuniversity/ricgraph.

There are two ways to run the Ricgraph Podman container:

1. You have not done a podman commit (see below).

Run the latest (newest) version of the Ricgraph Podman container, and download it if you don't have it:

podman run --pull=newer --name ricgraph -d -p 3030:3030 ghcr.io/utrechtuniversit

If you get an error like Error: creating container storage: the container name "ricgraph" is already in use, type:

podman run --pull=newer --replace --name ricgraph -d -p 3030:3030 ghcr.io/utrech

2. You have done a podman commit (see below).

Run the local version of the Ricgraph Podman container (do not use the --pull=newer flag):

podman run --name ricgraph -d -p 3030:3030 ghcr.io/utrechtuniversity/ricgraph:la

If you get an error like Error: creating container storage: the container name "ricgraph" is already in use, type:

podman run --replace --name ricgraph -d -p 3030:3030 ghcr.io/utrechtuniversity/n

Starting the container takes about ten seconds. Explore items with Ricgraph Explorer, in your browser, go to http://localhost:3030.

If you started podman run for the first time, the container does not have data in it. You can observe this by going to the Ricgraph Explorer home page (http://localhost:3030), by scrolling down to the "About Ricgraph" section, and looking for the number of nodes and edges. The easiest method for getting data in it, is to run the *multiple\_harvest\_demo.sh* script that harvests a selection of research information of Utrecht University (from the data repository Yoda and the Research Software Directory). These two repositories do not need authentication keys. Running this script will take several minutes to complete. Type

podman exec -it ricgraph make run\_bash\_script

It is a design decision to store all harvested items in the Ricgraph container (see Notes on the Ricgraph Podman container), so make them permanent (i.e. also available after restart of the container) by typing:

podman commit ricgraph ghcr.io/utrechtuniversity/ricgraph:latest

Now restart the Ricgraph container to see the results:

podman restart ricgraph

Explore the harvested items with Ricgraph Explorer, in your browser, go to http://localhost:3030.

You can stop all containers by typing:

podman stop -a

To (re)start, use the podman run command above.

# 9.4 Advanced use of the Ricgraph Podman container

For the following commands, you will need a *Terminal* or *Command line* or *PowerShell* window. If you use the Podman Windows command line version, first start *Windows Subsystem for Linux (WSL)* by typing wsl in the Windows *Terminal* or *PowerShell* window.

Get the status of all containers:

#### podman ps

Remove all podman containers:

podman rmi -a -f

Use a bash shell "in" the Ricgraph Podman container:

podman exec -it ricgraph /bin/bash

To harvest other sources then Yoda and the Research Software Directory, you will need to modify the Ricgraph Podman container. You might need to add API keys to the Ricgraph initialization file *ricgraph.ini*. Follow these steps:

- Execute a bash shell in the Ricgraph Podman container (see above). Do the following commands in the container (in the bash shell you just started).
- Add API keys to *ricgraph.ini*, if necessary for the sources you would like to harvest. It is in */usr/local* in the container. You can use vim (or vi) to edit it.
- Go to the directory with harvest scripts:
  - cd /app/ricgraph/harvest
- Run a harvest script (they start with 'harvest\_'):

python [name of harvest script]

- or create your own harvest script. For more information, read Ricgraph harvest scripts. You can also use the scripts that harvest multiple sources.
- Make the data permanent in the container, see podman commit above. Note that the size of your container may explode if you harvest a lot of items.
- Restart the container, see podman restart above.

### You can also use Ricgraph Makefile targets.

For other useful commands, read the comment section of the Containerfile.

# 10 Ricgraph with Neo4j graph database backend

Ricgraph can use two graph database backends: Neo4j and Memgraph. This page describes how to install Ricgraph with Neo4j graph database backend. Neo4j has several products:

- Neo4j Community Edition, allows to explore the graph using Cypher queries only.
- Neo4j Desktop;
- (optional) Neo4j Bloom graph visualization tool, included with Neo4j Desktop.

To use Ricgraph, if you choose the Neo4j graph database backend (which is recommended for now), you need to choose between Neo4j Community Edition and Neo4j Desktop. On this page you can find:

- Neo4j Community Edition
- Neo4j Desktop
- Dumping, restoring, and emptying the Ricgraph database
  - Create a Neo4j Desktop database dump of Ricgraph
  - Create a Neo4j Community Edition database dump of Ricgraph
  - Restore a Neo4j Desktop database dump of Ricgraph in Neo4j Desktop
  - Restore a Neo4j Desktop database dump of Ricgraph in Neo4j Community Edition
  - Restore a Neo4j Community Edition database dump of Ricgraph in Neo4j Community Edition
  - Empty a Neo4j Community Edition database
- How to reset the Neo4j Community Edition password
- How to solve an AttributeError: Neo4jDriver object has no attribute executequery

#### Return to main README.md file.

### 10.1 Neo4j Community Edition

#### 10.1.1 Install and start Neo4j Community Edition

To do this, you can either use the Ricgraph Makefile and execute as user *root* command

make install\_enable\_neo4j\_community

#### or follow the steps below.

- Login as user root.
- Install Neo4j Community Edition (it is free). To do this, go to the Neo4j Deployment Center. Go to section "Graph Database Self-Managed", choose "Community". Choose the latest version of Neo4j. Then choose your favorite package format:
  - OpenSUSE/Fedora: "Red Hat Linux Package (rpm)".
  - Debian/Ubuntu: "Debian/Ubuntu Package (deb)".

Download the package and install it. You might get an error message about a failed dependency on *cypher-shell*, or on other dependencies.

- OpenSUSE/Fedora: use either rpm -i <packagename> (first install) or rpm -U <packagename> (update).
- Debian/Ubuntu: use apt install <packagename>
- If you get an error message about a failed dependency on *cypher-shell*, install *cypher-shell* separately as follows:

- \* Go to the Tools tab of the Neo4j Deployment Center. Go to section "Cypher Shell", choose the version of Cypher Shell that matches the version of the Neo4j Community Edition you have downloaded above. Then choose the version that fits your Linux version:
  - · OpenSUSE: "Linux cypher-shell\_X.YY.0-Z.noarch.rpm".
  - · Debian/Ubuntu: "Linux cypher-shell\_X.YY.0\_all.deb".
- \* Click "Download" and install it.
- \* Install again Neo4j Community Edition (see above).
- If you get an error message about failed other dependencies, install these other packages.
- You need to reset the default Neo4j password. This password is necessary in section Create a Python virtual environment and install Ricgraph in it. Run:

 $/usr/bin/neo4j-admin\ dbms\ set-initial-password\ [the\ new\ password]$ 

Note that this command only works if you have not started Neo4j yet. Otherwise, follow the steps in the next section *Post-install steps Neo4j Community Edition*.

• If the installation has finished, make sure it runs by typing:

systemctl enable neo4j.service
systemctl start neo4j.service

Check the log for any errors, use one of:

systemctl -l status neo4j.service
journalctl -u neo4j.service

• Exit from user *root*.

Skip the next section *Post-install steps Neo4j Community Edition* (you have already reset the password above).

#### 10.1.2 Post-install steps Neo4j Community Edition

If you use the Ricgraph Makefile, you do not need to do this. Otherwise, for your first run of Neo4j Community Edition only, you need to change the default username and password of Neo4j:

- In your web browser, go to http://localhost:7474/browser.
- Neo4j will ask you to login, use username neo4j and password neo4j.
- Neo4j will ask you to change your password. Change it. You will need this new password in section Create a Python virtual environment and install Ricgraph in it.

## 10.2 Neo4j Desktop

#### 10.2.1 Install Neo4j Desktop

To do this, you can either use the Ricgraph Makefile and execute command

make install\_neo4j\_desktop

or follow the steps below.

- Install Neo4j Desktop Edition (it is free). To do this, go to the Neo4j Deployment Center. Go to section "Neo4j Desktop". Choose the latest version of Neo4j Desktop. Download the Linux version. It is an AppImage, so it can be installed and used without root permissions. You will be asked to fill in a form before you can download. In the following screen you will be given a "Neo4j Desktop Activation Key". Save it.
- 2. The downloaded file is called something like *neo4j-desktop-X.Y.Zx86\_64.AppImage*, where *X.Y.Z* is a version number. Make it executable using "chmod 755 [filename]".

#### 10.2.2 Post-install steps Neo4j Desktop

- 1. Start Neo4j Desktop by clicking on the downloaded file.
- 2. Accept the license.
- 3. Enter your activation key in the right part of the screen. Click "Activate". If you do not have a key, fill in the left part of the screen. Click "Register with Email". Wait awhile.
- 4. Choose whether you would like to participate in anonymous reporting.
- 5. You may be offered updates for Neo4j Desktop components, please update.
- 6. Move your mouse to "Example Project" in the left column. A red trash can icon appears. Click it to remove the Example Project database "Movie DBMS". Confirm. Then wait awhile.
- 7. The text "No projects found" will appear. Create a project by clicking the button "+ New Project".
- 8. The text "Project" appears with the text "Add a DBMS to get started". Click on the "+ Add" button next to it and select "Local DBMS". Leave the name as it is ("Graph DBMS") and fill in a password. Click "Create".
- 9. [This step is not necessary if you use the Ricgraph Makefile] Insert the password in field *graphdb\_password* in the Ricgraph initialization file.
- 10. Exit Neo4j Desktop using the "File" menu and select "Quit". If your database was active a message similar to "Your DBMS [name] is running, are you sure you want to quit" appears, choose "Stop DBMS, then quit".11. Ready.

Now we need to find the port number which Neo4j Desktop is using:

- 1. Start Neo4j Desktop. Start the Graph DBMS.
- 2. Click on the words "Graph DBMS". At the right (or below, depending on the width of the Neo4j Desktop window) a new screen appears. Look at the tab "Details". Note the port number next to "Bolt port" (the default value is 7687). Only if it is a different value, you will need to insert this port number in field *graphdb\_port* in the Ricgraph initialization file, see below.
- 3. Ready.

#### 10.2.3 Start Neo4j Desktop

- 1. Click on the downloaded AppImage. It will be called something like *neo4j*-*desktop*-*X*.*Y*.*Z*-*x*86\_64.*AppImage*, where *X*.*Y*.*Z* is a version number.
- 2. At the top right there is a text "No active DBMS".
- 3. Move your mouse to the text "Graph DBMS". When you hoover it, a button will appear with the text "Start". Click it.
- 4. Wait until the Neo4j graph database engine has started. It may ask for a password that has been changed. Enter the password you have used while creating your database. Click "Save".
- 5. Now, next to the text "Graph DBMS" a green icon appears with the text "ACTIVE". Your graph database engine is active and ready for use.

#### 10.2.4 Neo4j Desktop with Bloom (optional)

This is only necessary if you plan to use Bloom. If you don't know (yet), skip this step for now, you can come back to it later.

Bloom is Neo4j Desktop's graph visualization tool. It is included with Neo4j Desktop. According to Neo4j it is: "A beautiful and expressive data visualization tool to quickly explore and freely interact with Neo4j's graph data platform with no coding required". Neo4j has extensive documentation how to use Bloom and a Bloom overview.

#### 10.2.4.1 Install Bloom configuration for Neo4j Desktop

- 1. Start Neo4j Desktop.
- 2. Click on the icon on the left side of Neo4j Desktop.
- 3. Click on "Neo4j Bloom". A new window appears.
- 4. In this window, click on the icon at the top left. A Bloom "Perspective" slides out (Neo4j has an extensive description how to use it).
- 5. Click on "neo4j > Untitled Perspective 1".
- 6. A new window appears. Right of the words "Untitled Perspective 1" there are three vertical dots. Click on it. Click on "Delete". The perspective "Untitled Perspective 1" is removed.

- 7. In the same window, right of the word "Perspectives" click on the word "Import". A file open window appears. Go to directory *neo4j\_config* that is part of Ricgraph and select file *ricgraph\_bloom\_config.json*. Click "Open". The perspective "ricgraph\_bloom\_config" is loaded.
- 8. Click on the text "ricgraph\_bloom\_config".
- 9. Note that the text "neo4j > Untitled Perspective 1" has been changed in "neo4j > ricgraph\_bloom\_config".
- 10. A few centimeters below "neo4j > ricgraph\_bloom\_config", just below the text "Add category", click on the oval "RicgraphNode". At the right, a new window will appear.
- 11. In this window, below the word "Labels", check if an oval box with the text "RicgraphNode" is shown. If not, click on "Add labels", click on "RicgraphNode".
- 12. Click on the icon to go back to the main screen of Bloom.
- 13. Click on the cog icon below , you might want to set "Use classic search" to "on".
- 14. Ready.

The following sections show some examples for a quick start using Bloom.

#### 10.2.4.2 Open Bloom

- 1. Start Neo4j Desktop.
- 2. Click on the icon on the left side of Neo4j Desktop.
- 3. Click on "Neo4j Bloom". A new window appears.

### 10.2.4.3 Execute queries in Bloom

The Ricgraph Bloom configuration file contains four different shortcuts for Cypher queries:

- "Node name [value of node to find]": finds a node where property *name* of a node has value [value of node to find].
- "Node category [value of node to find]": similar to "Node name ..." for property *category*.

- "Node value [value of node to find]": similar to "Node name ..." for property value.
- "Node comment [value of node to find]": similar to "Node name ..." for property *comment*.

These queries can be entered in the Bloom text box "Search graph", by typing e.g. "Node name ORCID" or "Node category data set". For more information see Bloom search bar and Boom pattern search.

Nodes found can be examined or expanded as described in the section Actions while clicking on a node. The result will be visualized as described in section Visualization of nodes.

#### 10.2.4.4 Actions while clicking on a node in Bloom

The following are some examples of actions while clicking on a node:

- Double left-click on a node: the properties of a node are shown in a window.
- Right-click right on a node, choose "Expand", choose "All": The node will be expanded with all nodes connected to it.
- Multiple nodes can be selected by selecting one node, holding the Control key and selecting other nodes.
- Right-click on a node, choose "Dismiss": This node will be removed from the visualization.
- Right-click on a node, choose "Dismiss other nodes": All other nodes will be removed from the visualization.
- For other actions, see Bloom actions.

#### 10.2.4.5 Visualization of nodes in Bloom

Nodes can be visualized in different ways, by changing e.g. their size or color. This can be changed as follows:

1. On the right side of the Bloom window, there is an icon . Click it (Neo4j has extensive documentation how to use it).

- 2. A new window appears. It shows the default settings for the display of nodes. You can change the color, size, the property to show on the node, and the icon.
- 3. In the tab "Rule-based" you can add your own rules.

The Ricgraph Bloom configuration file contains a few rules based on the value of properties. Rules which determine the color of a node:

- if property *category* = "person": color = blue.
- if property *category* = "data set": color = green.
- if property *category* = "journal article": color = yellow.
- if property *category* = "software": color = red.
- all other nodes: color = grey.

Rules which determine the size of a node:

- if property *url\_main* contains "uu01": size of the node = small. This indicates which nodes have been harvested from the data repository Yoda.
- if property *url\_other* contains "research-software-directory": size of the node = large. This indicates which nodes have been harvested from the Research Software Directory.
- all other nodes: size = medium.

# 10.3 Dumping, restoring, and emptying the Ricgraph database

Depending on your situation (whether you use Neo4j Desktop or Neo4j Community Edition), this section lists various methods for dumping, restoring, and emptying the Ricgraph database.

#### 10.3.1 Create a Neo4j Desktop database dump of Ricgraph

To create a Neo4j Desktop database dump of Ricgraph, follow these steps:

1. Start Neo4j Desktop if it is not running, or stop the graph database if it is running.

- 2. Hoover over the name of your graph database (probably "Graph DBMS"), and click on the three horizontal dots at the right.
- 3. Select "Dump".
- 4. Your graph database will be dumped. This may take a while. When it is ready, a message appears.
- 5. Ready.

# 10.3.2 Create a Neo4j Community Edition database dump of Ricgraph

To create a Neo4j Community Edition database dump of Ricgraph, either use the Ricgraph Makefile and execute command

make dump\_graphdb\_neo4j\_community

or follow the steps below.

- 1. Login as user root.
- 2. Stop Neo4j Community Edition:

systemctl stop neo4j.service

3. To be able to restore a Neo4j database dump you need to set several permissions on */etc/neo4j*:

chmod 640 /etc/neo4j/\* chmod 750 /etc/neo4j

4. Do the database dump:

neo4j-admin database dump --expand-commands system --to-path=[path to database dump --expand-commands neo4j --to-path=[path to database dump --expand-commands neo4j --to-path=[path to database dump database dump --expand-commands neo4j --to-path=[path to database dump database dump database dump neo4j --to-path=[path to database dump database du

5. Start Neo4j Community Edition:

systemctl start neo4j.service

6. Check the log for any errors, use one of:

systemctl -l status neo4j.service
journalctl -u neo4j.service

7. Exit from user *root*.

# 10.3.3 Restore a Neo4j Desktop database dump of Ricgraph in Neo4j Desktop

To restore a Neo4j Desktop database dump of Ricgraph in Neo4j Desktop, follow these steps:

- 1. Start Neo4j Desktop if it is not running, or stop the graph database if it is running.
- 2. Click on the button "Add" on the right side of "Project" and select "File".
- 3. Select the file "neo4j.dump" from a previous Neo4j Desktop database dump. This file will be added to the "File" section a little down the "Project" window.
- 4. Hoover over this file and click on the three horizontal dots at the right.
- 5. Select "Create new DBMS from dump".
- 6. Give it a name, e.g. "Graph DBMS from import file".
- 7. When asked, enter the password you have specified in the Ricgraph initialization file (this saves you from entering a new password in that file).
- 8. A new local graph database is being created. This may take a while.
- 9. Hoover over the newly created graph database and click "Start" to run it.
- 10. Optional: Once it is active, install the Bloom configuration.
- 11. Now you are ready to explore the data using Ricgraph Explorer or using Bloom.

# 10.3.4 Restore a Neo4j Desktop database dump of Ricgraph in Neo4j Community Edition

To restore a Neo4j Desktop database dump of Ricgraph in Neo4j Community Edition, follow these steps:

1. Login as user root.

#### 2. Stop Neo4j Community Edition:

systemctl stop neo4j.service

3. To be able to restore a Neo4j database dump you need to set several permissions on */etc/neo4j*:

chmod 640 /etc/neo4j/\* chmod 750 /etc/neo4j

4. Save the old database:

cd /var/lib/neo4j mv data/ data-old

5. Go back to your working directory and restore the database dump:

#### cd

neo4j-admin database load --expand-commands neo4j --from-path=[path to database

For *path to database dump directory*, specify the path, not the path and the name of the database dump file (this name is *neo4j.dump*, it will be inferred automatically by the *neo4j-admin* command).

6. Set the correct permissions on */var/lib/neo4j/data*:

cd /var/lib/neo4j chown -R neo4j:neo4j data

7. Start Neo4j Community Edition:

systemctl start neo4j.service

8. Check the log for any errors, use one of:

systemctl -l status neo4j.service
journalctl -u neo4j.service

- 9. In your web browser, go to http://localhost:7474/browser.
- 10. Neo4j will ask you to login, use username neo4j and password neo4j.

- 11. Neo4j will ask you to change your password, for the new password, enter the password you have specified in the Ricgraph initialization file (this saves you from entering a new password in that file).
- 12. Restart Ricgraph Explorer if you use Ricgraph in a multi-user environment:

systemctl restart ricgraph\_explorer\_gunicorn.service

13. Check the log for any errors, use one of:

systemctl -l status ricgraph\_explorer\_gunicorn.service
journalctl -u ricgraph\_explorer\_gunicorn.service

14. Done. If all works well you might want to remove your old database:

cd /var/lib/neo4j rm -r data-old

15. Exit from user root.

### 10.3.5 Restore a Neo4j Community Edition database dump of Ricgraph in Neo4j Community Edition

To restore a Neo4j Community Edition database dump of Ricgraph in Neo4j Community Edition, either use the Ricgraph Makefile and execute command

make restore\_graphdb\_neo4j\_community

or follow the steps below.

- 1. Login as user root.
- 2. Stop Neo4j Community Edition:

systemctl stop neo4j.service

3. To be able to restore a Neo4j database dump you need to set several permissions on */etc/neo4j*: chmod 640 /etc/neo4j/\* chmod 750 /etc/neo4j

4. Save the old database:

cd /var/lib mv neo4j/ neo4j-old mkdir /var/lib/neo4j

- 5. Go back to your working directory and restore the database dump:
  - cd

neo4j-admin database load --expand-commands system --from-path=[path to database neo4j-admin database load --expand-commands neo4j --from-path=[path to database

For *path to database dump directory*, specify the path, not the path and the name of the database dump file, it will be inferred automatically by the *neo4j-admin* command.

6. Set the correct permissions on */var/lib/neo4j/data*:

cd /var/lib chown -R neo4j:neo4j neo4j

7. Start Neo4j Community Edition:

systemctl start neo4j.service

8. Check the log for any errors, use one of:

systemctl -l status neo4j.service
journalctl -u neo4j.service

9. Restart Ricgraph Explorer if you use Ricgraph in a multi-user environment:

systemctl restart ricgraph\_explorer\_gunicorn.service

10. Check the log for any errors, use one of:

systemctl -l status ricgraph\_explorer\_gunicorn.service
journalctl -u ricgraph\_explorer\_gunicorn.service

11. Done. If all works well you might want to remove your old database:

```
cd /var/lib
rm -r neo4j-old
```

12. Exit from user root.

#### 10.3.6 Empty a Neo4j Community Edition database

To do this, use the Ricgraph Makefile and execute command

make empty\_graphdb\_neo4j\_community

#### 10.4 How to reset the Neo4j Community Edition password

To reset the Neo4j Community Edition password, follow the following steps:

- 1. Login as user root.
- 2. Stop Neo4j Community Edition, type:

systemctl stop neo4j

3. Edit file /etc/neo4j/neo4j.conf and change line

#dbms.security.auth\_enabled: "false"

#### to

dbms.security.auth\_enabled: "false"

4. Start Neo4j, type:

systemctl start neo4j

5. Start cypher-shell, type:

cypher-shell

and in cypher-shell, type:

ALTER USER neo4j SET PASSWORD '<new-password>'; :exit

6. Stop Neo4j, type:

systemctl stop neo4j

7. Edit file /etc/neo4j/neo4j.conf and change line dbms.security.auth\_enabled: "false"

oms.security.autn\_enabled: "I

to

#dbms.security.auth\_enabled: "false"

8. Start Neo4j, type:

systemctl start neo4j

9. Exit from user root.

# 10.5 How to solve an AttributeError: Neo4jDriver object has no attribute executequery

If, at some point while running Ricgraph scripts or Ricgraph Explorer in a virtual environment, you get an error message like:

Traceback (most recent call last):
 File "ricgraph\_explorer.py", line 2930, in <module>
 initialize\_ricgraph\_explorer()
 File "ricgraph\_explorer.py", line 2867, in initialize\_ricgraph\_explorer
 name\_all = rcg.read\_all\_values\_of\_property('name')
 File "/opt/ricgraph\_venv/ricgraph/ricgraph.py", line 1140, in read\_all\_values\_of\_]
 result = \_graph.execute\_query(cypher\_query,
AttributeError: 'Neo4jDriver' object has no attribute 'execute\_query'

or, after typing pip install -r requirements.txt, you get an error message: ERROR: Could not find a version that satisfies the requirement neo4j>=5.8 then this means that your version of the Python module *neo4j* is too old. Note that this is related to the Python module *neo4j*, not to the graph database backend Neo4j Desktop or Neo4j Community Edition. You need at least version 5.8 of the Python module *neo4j*. With an "old" version of Python (3.6 and earlier), an old version of module *neo4j* will be used. The only way to solve this is using a new version of Python while creating the Python virtual environment. You can do this by using the following command:

• Create a Python virtual environment: in */opt*, type:

python3.11 -m venv ricgraph\_venv

in section Create a Python virtual environment and install Ricgraph in it. For *python3.11* you can take any Python version that is installed on your computer, as long as it is at least Python 3.9.

65

# 11 Ricgraph with Memgraph graph database backend

Ricgraph can use two graph database backends: Neo4j and Memgraph. This page describes how to install Ricgraph with Memgraph graph database backend.

#### 11.1 Install and start Memgraph

As an alternative to Neo4j, you can also use Memgraph. Memgraph is an in memory graph database and therefore faster than Neo4j. However, it has not been tested extensively with Ricgraph.

- Login as user root.
- Make sure you have Docker. If not, install it:
  - Debian/Ubuntu: follow Install Docker using the apt repository.
- Install Memgraph Platform. To do this, follow the instructions on the Memgraph download page in the block 'Memgraph Platform'. Mem-graph will be started automatically. Stop it by typing *Control-C*.
- To start Memgraph, go to the directory *memgraph-platform*:
- cd memgraph-platform

and type:

docker compose up

If you want to stop Memgraph, type Control-C.

• In the log printed on the terminal, you might get a message like: Max virtual memory areas vm.max\_map\_count 65530 is too low, increase to at least 262144. To resolve this, create a file in */etc/sysctl.d* with the name 90-local.conf and the following content:

vm.max\_map\_count=262144

After you have done that, type:

sysctl --system

and the message should be gone. Start Memgraph as above.

- To use Memgraph Platform, go to http://localhost:3000.
- How to start Memgraph automatically at system startup, is a 'to be done'.

# 12 Ricgraph REST API

You can use the Ricgraph REST API to get data from Ricgraph, or to execute predefined queries in Ricgraph and Ricgraph Explorer, in such a way that your code is not dependent on any Ricgraph code. In your own code, you do not need to include any of the Ricgraph dependencies, nor include *ricgraph.py* or *ricgraph\_explorer.py*. You can use the Ricgraph REST API to programmatically get items from Ricgraph, as an alternative to using the user interface Ricgraph Explorer.

Note that this will only work if someone has a running Ricgraph and Ricgraph Explorer instance available, which you can access over the web. That someone can be yourself, where you run Ricgraph and Ricgraph Explorer on your own computer or use a cloud service such as SURF Research Cloud. Alternatively, it can be someone else, who is doing that for you or for a community of users.

If you use the Ricgraph REST API, will get these results in a JSON format. The Ricgraph REST API uses the OpenAPI standard. It gives access to Ricgraph function calls both in *ricgraph.py* and in *ricgraph\_explorer.py*. Read more about REST (representational state transfer), or read more about API (application programming interface).

On this page, you can learn more about:

- Installation of the Ricgraph REST API
- Use of the Ricgraph REST API
- Ricgraph REST API endpoint documentation
- Generate the Ricgraph REST API documentation page

Return to main README.md file.

# 12.1 Installation of the Ricgraph REST API

The Ricgraph REST API is part of Ricgraph Explorer. Read how to start Ricgraph Explorer.

Depending on your needs, you might also want to read:

- Install and use a service unit file to run Ricgraph Explorer and the Ricgraph REST API.
- Use Apache, WSGI, and ASGI to make Ricgraph Explorer and the Ricgraph REST API accessible from outside your virtual machine.

# 12.2 Use of the Ricgraph REST API

A call to a REST API consists of a hostname, sometimes a port number, the path */api/*, followed by a REST API endpoint. For example, in

http://127.0.0.1:3030/api/person/search?value=John+Doe

/person/search is the endpoint, value the name of a query parameter to the REST API, and John+Doe the value for the query parameter.

You can use the online documentation in Ricgraph Explorer for the REST API. It lists the various endpoints and parameters, and you can try out each endpoint with values for parameters as you like. To do this, click the "REST API doc" button in the top bar of Ricgraph Explorer, and you will get an explanation how to do this. See the figure below, that shows part of the Home page of Ricgraph Explorer.

	search for a (sub-)organization search for a skill, expertise area or resear search for anything (broad search	rch area				
tems, re 'ou can	elations that are not present in any of the use Ricgraph Explorer to explore Ricgrap search for a person	e separate source syste	ems. It is flexible and exten nethods to start exploring:	sible, and can be adapted t	o new application areas.	
elation: Ricgraph pelow fo	is between these items. h can store many types of items into a sin or the sources of the items in this instanc	igle graph (network). T e of Ricgraph). Ricgrap	hese items can be obtaine h facilitates reasoning abo	d from various systems and ut these items because it ir	i from multiple organizations	s (see

Figure 12.1: Home page of Ricgraph Explorer.

## 12.3 Ricgraph REST API endpoint documentation

Read the Ricgraph REST API documentation page. This page is automatically generated from the OpenAPI yaml specification file.

#### 12.4 Generate the Ricgraph REST API documentation page

Please read Create the Ricgraph REST API documentation.

This is documentation for version 2.12 of Ricgraph – Research in context graph. It is a pdf of docs.ricgraph.eu. For more information about Ricgraph, Ricgraph Explorer, and the Ricgraph REST API, please read the reference publication doi.org/10.1016/j.softx.2024.101736, visit the website www.ricgraph.eu, or go to the GitHub repository github.com/UtrechtUniversity/ricgraph.

# 13 Ricgraph - Research in context graph REST API

This REST API documentation has been generated automatically.

13.4 Base URL

**REST API version 1.0.0.** 

#### 13.1 What to find on this page?

This is the documentation page for the Ricgraph REST API. You can use the Ricgraph REST API to programmatically get items from Ricgraph, as an alternative to using the user interface. You will get these items in a JSON format.

#### 13.2 How to use

In the left column of this page, you can explore the calls (i.e., the REST API operations) that are available. You can also try them out, by selecting a call, entering values in the 'Query-string parameters' subsection of the 'Request' section, and clicking the 'Try' button. Next, a gray tabbed box will appear. In the 'Response' tab of that box, you will get the JSON response. In the 'CURL' tab of that box, you will get a curl call with an URL (web link) that you can use in a browser or in your code. You don't need to provide authentication to use these calls.

#### **13.3 Technicalities**

The Ricgraph REST API uses the OpenAPI standard. It gives access to Ricgraph function calls both in *ricgraph.py* and in *ricgraph\_explorer.py*. Read more about REST (representational state transfer), or read more about API (application programming interface).

URL	Description
/api	

## **13.5 Endpoints**

	Method	Endpoint	Description
	GET	GET /person/search	Search for a person
	GET	GET /per-	Show all information
		son/all_information	related to this person
	GET	GET /per-	Find persons that
		son/share_research_res	sshtare any share re-
			search result types
	CEM	CETT /nor	With this person
	GET	GET /per-	rind organizations
		son/conaborating_orga	laborates with
	GET	GET /person/enrich	Find information
		/ /	harvested from other
			source systems, not
			present in this source
			system
	GET	GET /organiza-	Search for a (sub-
	0.D.m.	tion/search	)organization
	GET	GET /organiza-	Show all information
			related to this organi-
	GET	GET /organiza-	Find any information
		tion/information perso	onfromespetrsons or their
		, <u> </u>	results in this organi-
			zation
	GET	GET /organiza-	Find information
		tion/enrich	harvested from other
			source systems, not
			present in this source
	CFT	GET /compo_	Systelli Soarch for a skill
	0E1	tence/search	expertise area or
		tenee, bearen	research area
	GET	GET /compe-	Show all information
		tence/all_information	related to this compe-
This is documentation for version 2.12 of Ricgraph – Research in context graph. It is a pdf of docs.ricgraph please read the reference publication doi org/10.1016/j softy 202/.101726 visit the website unaw ricgraph of	eu. For more information about l	Ricgraph, Ricgraph Explorer, and	the Ricgraph REST API, 70
preuse read the reference publication dotorg/10.1010/j.301tx.2024.101/30, visit the website www.itegraph.e	GET	GET /broad_search	Search for anything
			(broad search)
	(FET)	GET /ad-	Advanced search

#### 13.6 GET /person/search

#### Search for a person

#### Parameters

Name	Туре	Required	Description
value	string	False	Search for a value in Ric- graph field valuevalue
max_nr_items	string	False	The maximum number of items to return, or 0 to return all items
Responses			
• 200: OK			
• 250: Nothing f	ound		

• 251: Invalid search

Parameters

#### 13.7 GET /person/all\_information

Show all information related to this person

Name	Туре	Required	Description
key	string	True	Search for a value in Ric- graph field *_key*
max_nr_items	string	False	The maximum number of items to return, or 0 to return all items

#### Responses

- 200: OK
- 250: Nothing found
- 251: Invalid search

#### 13.8 GET /person/share\_research\_results

Find persons that share any share research result types with this person

#### Parameters

Name	Туре	Required	Description
key	string	True	Search for a value in Ric- graph field *_key*
max_nr_items	string	False	The maximum number of items to return, or 0 to return all items

#### Responses

#### • 200: OK

- 250: Nothing found
- 251: Invalid search

#### 13.9 GET /person/collaborating\_organizations

#### Find organizations that this person collaborates with

With this call you will get an overview of organizations that his person collaborates with. A person X from organization A collaborates with a person Y from organization B if X and Y have both contributed to the same research result.

#### Parameters

Name	Туре	Required	Description
key	string	True	Search for a value in Ric- graph field *_key*
max_nr_items	string	False	The maximum number of items to return, or 0 to return all items

#### Responses

• 200: OK

• 250: Nothing found

• 251: Invalid search

Find information harvested from other source systems, not present in this source system

13.10 GET /person/enrich

The process of improving or enhancing information in a source system is called *enriching* that source system. This is only possible if you have harvested more than one source system. By using information found in one or more other harvested systems, information in this source system can be improved or enhanced. With this call you can enter a name of one of your source systems. Ricgraph will show what information can be added to this source system, based on the information harvested from other source systems. This call enriches persons. Note that if you specify *category\_want*, you will need to add category *person* too, since otherwise you will not be able to find *person\_identifying\_nodes* (the nodes to add the newly found information to).

#### Parameters
Name	Туре	Required	Description
key	string	True	Search for a value in Ric- graph field * key*
name_want	array	False	Return only neighbor nodes whose field namename matches any value in the provided list; if the list is empty, return all neighbor nodes regard- less of their field namename
category_want	array	False	Return only neighbor nodes whose field categorycate- gory matches any value in the provided list; if the list is empty, return all neighbor nodes regard- less of their field catego- rycategory
source_system	string	True	The name of the source system you would like to enrich
max_nr_items	string	False	The maximum

#### Responses

- 200: OK
- 250: Nothing found
- 251: Invalid search

## 13.11 GET /organization/search

#### Search for a (sub-)organization

#### Parameters

Name	Туре	Required	Description
value	string	False	Search for a value in Ric- graph field valuevalue
max_nr_items	string	False	The maximum number of items to return, or 0 to return all items

#### Responses

- 200: OK
- 250: Nothing found
- 251: Invalid search

## 13.12 GET /organization/all\_information

Show all information related to this organization

#### Parameters

This is documentation for version 2.12 of Ricgraph - Research in context grammers of the second seco

return all items

Name	Туре	Required	Description	Name	Туре	Required	Description
key	string	True	Search for a value in Ric- graph field *_key*	key	string	True	Search for a value in Ric- graph field *_key*
max_nr_items Responses • 200: OK	string	False	The maximum number of items to return, or 0 to return all items	name_want	array	False	Return only neighbor nodes whose field namename matches any value in the provided list; if the list is empty, return all neighbor nodes regard- less of their field namename
<ul> <li>• 250: Nothing:</li> <li>• 251: Invalid se</li> <li>13.13 GET /org</li> </ul>	found earch ; <b>anization/ir</b>	nformation_perso	ons_results	category_want	array	False	Return only neighbor nodes whose field categorycate- gory matches any value in the provided list; if the list is empty, return all neighbor
Find any informat Parameters	ion from perso	ons or their results in	this organization	max_nr_items	string	False	less of their field catego- rycategory The maximum number of items to return, or 0 to return all items

This is documentation for version 2.12 of Ricgraph – Research in context graph. It is a pdf of docs.ricgraph.eu. For more information about Ricgraph, Ricgraph Explorer, and the Ricgraph REST API, please read the reference publication doi.org/10.1016/j.softx.2024.101736, visit the website www.ricgraph.eu, or go to the GitHub repository github.com/UtrechtUniversity/ricgraph.

· 200: OK       key       string       True       Search for a value in RE graph field "key"         · 250: Nothing found       name_want       array       False       Return only neighbor nodes whose field namename matches any value in the provided lists if the list is empty, return all neighbor nodes regard less of their field namename matches any value in the provided lists if the list is category_want array       False         Find information harvested from other source systems, not present in this source system       category_want       array       False         The process of improving or enhancing information in a source system is called <i>enriching</i> that source system. This is only possible if you have harvested system, information in this source system. This is only possible if you have harvested system, information in this source system. This is only possible if sou have harvested system, information in this source system. This is only possible if you have harvested system, information in this source system. This is only possible if you have harvested system, information in this source system. This call of your source system. This is only possible if you have harvested system, information harvested system. This called enriching that source system. This and added to this source system and allor et uris information that allows to you find the node who what information harvested system. This called or other source system. This called number source system. This called article sorter source system. This called article arteport source sorter were sorter who what information that allows to you find the node who what information (the lates or the source system and you would like the list is the list is the list or the source system is called article arteport than the source system article artepor	Responses	Name	Туре	Required	Description
<ul> <li>250: Nothing found         <ul> <li>ame_want</li> <li>array</li> <li>False</li> <li>Return only neighbor nodes whose field namename matches any value in the provided list; if the list is empty, return all neighbor nodes regard less of their field namename field namename matches any value in the provided list; if the list is source system</li> </ul> </li> <li>Find information harvested from other source systems, not present in this source system is category_want array</li> <li>False</li> <li>False</li> <li>Return only neighbor nodes whose field is source system. This is only possible if you have harvested systems. By using information in a source system is called enriching that source system. This is only possible if you have harvested systems. Information in this source system can be improved or enhanced. With this call you can enter a name of one of your source with source system, based on the information harvested from other source systems. This source system can be improved from other compared to the REST API call /person/enrich, since that call also returns information that allows to you find the node to enrich in the source system that information (yet).</li> <li>Farameters</li> </ul>	• 200: OK	key	string	True	Search for a value in Ric- graph field * key*
<ul> <li>• 251: Invalid search</li> <li>and the se</li></ul>	• 250: Nothing found	name_want	array	False	Return only neighbor nodes
13.14 GET /organization/enrichall neighbor nodes regard less of their field namename source systemFind information harvested from other source systems, not present in this source systemcategory_want arrayarrayFalseReturm only neighbor nodes whose field categorycate- 	• 251: Invalid search				namename matches any value in the provided list; if the list is
Find information harvested from other source systems, not present in this source systemcategory_wantarrayFalseReturnonly neighbor nodes whoseSource systemThe process of improving or enhancing information in a source system is called <i>enriching</i> that source system. This is only possible if you have har- vested more than one source system. By using information found in one or more other harvested systems, information in this source system can be im- proved or enhanced. With this call you can enter a name of one of your source system, based on the information harvested from other source systems. This call enriches organizations. It is different compared to the REST API call /per- son/enrich, since that call also returns information (yet).Returnonly neighbor nodes whoseReturnonly 	13.14 GET /organization/enrich				empty, return all neighbor nodes regard- less of their field namename
The process of improving or enhancing information in a source system is called <i>enriching</i> that source system. This is only possible if you have har- vested more than one source system. By using information found in one or more other harvested systems, information in this source system can be im- proved or enhanced. With this call you can enter a name of one of your source system. Based on the information harvested from other source systems. This call enriches organizations. It is different compared to the REST API call /per- son/enrich, since that call also returns information that allows to you find the node to enrich in the source system you specify. The REST API call /organi- zation/enrich does NOT returns that information (yet). Parameters True Parameters	Find information harvested from other source systems, not present in this source system	category_want	array	False	Return only neighbor nodes whose field
zation/enrich does NOT returns that information (yet). source_system string True The name of the source system you would like you would like to enrich	The process of improving or enhancing information in a source system is called <i>enriching</i> that source system. This is only possible if you have harvested more than one source system. By using information found in one or more other harvested systems, information in this source system can be improved or enhanced. With this call you can enter a name of one of your source systems. Ricgraph will show what information can be added to this source system, based on the information harvested from other source systems. This call enriches organizations. It is different compared to the REST API call /person/enrich, since that call also returns information that allows to you find the node to enrich in the source system you specify. The REST API call /organi-				gory matches any value in the provided list; if the list is empty, return all neighbor nodes regard- less of their field catego- rycategory
	zation/enrich does NOT returns that information (yet).	source_system	string	True	The name of the source system you would like
max nr items string False The maximum	Parameters	max nr items	string	False	to enrich The maximum

This is documentation for version 2.12 of Ricgraph – Research in context graph. It is a pdf of docs.ricgraph.eu. For more information about Ricgraph Explorer, and the Ricgraph Rest APR/Le113/5 please read the reference publication doi.org/10.1016/j.softx.2024.101736, visit the website www.ricgraph.eu, or go to the GitHub repository github.com/UtrechtUniversity/ricgraph. to return, or 0 to return or 0 to return all items

#### Responses

- 200: OK
- 250: Nothing found
- 251: Invalid search

## 13.15 GET /competence/search

#### Search for a skill, expertise area or research area

#### Parameters

Name	Туре	Required	Description
value	string	False	Search for a value in Ric- graph field valuevalue
max_nr_items	string	False	The maximum number of items to return, or 0 to return all items
Responses			
• 200: OK			
• 250: Nothing f	ound		

• 251: Invalid search

Parameters

# 13.16 GET /competence/all\_information

Show all information related to this competence

Name	Туре	Required	Description
key	string	True	Search for a value in Ric- graph field *_key*
max_nr_items	string	False	The maximum number of items to return, or 0 to return all items

#### Responses

- 200: OK
- 250: Nothing found
- 251: Invalid search

# 13.17 GET /broad\_search

## Search for anything (broad search)

#### Parameters

Name	Туре	Required	Description
value	string	False	Search for a value in Ric- graph field valuevalue
max_nr_items	string	False	The maximum number of items to return, or 0 to return all items

#### Responses

#### • 200: OK

- 250: Nothing found
- 251: Invalid search

# 13.18 GET /advanced\_search

#### Advanced search

The fields you enter are case-sensitive and use exact match search. If you enter values in more than one field, these fields are combined using AND.

#### **Parameters**

Name	Туре	Required	Description	
name	string	False	Search for a value in Ric- graph field namename	max_nr_items string
category	string	False	Search for a value in Ric- graph field	Responses
			categorycate- gory	• 200: OK
value	string	False	Search for a value in Ric- graph field	• 250: Nothing found
max_nr_items	string	False	valuevalue The maximum number of items to return, or 0 to	• 251: Invalid search
lesponses			return an items	13.20 GET /get_all_neig
• 200: OK				
• 250: Nothing f	found			Find all neighbor nodes of th
• 251: Invalid se	arch			Parameters

# 13.19 GET /get\_all\_personroot\_nodes

Find all person-root nodes of this node

#### **Parameters**

Name	Туре	Required	Description
key	string	True	Search for a value in Ric- graph field *_key*
max_nr_items	string	False	The maximum number of items to return, or 0 to return all items

# hbor\_nodes

is node

Name	Туре	Required	Description
key	string	True	Search for a value in Ric- graph field * key*
name_want	array	False	Return only neighbor nodes whose field namename matches any value in the provided list; if the list is empty, return all neighbor nodes regard- less of their field namename
name_dontwant	array	False	Return only neighbor nodes whose field na- mename does notnot match any value in the provided list; if the list is empty, return all neighbor nodes regard- less of their field namename
category_want	array	False	Return only neighbor nodes whose field categorycate- gory matches any value in

#### Responses

- 200: OK
- 250: Nothing found
- 251: Invalid search

# 13.21 GET /get\_ricgraph\_list

Get the values of an internal global Ricgraph list

#### Parameters

Name	Туре	Required	Description
ricgraph_list	_nansetring	True	Return the values in the specified inter- nal Ricgraph list. These are dependent on the data in your Ricgraph in- stance and on the systems you have har- vested. Allowed Ricgraph lists are:

• name\_all: all possible values of the *name* field in a Ricgraph node.

• name\_personal\_all: all possible values of the *name* field that contain personal data in a Ricgraph node.

- category\_all: all possible values of the *category* field in a Ricgraph node.
- personal\_types\_all: all category values in list *category\_all* that are applicable to a person.

This is documentation for version 2.12 of Ricgraph – Research in context graph: It is **provided**s.ricgraph.eu. For more information about Ricgraph, Ricgraph Explorer, and the Ricgraph REST API, please read the reference publication doi.org/10.1016/j.softx.2024.101736, visit of version 2.12 of Ricgraph.eu, or go to the GitHub repository github.com/UtrechtUniversity/ricgraph.eu 78

- remainder\_types\_all: all other category values in list category\_all, that is all values in list category\_all minus those in personal\_types\_all.
- source\_all: the names of all the harvested source systems.
- resout\_types\_all: all research result types defined in file ricgraph.py. |

#### Responses

- 200: OK
- 250: Nothing found
- 251: Invalid search

This is documentation for version 2.12 of Ricgraph – Research in context graph. It is a pdf of docs.ricgraph.eu. For more information about Ricgraph, Ricgraph Explorer, and the Ricgraph REST API, please read the reference publication doi.org/10.1016/j.softx.2024.101736, visit the website www.ricgraph.eu, or go to the GitHub repository github.com/UtrechtUniversity/ricgraph.

# 14 Implementation details

On this page, you can find more information about:

- Person identifiers.
- The Person-root node in Ricgraph.
- Properties of nodes in Ricgraph.

You can also look at Example research questions.

Return to main README.md file.

## 14.1 Person identifiers

In the research world, persons can have any number of different identifiers. Some of these are standard, generally accepted and more-or-less unique identifiers over the lifetime of a person. These are called <u>persistent identi-</u> fiers. Others are non-unique, some are specific to an organization and some are specific to a company. Examples are:

- persistent identifiers: ORCID, ISNI;
- non-unique identifiers: full name (there are persons with the same name);
- organization identifiers: employee ID, email address (will change when a person leaves an organization);
- company identifiers: Scopus Author ID.

# 14.2 Person-root node in Ricgraph

Ricgraph uses a special node *person-root*. This node is connected to all the different person identifiers which have been harvested. *Person-root* "represents" a person. Research outputs from a person will also be connected to

this *person-root* node. The following figures shows two examples. *B* is the *person-root* node.



A person can have any number of identifiers. The person in the left figure has one *ORCID*, one *ISNI* and one *FULL\_NAME*. The person in the right figure has a lot more identifiers, and some identifiers appear more than once. E.g. this person has two different ORCIDs and two FULL\_NAMEs (with different spellings).

# 14.3 Properties of nodes in Ricgraph

All nodes in Ricgraph have the following properties:

- name: name of the node, e.g. ISNI, ORCID, DOI, FULL\_NAME, SCO-PUS\_AUTHOR\_ID, etc.;
- category: category of the node, e.g. person, person-root, book, journal article, data set, software, etc.;
- value: value of the node;
- \_key: key value of the node, not to be modified by the user;
- \_source: sorted list of sources a record has been harvested from, not to be modified by the user.

• \_history: list of history events of the node, not to be modified by the user.

Additional properties for nodes can be added by changing an entry in the Ricgraph initialization file. In the default configuration, the following properties are included:

- comment: comment for a node;
- year: year of a research output;
- url\_main: main URL for a node, pointing to e.g. the corresponding ISNI, ORCID or DOI record on the web;
- url\_other: other URL for a node, pointing to e.g. the originating record in the source system;
- source\_event: an event to be added to \_source.
- history\_event: an event to be added to \_history.

# **15** Ricgraph miscellaneous scripts

This page describes scripts for importing and exporting items from Ricgraph, about scripts to enhance (deleting personal data, renaming (sub-)organizations, etc.) information in Ricgraph, and about Ricgraph maintenance scripts. Read more about scripts for harvesting sources and inserting the results in Ricgraph, or about writing your own scripts.

#### On this page, you can find:

- Scripts for importing and exporting (directory *import\_export*):
  - Construct a Ricgraph from a csv file (construct\_ricgraph\_from\_csv)
  - Import nodes and edges from a csv file, raw version (ricgraph\_import\_raw\_from\_csv)
  - Export nodes and edges to a csv file, raw version (ricgraph\_export\_raw\_to\_csv)
  - Count the number of organizations that contributed to a category (count\_organizations\_contributed\_to\_category)
  - Export nodes to a file (export\_person\_identifiers and export\_person\_node\_properties)
- Scripts to enhance (finding, enriching, etc.) information (directory *enhance*):
  - Delete personal data from Ricgraph (delete\_personal\_data)
  - Rename (sub-)organizations in Ricgraph (rename\_organizations)
  - Script to enrich persons (enrich\_orcids\_scopusids)
  - Script to find person identifiers pointing to different persons (find\_double\_pids)
- Ricgraph maintenance scripts (directory maintenance):
  - Create a table of contents of the Ricgraph documentation (create\_toc\_documentation)

- Create an index of the Ricgraph documentation (create\_index\_documentation)
- Create the Ricgraph REST API documentation (convert\_openapi\_to\_mddoc)

All code is documented and hints to use it can be found in the source files.

Return to main README.md file.

# 15.1 Construct a Ricgraph from a csv file (construct\_ricgraph\_from\_csv)

To construct a Ricgraph from a csv file, use the script *con-struct\_ricgraph\_from\_csv.py*. You can find this script in the directory *import\_export*.

Nodes and edges are inserted with Ricgraph calls. That means, that nodes are inserted in pairs connected by an edge. *Person-root* nodes cannot be inserted with this script, they will be created whenever necessary. If a node in the nodes import file is not connected to another node by an edge in de the edge import file, it will not be created. This is due to the way Ricgraph works.

This script is different compared to Import nodes and edges from a csv file, raw version and Export nodes and edges to a csv file, raw version, since the "raw" scripts import and export with Cypher queries.

Usage:

construct\_ricgraph\_from\_csv.py [options]

Options:

--empty\_ricgraph <yes|no>

'yes': Ricgraph will be emptied before importing.

'no': Ricgraph will not be emptied before importing.

If this option is not present, the script will prompt the user what to do.

--filename <filename>

Import nodes and edges from a csv file starting with <filename>.

The file with nodes is <filename>-nodes.csv.

The file with edges is <filename>-edges.csv.

The import file containing nodes should be a csv file. At least the following columns should be present:

- name
- category
- value

Other fields that may be present:

• All fields in parameter *ricgraph\_properties\_additional* in the Ricgraph initialization file, but not the fields *source\_event* and *history\_event*.

The import file containing edges should be a csv file containing exactly four columns:

- name\_from, value\_from: the from node for the edge.
- name\_to, value\_to: the to node for the edge.

# 15.2 Import nodes and edges from a csv file, raw version (ricgraph\_import\_raw\_from\_csv)

To import nodes and edges from a csv file, use the script *ric-graph\_import\_raw\_from\_csv.py*. You can find this script in the directory *import\_export*.

This is a "raw" import, because *person-root* nodes are also imported, as are the connections between e.g. an *ORCID* node and its *person-root* node. When

you do the import, all nodes and edges will be inserted directly in the graph database backend using a Cypher query. That means that no checking is done at all if the resulting nodes and edges conform to the "Ricgraph model". This may result in a graph not consistent with the Ricgraph model. Due to this, Ricgraph Explorer may not work as expected.

This script forms a pair with Export nodes and edges to a csv file, raw version.

#### Usage:

ricgraph\_import\_raw\_from\_csv.py [options]

#### Options:

--empty\_ricgraph <yes|no>

'yes': Ricgraph will be emptied before importing.

'no': Ricgraph will not be emptied before importing.

If this option is not present, the script will prompt the user what to do.

--filename <filename>

Import nodes and edges from a csv file starting with <filename>.
The file with nodes is <filename>-nodes.csv.

The file with edges is <filename>-edges.csv.

The import file containing nodes should be a csv file. At least the following columns should be present:

- name
- category
- $\cdot$  value
- \_key

Other fields that may be present:

- The remaining fields in parameter *ricgraph\_properties\_hidden* in the Ric-graph initialization file.
- The fields in parameter *ricgraph\_properties\_additional* in the Ricgraph initialization file.

The import file containing edges should be a csv file containing exactly four columns:

- name\_from, value\_from: the from node for the edge.
- name\_to, value\_to: the to node for the edge.

For an example import file, export the nodes and edges in Ricgraph using Export nodes and edges to a csv file, raw version.

# 15.3 Export nodes and edges to a csv file, raw version (ricgraph\_export\_raw\_to\_csv)

To export nodes and edges to a csv file, use the script *ric-graph\_export\_raw\_to\_csv.py*. You can find this script in the directory *import\_export*.

This is a "raw" export, because person-root nodes are also exported, as are the connections between e.g. an ORCID node and its person-root node. The export is done using a Cypher query. When you import the export generated by this script, all nodes and edges will be inserted directly in the graph database backend using a Cypher query. That means that no checking is done at all if the resulting nodes and edges conform to the "Ricgraph model". This may result in a graph not consistent with the Ricgraph model. Due to this, Ricgraph Explorer may not work as expected.

This script forms a pair with Import nodes and edges from a csv file, raw version.

Usage: ricgraph\_export\_raw\_to\_csv.py [options]

#### Options:

--filename <filename>

Export all nodes and edges in Ricgraph to a csv file starting with <filename>. The file with nodes is <filename>-nodes.csv. The file with edges is <filename>-edges.csv. The export file containing nodes will be a csv file. All fields in Ricgraph will be exported.

The export file containing edges will be a csv file containing exactly four columns:

- name\_from, value\_from: the from node for the edge.
- name\_to, value\_to: the to node for the edge.

# 15.4 Count the number of organizations that contributed to a category (count\_organizations\_contributed\_to\_category)

To count the number of organizations that contributed to a category, use the script *count\_organizations\_contributed\_to\_category.py*. You can find this script in the directory *import\_export*.

This script counts the (sub-)organizations of persons who have contributed to all nodes of a specified category (e.g., *data set* or *software*). Both a histogram and a collaboration table will be computed and written to a file. The histogram contains the count of (sub-)organizations of all nodes of the specified category. The collaboration table contains the count of (sub-)organizations who have worked together on all nodes of the specified category.

What makes this script interesting, is that it also counts collaborations of suborganizations, if you have harvested them. For example, the Research Information System Pure contains a full organization hierarchy for persons. After harvesting Pure, Ricgraph contains this organization hierarchy. That is, not only the top level organization, such as *Utrecht University*, but also faculties, departments, units, and chairs. Using these sub-organizations, this script is able to show collaborations between e.g. different departments in the same organization. In case you have harvested organization hierarchies from different organizations, collaborations between e.g. departments of two universities can be shown. Usage:

count\_organizations\_contributed\_to\_category.py [options]

#### Options:

--sort\_organization <organization name>

Sort the collaboration table on this organization name.

If the name has one or more spaces, enclose it with "...".

If this option is not present, the script will prompt the user for a name.

#### --category <category>

Compute histogram and collaboration table for given category. If the name has one or more spaces, enclose it with "...". If this option is not present, the script will prompt the user for a name.

# 15.5 Export nodes to a file (export\_person\_identifiers and export\_person\_node\_properties)

[This is an old script, you might want to use Export nodes and edges to a csv file, raw version].

There are two scripts which allow to export *person* nodes to a csv file. These can be found in the directory *import\_export*.

- export\_person\_identifiers.py: exports all person identifiers connected to a person-root node.
- export\_person\_node\_properties.py: exports all node properties for every person node connected to a person-root node.

Use the parameter *EXPORT\_MAX\_RECS* for the number of records to export and *EXPORT\_FILENAME* for the filename to export at the start of both scripts.

# 15.6 Delete personal data from Ricgraph (delete\_personal\_data)

This script deletes all personal data of one or more persons from Ricgraph. The bash script *delete\_personal\_data.sh* is a wrapper for the Python script *delete\_pers\_data.py*. Both can be found in the directory *enhance*. These persons need to be listed in a csv file. An example csv file *pers\_to\_delete.csv* can also be found in directory *enhance*.

The script will delete all nodes of category *person* that are related with the person whose personal data have to be deleted (except for the *person-root* node, this node will not be deleted since it does not contain personal information).

Bash script:

Usage: delete\_personal\_data.sh [options]

Options:

-o, --organization [organization]

The organization to harvest. Specify the organization abbreviation.

-e, --empty\_ricgraph [yes|no]

Whether to empty Ricgraph before harvesting the

first organization. If absent, Ricgraph will not be emptied.

-c, --python\_cmd [python interpreter]

The python interpreter to use. If absent, and a python virtual environment is used, that interpreter is used.

-p, --python\_path [python path]

The value for PYTHONPATH, the path to python libraries.

- If absent, the current directory is used.
- -h, --help

Show this help text.

Usage:

delete\_pers\_data.py [options]

#### Options:

--filename <filename>

Specifies a csv file that has columns 'name' and 'value'. Every row in this file contains a personal identifier of a person whose personal data needs to be deleted from Ricgraph. --are\_you\_sure <yes>

Safety check since the script will delete items from Ricgraph. 'yes': This script will run.

any other value: This script will not run.

If this option is not present, the script will prompt the user whether to run the script.

The file *filename* contains identifiers for persons whose personal data have to be deleted from Ricgraph. It contains exactly two columns and can have as many rows as necessary. The columns are:

• name, value: values to identify the person in Ricgraph.

# 15.7 Rename (sub-)organizations in Ricgraph (rename\_organizations)

This script renames (sub-)organizations in Ricgraph that are listed in a csv file. The bash script *rename\_organizations.sh* is a wrapper for the Python script *rename\_orgs.py*. Both can be found in the directory *enhance*. Example corresponding csv files can also be found in directory *enhance*.

The script will read a line from the csv file. Then it will rename the (sub-)organization.

Bash script:

Usage: rename\_organizations.sh [options]

#### Options:

-o, --organization [organization]

The organization to harvest. Specify the organization abbreviation.

-e, --empty\_ricgraph [yes|no]

Whether to empty Ricgraph before harvesting the first organization. If absent, Ricgraph will not be emptied.

-c, --python\_cmd [python interpreter]

The python interpreter to use. If absent, and a python virtual environment is used, that interpreter is used.

#### -p, --python\_path [python path]

The value for PYTHONPATH, the path to python libraries. If absent, the current directory is used.

-h, --help

Show this help text.

#### Python script:

Usage: rename\_orgs.py [options]

#### Options:

--filename <filename>

Specifies a csv file that has columns 'orgname\_old' and 'orgname\_new'.

Every row in this file contains an organization name

that has to be renamed to a new organization name.

#### --are\_you\_sure <yes>

Safety check since the script will modify Ricgraph.

'yes': This script will run.

any other value: This script will not run.

If this option is not present, the script will prompt the user whether to run the script.

The file *filename* contains (sub-)organization names. It contains exactly two columns and can have as many rows as necessary. The columns are:

- orgname\_old: the old (sub-)organization name;
- orgname\_new: the new (sub-)organization name.

# 15.8 Script to enrich persons (enrich\_orcids\_scopusids)

With the script *enrich\_orcids\_scopusids.py*, you can enrich persons having an ORCID but no SCOPUS\_AUTHOR\_ID (using OpenAlex), or vice versa (using the Scopus API). Note that Scopus has a rate limit, and that you have to set some parameters in *ricgraph.ini*. You can find this script in the directory *enhance*.

# 15.9 Script to find person identifiers pointing to different persons (find\_double\_pids)

With the script *find\_double\_pids.py*, you can check if there are any personal identifiers that are pointing to two or more different persons. You can find this script in the directory *enhance*.

# 15.10 Create a table of contents of the Ricgraph documentation (create\_toc\_documentation)

To create a table of contents of the Ricgraph documentation use the script *create\_toc\_documentation.py*. You can find this script in the directory *maintenance*. The table of contents will be created in file ric-graph\_toc\_documentation.md.

Usage: create\_toc\_documentation.py

# 15.11 Create an index of the Ricgraph documentation (create\_index\_documentation)

To create an index of the Ricgraph documentation use the script *create\_index\_documentation.py*. You can find this script in the directory *maintenance*. The index will be created in file ric-graph\_index\_documentation.md.

Usage:

 $\verb|create_index_documentation.py||$ 

# 15.12 Create the Ricgraph REST API documentation (convert\_openapi\_to\_mddoc)

To create the Ricgraph REST API documentation use the script *convert\_openapi\_to\_mddoc.py*. This documentation is based on the Ricgraph OpenAPI yaml file *openapi.yaml* in the directory *ricgraph\_explorer/static*. You can find this script in the directory *maintenance*. The REST API documentation will be created in file ricgraph\_restapi\_gendoc.md.

Usage: convert\_openapi\_to\_mddoc.py

# 16 Ricgraph script writing

This page describes how to write scripts for Ricgraph. Read more about scripts for harvesting sources and inserting the results in Ricgraph, about scripts to import and export items from Ricgraph, or about scripts to enhance (finding, enriching, etc.) information in Ricgraph.

#### On this page, you can find:

- How to make your own harvesting scripts
- General program structure of a Python script using Ricgraph
- Structure of a Python script that is harvesting data
- Function calls for inserting nodes
- Function call for unifying personal identifiers
- Function calls to create, read (find), update and delete (CRUD) nodes
- Function calls to get neighbors of nodes

You can also adapt the code of Ricgraph in directory *ricgraph*, or the code of Ricgraph Explorer in directory *ricgraph\_explorer*.

All code is documented and hints to use it can be found in the source files.

Return to main README.md file.

# 16.1 How to make your own harvesting scripts

You can make your own harvesting script of your favorite source. The easiest way to do so is to take one of the harvesting scripts as an example. For example, if you use the script *harvest\_pure\_to\_ricgraph.py*, you'll recognize the three parts:

 Code for harvesting. This is done with harvest\_json\_and\_write\_to\_file() which also writes the harvested json data to a file. It gets data from a source.

- 2. Code for parsing. This is done with parse\_pure\_persons(), parse\_pure\_organizations() and parse\_pure\_resout() for persons, organizations and research outputs from Pure. It does data processing to get harvested results in a "useful" shape for inserting nodes and edges in Ricgraph.
- 3. Code for inserting the parsed results in Ricgraph. This is done with parsed\_persons\_to\_ricgraph(), parsed\_organizations\_to\_ricgraph() and parsed\_resout\_to\_ricgraph(). It inserts the nodes and edges in Ricgraph.

You can adapt each of these parts as suits the source you would like to harvest.

# 16.2 General program structure of a Python script using Ricgraph

#### import ricgraph as rcg

```
rcg.open_ricgraph()
rcg.empty_ricgraph() # use this only if you need to empty the graph
# some things happen
rcg.close_ricgraph()
```

# 16.3 Structure of a Python script that is harvesting data

This structure is used in the programming examples in the directory *har*-*vest*.

#### import ricgraph as rcg

rcg.open\_ricgraph() rcg.empty\_ricgraph() # use this only if you need to empty the graph

# Harvesting code: code to get data from a system

# Parsing code: post process the data found, and put it in a format that can easily be processed in Python, e.g. in a DataFrame

# Code to store the post processed results in Ricgraph

rcg.close\_ricgraph()

#### 16.4 Function calls for inserting nodes

Ricgraph stores objects and relations to objects. Therefore, most calls to insert nodes in have two nodes as parameter that are to be connected. Or two sets of nodes. Examples of these calls are (without the opening, emptying and closing of the graph):

```
rcg.read_node() # read (find) a node and return one
import ricgraph as rcg
                                                                              rcg.delete_node() # delete a node
  # example 1
rcg.create_two_nodes_and_edge() # create two nodes and connect with one edge
  # example 2
rcg.create_nodepairs_and_edges_df() # the same, now using a DataFrame to insert. 7 Function calls to get neighbors of nodes
                                     # a number of node pairs and their edges in one go
                                                                              There are several function calls to get neighbors of nodes. For a more exten-
  # example 3
                                                                              sive description how to use these, see the code comments in file ricgraph.py
rcg.create_nodepairs_and_edges_params() # the same, now using Python Dicts
                                                                             in directory ricgraph or the code examples in file ricgraph_explorer.py in direc-
```

# a number of node pairs and their

#### 16.5 Function call for unifying personal identifiers

Unification is the process of making sure that every personal identifier found for a certain person is connected to every other, via the person-root node. E.g., if there are four identifiers for a person: ORCID, ISNI, FULL NAME and SCO-PUS AUTHOR ID, they have to be unified pairwise. There is a function call to make this easier:

import ricgraph as rcg

rcg.unify\_personal\_identifiers() # takes a DataFrame with all identifiers to be un

# 16.6 Function calls to create, read (find), update and delete (CRUD) nodes

Of course, there are function calls to create, read, update and delete (CRUD) nodes. "Read" is used as term for "Find" or "Search".

```
import ricgraph as rcg
```

edges in an apple explorer.

rcg.create\_update\_node() # create or update a node rcg.read\_all\_nodes() # read (find) nodes and return all nodes found

This is documentation for version 2.12 of Ricgraph – Research in context graph. It is a pdf of docs.ricgraph.eu. For more information about Ricgraph, Ricgraph Explorer, and the Ricgraph REST API, 89 please read the reference publication doi.org/10.1016/j.softx.2024.101736, visit the website www.ricgraph.eu, or go to the GitHub repository github.com/UtrechtUniversity/ricgraph.

#### import ricgraph as rcg

rcg.get\_personroot\_node()
rcg.get\_all\_personroot\_nodes()
rcg.get\_all\_neighbor\_nodes()

# get a 'person-root' node starting from any 'person' node # get all 'person-root' nodes (there should be only one) # get all neighbor nodes connected to a node.

# it is possible to restrict to nodes having

# a certain property 'name' or 'category'

90

# 17 Ricgraph comparison

There are a number of approaches that collect research related information from various sources and combine them into one data structure. This section will give a short explanation of some of these approaches and how they compare to Ricgraph. This comparison was last updated in spring 2024.

Return to main README.md file.

## 17.1 General overview

All of these systems only store metadata, they do not store objects (e.g. pdfs, data files, software, etc.). Often they store the link to the object.

name		data struc- ture	how does it obtain data?	fields (enti- ties) in sys- tem	maturity
Ricgrap	h	graph	harvesting of any source, not only research related entities, requires creation or adaptation of harvest script	any field (config- urable), in standard config- uration: research outputs, people (with any identifier), organiza- tions	proof of con- cept
EOSC search discover graph	re- ry	probably graph	in develop- ment	in develop- ment	does not ex- ist yet
EOSC	PID	probably graph	in develop- ment	in develop- ment	does not ex- ist vet
Freya graph	PID	graph	unclear	unclear	proof of con- cept, project finished
Lens		relational	harvesting of many sources, e.g. Microsoft Academic, Crossref, ORCID, Pubmed, EPO, USPTO,	research outputs, people, publishers, patents, or- ganizations	mature

A big difference between on the one hand Lens, OpenAire graph, OpenAlex and Research.fi and on the other Ricgraph is their scale: the first group harvests a large number of sources. They also offer one single place of access for anyone. That also means that it will not be easy to extend these to your own needs (see next section), since one will need large computer facilities to host these systems and the data they contain.

Ricgraph follows a different approach: by selecting sources to harvest that are important to a user or organization, one is able to create a system that perfectly suits a certain information need of that person or organization. In creating harvest scripts, it is possible to harvest only that information that is relevant for a certain purpose. For example, one of the example harvest scripts harvests the Utrecht University staff pages. These pages cannot be harvested by other organizations due to the privileges required. Also, it is possible to harvest a source that is internal to an organization. Ricgraph can be installed on any internal or external accessible system according to your needs, so the data in Ricgraph is only accessible for persons of a certain organization, or for anyone.

The indecimentation of the provided in the reference publication doi.org/10.1016/j.softx.202/.101736, visit the website www.ricgraph.eu, or go to the GitHub repository github.com/UtrechtUniversity/ricgraph. 92

tutional, people, ordata, and ganizations,

# 17.2 Extendability

name	open source	extendabl	ecreate your own visual- izers or explor- ers	example code	harvest your own sources	additional infor- mation
Ricgraph	GitHub	yes	yes	yes	yes	read
EOSC re- search dis- covery graph	probably in the future					more in de- velop- ment, does not exist vet
EOSC PID graph	probably in the future					in de- velop- ment, does not exist vet
Freya PID graph	unknown	unknown	unknown	unknown	unknown	project has fin- ished
Lens	no	no	no	no	no	read
	_	_	_	_		more
OpenAire graph	unknown	unknown	probably not	probably not	no	read more
OpenAlex	GitHub	probably	probably	probably	no	read
		not	not	not		more
Research.	fiGitHub	probably	probably	probably	no	read
		not	not	not		more

This is documentation for version 2.12 of Ricgraph – Research in context graph. It is a pdf of docs.ricgraph.eu. For more information about Ricgraph, Ricgraph Explorer, and the Ricgraph REST API, please read the reference publication doi.org/10.1016/j.softx.2024.101736, visit the website www.ricgraph.eu, or go to the GitHub repository github.com/UtrechtUniversity/ricgraph.

As indicated above, systems such as Lens, OpenAire graph, OpenAlex and Research.fi are difficult to extend due to their size. For the Freya PID graph the author could not find information, and the EOSC research discovery graph and PID graph do not exist yet.

Ricgraph is easy to extend: the code is concise and can be found on GitHub. Also, it is possible to traverse the graph that has been constructed, either with Ricgraph Explorer, the Ricgraph REST API or with any other visualizer or explorer that someone builds. Also, Ricgraph can contain any field (entity) by changing the Ricgraph initialization file, creating a harvest script to fill this field, and modifying Ricgraph Explorer to show this field.

This is documentation for version 2.12 of Ricgraph – Research in context graph. It is a pdf of docs.ricgraph.eu. For more information about Ricgraph, Ricgraph Explorer, and the Ricgraph REST API, please read the reference publication doi.org/10.1016/j.softx.2024.101736, visit the website www.ricgraph.eu, or go to the GitHub repository github.com/UtrechtUniversity/ricgraph.

# 18 Ricgraph known bugs

## 18.1 Error while emptying Ricgraph with Neo4j

While deleting nodes and edges in Ricgraph you might get a Python error, similar to:

Deleting all nodes and edges in Ricgraph...

would use more than the limit [number] MiB. Currently using [number] MiB. dbms.memory.transaction.total.max threshold reached)

Traceback (most recent call last):

File "[path]/harvest/harvest\_pure\_to\_ricgraph.py",

line 729, in <module> rcg.empty\_ricgraph()

File "[path]/ricgraph/ricgraph.py", line [number], in empty\_ricgraph

\_graph.execute\_query('MATCH (node) DETACH DELETE node', database\_=GRAPHDB\_NAME) [...]

[module].errors.TransientError: [General.MemoryPoolOutOfMemoryError] The allocation of an extra 2.0 MiB would use more than the limit 716.8 MiB. Currently using 715.0 MiB. dbms.memory.transaction.total.max threshold reached.

This is caused by the free-to-use version of Neo4j, which does not have a statement similar to DROP [database]. If you empty the graph database, it needs to delete every node and edge and this fails when there are a lot of nodes and edges. This happens when you use Neo4j Desktop, and might happen sometimes when you are using Neo4j Community Edition.

# 18.1.1 Solution for Neo4j Community Edition

Read Empty a Neo4j Community Edition database.

## 18.1.2 Solution for Neo4j Desktop

You can do the following for Neo4j Desktop and Ricgraph:

- 1. Start Neo4j Desktop if it is not running yet.
- 2. Move your mouse to "Project" in the left column. A red trash can icon appears. Click it to remove the database. Confirm.
- 3. The text "No projects found" will appear. Create a project by clicking the upunder upunder and will be retried in [number]s (The allocation of an extra [number] (MiBran Durte t)
  - 4. The text "Project" appears with the text "Add a DBMS to get started". Click on the "+ Add" button next to it and select "Local DBMS". Leave the name as it is ("Graph DBMS") and fill in a password. Choose an easy to type and short one since the DBMS will only be accessible from your own machine. Click "Create". Also, insert the password in field graphdb\_password in the Ricgraph initialization file.

# 19 Table of contents Ricgraph documentation

This table of contents has been generated automatically.

- Ricgraph Research in context graph
  - What is Ricgraph?
  - Motivation
    - \* Use case for a journalist
    - \* Use case for a librarian
    - \* Use case for a researcher
  - Main contributions of Ricgraph
  - Why Ricgraph?
  - Examples
  - Ricgraph in bullet points
  - Next steps
    - \* Further information about Ricgraph
    - \* Steps to take if you would like to install Ricgraph and harvest data
    - \* Steps to take if you would like to use Ricgraph
  - Contact
- Ricgraph as a server on Linux
  - Fast and recommended way to install Ricgraph as a server
  - Run Ricgraph scripts from the command line or as a cronjob
    - \* Using the Makefile
    - \* In case you have installed Ricgraph as a server
    - \* In case you have installed Ricgraph for a single user
  - Use a service unit file to run Ricgraph Explorer and the Ricgraph REST API
  - Use Apache, WSGI, and ASGI to make Ricgraph Explorer and the Ricgraph REST API accessible from outside your virtual machine

- \* Introduction Apache webserver
- \* Installation Apache
- \* Post-install steps Apache
- Use Nginx, WSGI, and ASGI to make Ricgraph Explorer and the Ricgraph REST API accessible from outside your virtual machine
  - \* Introduction Nginx webserver
  - \* Installation Nginx
  - \* Post-install steps Nginx
- How to install Ricgraph and Ricgraph Explorer on SURF Research Cloud
  - \* Preliminaries
  - \* Create a SURF Research Cloud workspace
  - \* Install Ricgraph in a SURF Research Cloud workspace
  - \* Pause and resume a SURF Research Cloud workspace
  - \* Access a SURF Research Cloud workspace
- Steps to take to install Ricgraph as a server by hand
  - \* Install your graph database backend
  - \* Create a ricgraph user and group
  - \* Create a Python virtual environment and install Ricgraph in it
- Ricgraph comparison
  - General overview
  - Extendability
- Ricgraph in a container
  - Notes on the Ricgraph Podman container
  - Install Podman
  - Install and run the Ricgraph Podman container
  - Advanced use of the Ricgraph Podman container

- Implementation details
  - Person identifiers
  - Person-root node in Ricgraph
  - Properties of nodes in Ricgraph
- Ricgraph Explorer
  - How to start Ricgraph Explorer
  - Working with Ricgraph Explorer
    - \* A usage flow through Ricgraph Explorer
    - \* Example research questions
  - Home page
  - Search page
  - Person options page
  - Organization options page
  - Results page
  - Browsing Ricgraph
  - Showing a privacy statement or privacy measures document
  - Video Ricgraph Explorer
- Index Ricgraph documentation
- Ricgraph known bugs
  - Error while emptying Ricgraph with Neo4j
    - \* Solution for Neo4j Community Edition
    - \* Solution for Neo4j Desktop
- Ricgraph REST API
  - Installation of the Ricgraph REST API
  - Use of the Ricgraph REST API
  - Ricgraph REST API endpoint documentation
  - Generate the Ricgraph REST API documentation page
- Ricgraph Research in context graph REST API
  - What to find on this page?
  - How to use
  - Technicalities
  - Base URL

- Endpoints
- GET /person/search
- GET /person/all\_information
- GET /person/share\_research\_results
- GET /person/collaborating\_organizations
- GET /person/enrich
- GET /organization/search
- GET /organization/all\_information
- GET /organization/information\_persons\_results
- GET /organization/enrich
- GET /competence/search
- GET /competence/all\_information
- GET /broad\_search
- GET /advanced\_search
- GET /get\_all\_personroot\_nodes
- GET /get\_all\_neighbor\_nodes
- GET /get\_ricgraph\_list
- Ricgraph script writing
  - How to make your own harvesting scripts
  - General program structure of a Python script using Ricgraph
  - Structure of a Python script that is harvesting data
  - Function calls for inserting nodes
  - Function call for unifying personal identifiers
  - Function calls to create, read (find), update and delete (CRUD) nodes
  - Function calls to get neighbors of nodes
- Table of contents Ricgraph documentation
- Ricgraph with Memgraph graph database backend
  - Install and start Memgraph
- Install and configure Ricgraph
  - Fast and recommended way to install Ricgraph for a single user
    - \* You can change to user root
    - \* You cannot change to user root
  - Requirements for Ricgraph

- Ricgraph Makefile
- Ricgraph initialization file
  - \* Settings for the graph database backend
  - \* Extending Ricgraph with new properties in the nodes
  - \* RICGRAPH\_NODEADD\_MODE
- Ricgraph on Windows
- Steps to take to install Ricgraph for a single user by hand
  - \* Install your graph database backend
  - \* Download Ricgraph
  - \* Use a Python virtual environment and install Python requirements
    - · Using Python's venv module
    - $\cdot$  Using a Python Integrated development environment (IDE)
- Documentation Ricgraph Research in context graph
  - What is Ricgraph?
  - What to find on the Documentation website for Ricgraph?
  - Contact
- Tutorial Ricgraph Research in context graph
  - Ricgraph introduction
  - Motivation
    - \* Use case for a journalist
    - \* Use case for a librarian
    - \* Use case for a researcher
  - Main contributions of Ricgraph
  - Installation guide
  - Harvesting source systems
  - Browsing using Ricgraph Explorer
    - \* A usage flow through Ricgraph Explorer
    - \* Example research questions
  - More information
  - Contact
- Ricgraph publications, presentations, use, and mentions

- Ricgraph publications
- Ricgraph presentations
- Ricgraph newsletters
- Ricgraph projects with students
- Ricgraph use
- Ricgraph mentions
- Arjan Sieverink presents Ricgraph at Pure International Conference 2023
- Ricgraph harvest scripts
  - Introduction to harvest scripts
  - Organization abbreviation
  - Scripts that harvest multiple sources
  - Scripts that harvest a single source
    - \* Harvest of OpenAlex (harvest\_openalex\_to\_ricgraph)
    - \* Harvest of Pure (harvest\_pure\_to\_ricgraph)
      - $\cdot\,$  Limit the amount of data to harvest from Pure
      - Pure READ and Pure CRUD API
      - Pure harvesting of projects
    - \* Harvest of data sets from Yoda-DataCite (harvest\_yoda\_datacite\_to\_ricgraph)
    - \* Harvest of Utrecht University staff pages (harvest\_uustaffpages\_to\_ricgraph)
    - \* Harvest of software from the Research Software Directory (harvest\_rsd\_to\_ricgraph)
  - Order of running the harvest scripts
  - How to make your own harvesting scripts
- Ricgraph miscellaneous scripts
  - Construct a Ricgraph from a csv file (construct\_ricgraph\_from\_csv)
  - Import nodes and edges from a csv file, raw version (ricgraph\_import\_raw\_from\_csv)
  - Export nodes and edges to a csv file, raw version (ricgraph\_export\_raw\_to\_csv)

- Count the number of organizations that contributed to a category (count\_organizations\_contributed\_to\_category)
- Export nodes to a file (export\_person\_identifiers and export\_person\_node\_properties)
- Delete personal data from Ricgraph (delete\_personal\_data)
- Rename (sub-)organizations in Ricgraph (rename\_organizations)
- Script to enrich persons (enrich\_orcids\_scopusids)
- Script to find person identifiers pointing to different persons (find\_double\_pids)
- Create a table of contents of the Ricgraph documentation (create\_toc\_documentation)
- Create an index of the Ricgraph documentation (create\_index\_documentation)
- Create the Ricgraph REST API documentation (convert\_openapi\_to\_mddoc)
- Ricgraph with Neo4j graph database backend
  - Neo4j Community Edition
    - \* Install and start Neo4j Community Edition
    - \* Post-install steps Neo4j Community Edition
  - Neo4j Desktop
    - \* Install Neo4j Desktop
    - \* Post-install steps Neo4j Desktop
    - \* Start Neo4j Desktop
    - \* Neo4j Desktop with Bloom (optional)
      - · Install Bloom configuration for Neo4j Desktop
      - $\cdot$  Open Bloom
      - Execute queries in Bloom
      - · Actions while clicking on a node in Bloom
      - · Visualization of nodes in Bloom
  - Dumping, restoring, and emptying the Ricgraph database
    - \* Create a Neo4j Desktop database dump of Ricgraph
    - \* Create a Neo4j Community Edition database dump of Ricgraph
    - \* Restore a Neo4j Desktop database dump of Ricgraph in Neo4j Desktop

- \* Restore a Neo4j Desktop database dump of Ricgraph in Neo4j Community Edition
- \* Restore a Neo4j Community Edition database dump of Ricgraph in Neo4j Community Edition
- \* Empty a Neo4j Community Edition database
- How to reset the Neo4j Community Edition password
- How to solve an AttributeError: Neo4jDriver object has no attribute executequery

# 20 Index Ricgraph documentation

This index has been generated automatically.

#### А

- A usage flow through Ricgraph Explorer
- A usage flow through Ricgraph Explorer
- Access a SURF Research Cloud workspace
- Actions while clicking on a node in Bloom
- Advanced use of the Ricgraph Podman container
- Arjan Sieverink presents Ricgraph at Pure International Conference 2023

#### B

- Base URL
- Browsing Ricgraph
- Browsing using Ricgraph Explorer

#### С

- Construct a Ricgraph from a csv file (construct\_ricgraph\_from\_csv)
- Contact
- Contact
- Contact
- Count the number of organizations that contributed to a category (count\_organizations\_contributed\_to\_category)
- Create a Neo4j Community Edition database dump of Ricgraph
- Create a Neo4j Desktop database dump of Ricgraph
- Create a Python virtual environment and install Ricgraph in it
- Create a ricgraph user and group
- Create a SURF Research Cloud workspace
- Create a table of contents of the Ricgraph documentation (create\_toc\_documentation)

- Create an index of the Ricgraph documentation (create\_index\_documentation)
- Create the Ricgraph REST API documentation (convert\_openapi\_to\_mddoc)

#### D

- Delete personal data from Ricgraph (delete\_personal\_data)
- Documentation Ricgraph Research in context graph
- Download Ricgraph
- Dumping, restoring, and emptying the Ricgraph database

#### Е

- Empty a Neo4j Community Edition database
- Endpoints
- Error while emptying Ricgraph with Neo4j
- Example research questions
- Example research questions
- Examples
- Execute queries in Bloom
- Export nodes and edges to a csv file, raw version (ricgraph\_export\_raw\_to\_csv)
- Export nodes to a file (export\_person\_identifiers and export\_person\_node\_properties)
- Extendability
- Extending Ricgraph with new properties in the nodes
- F
- Fast and recommended way to install Ricgraph as a server
- Fast and recommended way to install Ricgraph for a single user
- Function call for unifying personal identifiers

- Function calls for inserting nodes
- Function calls to create, read (find), update and delete (CRUD) nodes
- Function calls to get neighbors of nodes
- Further information about Ricgraph

## G

- General overview
- General program structure of a Python script using Ricgraph
- Generate the Ricgraph REST API documentation page
- GET /advanced\_search
- GET /broad\_search
- GET /competence/all\_information
- GET /competence/search
- GET /get\_all\_neighbor\_nodes
- GET /get\_all\_personroot\_nodes
- GET /get\_ricgraph\_list
- GET /organization/all\_information
- GET /organization/enrich
- GET /organization/information\_persons\_results
- GET /organization/search
- GET /person/all\_information
- GET /person/collaborating\_organizations
- GET /person/enrich
- GET /person/search
- GET /person/share\_research\_results

## Н

- Harvest of data sets from Yoda-DataCite (harvest\_yoda\_datacite\_to\_ricgraph)
- Harvest of OpenAlex (harvest\_openalex\_to\_ricgraph)
- Harvest of Pure (harvest\_pure\_to\_ricgraph)
- Harvest of software from the Research Software Directory (harvest\_rsd\_to\_ricgraph)
- Harvest of Utrecht University staff pages (harvest\_uustaffpages\_to\_ricgraph)
- Harvesting source systems

- Home page
- How to install Ricgraph and Ricgraph Explorer on SURF Research Cloud
- How to make your own harvesting scripts
- How to make your own harvesting scripts
- + How to reset the Neo4j Community Edition password
- How to solve an AttributeError: Neo4jDriver object has no attribute executequery
- How to start Ricgraph Explorer
- How to use
- Ι
  - Implementation details
  - Import nodes and edges from a csv file, raw version (ricgraph\_import\_raw\_from\_csv)
  - In case you have installed Ricgraph as a server
  - In case you have installed Ricgraph for a single user
  - Index Ricgraph documentation
- Install and configure Ricgraph
- Install and run the Ricgraph Podman container
- Install and start Memgraph
- Install and start Neo4j Community Edition
- Install Bloom configuration for Neo4j Desktop
- Install Neo4j Desktop
- Install Podman
- Install Ricgraph in a SURF Research Cloud workspace
- Install your graph database backend
- Install your graph database backend
- Installation Apache
- Installation guide
- Installation Nginx
- Installation of the Ricgraph REST API
- Introduction Apache webserver
- Introduction Nginx webserver
- Introduction to harvest scripts

L

• Limit the amount of data to harvest from Pure

#### М

- Main contributions of Ricgraph
- Main contributions of Ricgraph
- More information
- Motivation
- Motivation

#### Ν

- Neo4j Community Edition
- Neo4j Desktop
- Neo4j Desktop with Bloom (optional)
- Next steps
- Notes on the Ricgraph Podman container

#### 0

- Open Bloom
- Order of running the harvest scripts
- Organization abbreviation
- Organization options page

#### Р

3 Ricgrap

- Pause and resume a SURF Research Cloud workspace
- Person identifiers
- Person options page
- Person-root node in Ricgraph
- Post-install steps Apache
- Post-install steps Neo4j Community Edition
- Post-install steps Neo4j Desktop
- Post-install steps Nginx
- Preliminaries
- Properties of nodes in Ricgraph
- Pure harvesting of projects
- Pure READ and Pure CRUD API

## R

- Rename (sub-)organizations in Ricgraph (rename\_organizations)
- Requirements for Ricgraph
- Restore a Neo4j Community Edition database dump of Ricgraph in Neo4j Community Edition
- Restore a Neo4j Desktop database dump of Ricgraph in Neo4j Community Edition
- Restore a Neo4j Desktop database dump of Ricgraph in Neo4j Desktop
- Results page
- Ricgraph Research in context graph
- Ricgraph Research in context graph REST API
- Ricgraph as a server on Linux
- Ricgraph comparison
- Ricgraph Explorer
- Ricgraph harvest scripts
- Ricgraph in a container
- Ricgraph in bullet points
- Ricgraph initialization file
- Ricgraph introduction
- Ricgraph known bugs
- Ricgraph Makefile
- Ricgraph mentions
- Ricgraph miscellaneous scripts
- Ricgraph newsletters
- Ricgraph on Windows
- Ricgraph presentations
- Ricgraph projects with students
- Ricgraph publications
- Ricgraph publications, presentations, use, and mentions
- Ricgraph REST API
- Ricgraph REST API endpoint documentation
- Ricgraph script writing
- Ricgraph use
- Ricgraph with Memgraph graph database backend
- Ricgraph with Neo4j graph database backend
- RICGRAPH\_NODEADD\_MODE

- Run Ricgraph scripts from the command line or as a cronjob
- S
- Script to enrich persons (enrich\_orcids\_scopusids)
- Script to find person identifiers pointing to different persons (find\_double\_pids)
- Scripts that harvest a single source
- Scripts that harvest multiple sources
- Search page
- Settings for the graph database backend
- Showing a privacy statement or privacy measures document
- Solution for Neo4j Community Edition
- Solution for Neo4j Desktop
- Start Neo4j Desktop
- · Steps to take if you would like to install Ricgraph and harvest data
- Steps to take if you would like to use Ricgraph
- Steps to take to install Ricgraph as a server by hand
- Steps to take to install Ricgraph for a single user by hand
- Structure of a Python script that is harvesting data

#### т

- Table of contents Ricgraph documentation
- Technicalities
- Tutorial Ricgraph Research in context graph

#### U

- Use a Python virtual environment and install Python requirements
- Use a service unit file to run Ricgraph Explorer and the Ricgraph REST API
- Use Apache, WSGI, and ASGI to make Ricgraph Explorer and the Ricgraph REST API accessible from outside your virtual machine
- Use case for a journalist
- Use case for a journalist
- Use case for a librarian
- Use case for a librarian
- Use case for a researcher

- Use case for a researcher
- Use Nginx, WSGI, and ASGI to make Ricgraph Explorer and the Ricgraph REST API accessible from outside your virtual machine
- Use of the Ricgraph REST API
- Using a Python Integrated development environment (IDE)
- Using Python's venv module
- Using the Makefile

## V

- Video Ricgraph Explorer
- Visualization of nodes in Bloom

## w

- What is Ricgraph?
- What is Ricgraph?
- What to find on the Documentation website for Ricgraph?
- What to find on this page?
- Why Ricgraph?
- Working with Ricgraph Explorer

# Y

- You can change to user root
- You cannot change to user root

