**Technical**

**MANUAL**

*GRAVITATE Semantic Enrichment Tool*

Release: v1

Date: *23-11-2018*

*Authors: Stuart E. Middleton*

[[1]](#footnote-1)

# TABLE OF CONTENTS

Contents

[1 TABLE OF CONTENTS 2](#_Toc531707399)

[2 GENERAL INFORMATION 3](#_Toc531707400)

[2.1 Release notes 3](#_Toc531707401)

[2.2 System Overview 3](#_Toc531707402)

[2.3 Authorized Use Permission 3](#_Toc531707403)

[2.4 Points of Contact and Support 3](#_Toc531707404)

[3 Getting Started 4](#_Toc531707405)

[3.1 System prerequisites and software download 4](#_Toc531707406)

[3.1.1 3rd party software pre-requisites necessary to install the module 4](#_Toc531707407)

[3.1.2 Distribution 5](#_Toc531707408)

[4 Installation 6](#_Toc531707409)

[5 System Usage 7](#_Toc531707410)

# GENERAL INFORMATION

## Release notes

* Release v1 has been developed within the GRAVITATE project, funded by the European Union’s Horizon 2020 research and innovation programme under grant agreement 665155.
* IP and usage rights are described in the related section of this document.

## System Overview

The semantic enrichment command line tool will take a set of artifact free text descriptions and generate a set of CIDOC-CRM assertions, encoded using RDF, suitable for ingestion into the GRAVIATE database via the GRAVITATE ETL process.

## Authorized Use Permission

The GRAVITATE semantic enrichment tool has a BSD license with an attribution clause and restrictions to research, education or evaluation purposes. A commercial license is available royalty free on request.

*Copyright (c) 2018, University of Southampton All rights reserved. Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:*

1. *Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.*
2. *Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.*
3. *All advertising and publication materials containing results from the use of this software must acknowledge the University of Southampton:*
4. *Neither the name of the University of Southampton nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.*
5. *This software should only be used for research, education or evaluation purposes. A royalty free commercial license is available on request to sem03@soton.ac.uk.*

*THIS SOFTWARE IS PROVIDED BY University of Southampton ''AS IS'' AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL University of Southampton BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWATE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.*

## Points of Contact and Support

Contact email: sem03@soton.ac.uk

Type of support: Best Efforts Only

# Getting Started

## System prerequisites and software download

The following instructions need to be followed to prepare your system for installing the GRAVITATE software.

### 3rd party software pre-requisites necessary to install the module

R v3.4.4

The R Foundation

GPL v2

https://cran.r-project.org/

Stanford POS tagger

Stanford University

GPL v2

https://nlp.stanford.edu/software/tagger.shtml

Stanford Parser

Stanford University

GPL v2

https://nlp.stanford.edu/software/tagger.shtml

Python 2.7.8

Python Software Foundation

Python Software Foundation License

http://www.python.org/psf/license

numpy 1.13.1+mkl

Author: NumPy Developers

BSD

http://www.numpy.org

nltk 3.2.1

NLTK Project

Apache License Version 2.0

http://www.nltk.org

openiepy 0.1.0, lexicopy 0.1.0, soton\_corenlppy 0.1.0

University of Southampton

BSD variant

Shipped with release

### Distribution

https://github.com/GRAVITATE-EU/semantic-enrichment

# Installation

Copy the GRAVITATE semantic-enrichment release files to <install dir>

Install Python 2.7 and Pip

Install Python lib NLTK 3.2.1 (download all corpus data)

Install Python lib Numpy 1.13.1+mk1

Install soton\_corenlppy, lexicopy, openiepy

cd <install dir>

pip install --upgrade --upgrade-strategy “only-if-needed” soton\_corenlppy-0.1.0-py27-none-any.whl

pip install --upgrade --upgrade-strategy “only-if-needed” lexicopy-0.1.0-py27-none-any.whl

pip install --upgrade --upgrade-strategy “only-if-needed” openiepy-0.1.0-py27-none-any.whl

Install R v3.4.4 >> C:\Program Files\R\R-3.4.4

Install Stanford POS tagger >> C:\stanford-parser-full

Install Stanford parser >> C:\stanford-postagger-full

Install Stanford english model >> C:\stanford-postagger-full\stanford-english-corenlp-2016-10-31-models.jar

Note: Installation of pre-requisites to other locations than specified will require edits to the file ch\_information\_extraction\_app.ini

# System Usage

Create your input data by running a SPARQL query against GRAVIATE BlazeGraph

curl -X POST http://localhost:9999/blazegraph/sparql --data-urlencode query@query-artifact-text.txt -H 'Accept:application/json;charset=UTF-8' -o "artifact-text.json"

Copy artifact-text.json to <install dir>

cd <install dir>

python ch\_information\_extraction\_app.py ch\_information\_extraction\_app.ini

The output file productions.trig will be generated in the <install dir>

1. [↑](#footnote-ref-1)