

Ferramentas de descrição e anotação no workflow GDI: caso do estudo com o Dendro e serviços EUDAT

Yulia Karimova^a, João Aguiar Castro^a, João Rocha da Silva^a

^aINESC TEC, Faculdade de Engenharia, Universidade do
Porto, {ylaleo, joaoaguiarcastro, joaorosilva} @gmail.com

F
Findable

A
Accessible

I
Interoperable

R
Reusable



Plataforma Dendro

Dendro^β Find ▾

Plugins ▾

Search Dendro

SUBMIT

Login

Welcome to Dendro Beta

This Dendro is in Demo Mode!

This is a demo instance of Dendro. To try out the platform please use the following credentials:

Username : demouser1

Password : demouserpassword2015

If you have any issues or enquiries, please contact the researchers at the [Dendro home page](#).

A linked, open and collaborative research data management platform.



Read papers

Find relevant bibliography by searching on high-quality metadata. The papers and other documents are contributed by researchers like yourself.



Browse datasets

Find high-quality, community-managed datasets, including domain-specific descriptions



Contact researchers

Contact researchers who shared data on this Dendro

Welcome to the INESC TEC research data repository.

This data repository showcases datasets produced or used by INESC TEC researchers and their partners. It is an embodiment of our institutional commitment to Open Data in research.



This is a featured section

CS: Computer Science

The Computer Science Cluster mission is to...

Electric field data from ENVRiplus TNA campaign RELECT at SMEAR II – HYYTIÄLÄ...

Vertical electric field measurements (in V/m), 1-min averages. Under construction (weekly datafiles added as the...)

Radon data from ENVRiplus TNA campaign RELECT at SMEAR II – HYYTIÄLÄ multi-di...

Radon concentration measurements (in Bq/m³) every 2-hours. Under construction (weekly datafiles added as the campaign...)

About CKAN

CKAN API

Open Knowledge Foundation

[OPEN DATA](#)

Search data

E.g.



Search records for...

SEARCH

[HELP](#) [COMMUNITIES](#) [UPLOAD](#) [CONTACT](#)

Login

Store and publish your research data

Search in public datasets or register as a user to upload and publish your data!

Please use <https://trng-b2share.eudat.eu> for testing or training.

[Login or Register](#)

Create Record

Create a new record

Latest Records

1-D beams crossing a relaxed simulated gal.cluster

13 Oct 2017 by [Franco Vazza](#);
150 lines of sight through the volume of a simulated relaxed galaxy cluster at z=0, produced by FVazza et al. running ENZO-MHD on the Jureca supercluster in Juelich. Each line is simple tabulated dat

Simulated cosmic volume - dynamo amplification

11 Oct 2017 by [Franco Vazza](#);
This is one subset (300³ cells) derived from a larger 1024³ cosmological MHD simulation produced using ENZO on Piz-Daint (CSCS) by FVazza& collaborators. In these 'dynamo' models we impose a negl

Simulated cosmic volume - Primordial magnetic seeding models.

11 Oct 2017 by [Franco Vazza](#);
These are 2 subsets (300³ cells each) derived from larger 1024³ cosmological MHD simulations produced using ENZO on Piz-Daint (CSCS) by FVazza& collaborators. In these 'primordial' models we impose

EUDAT Services Terms of Use v1.1

27 Sep 2017 by [Rouchon, Olivier](#); [Massol, Marion](#); [Kaila, Urpo](#); [Elbers, Willem](#); [Thiemann, Hannes](#); [Brzezniak, Maciej](#); [van de Sanden, Mark](#); [Reetz, Johannes](#);
This document describes the Terms of Use (ToU) of EUDAT resources and services. Before using any EUDAT services and resources each user must have read and agreed to understand and follow these terms.

Soil_moisture_LTER_EU_SP_001_Doñana_LTSE

21 Aug 2017 by [Ricardo Diaz-Deigado](#);

1-D beams crossing a perturbed simulated gal.cluster

13 Oct 2017 by [Franco Vazza](#);
150 lines of sight through the volume of a simulated galaxy cluster at z=0, produced by FVazza et al. running ENZO-MHD on the Jureca supercluster in Juelich. Each line is simple tabulated dataset con

Simulated cosmic volume - astrophysical seeding scenario

11 Oct 2017 by [Franco Vazza](#);
These are 3 subsets (300³ cells each) derived from larger 1024³ cosmological MHD simulations produced using ENZO on Piz-Daint (CSCS) by FVazza& collaborators. In these 'astrophysical' models we imp

Statistics of star formation and magnetic seeding

10 Oct 2017 by [Franco Vazza](#);
Tabular data for the cosmic star formation history in cosmological simulations, and of the relative magnetisation output of stellar activity. For meaning of the models please refer to <http://cosmosimf>

EUDAT DB.2: Report on technology watch

26 Aug 2017 by [Le Franc, Yann](#); [Dima, Emanuel](#); [Rajapakse, Asela](#); [Schentz, Herbert](#); [Goldfarb, Doron](#); [Wohner, Christoph](#); [Queral, Anna](#); [Miranda, Alberto](#); [Cortes, Toni](#);
W/P8 aims at developing proof-of-concept services and data models to complement and extend the EUDAT service offer to fill the gaps in supporting communities Data Life Cycles (DLCs). Part of these gaps

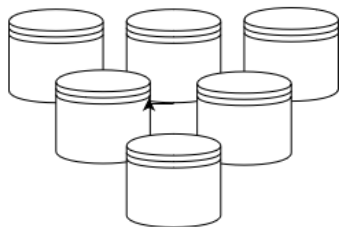
asdasd

7 Aug 2017

Workflow GDI com a plataforma Dendro e os serviços do EUDAT B2SHARE e B2NOTE

STEP 1

Data Production

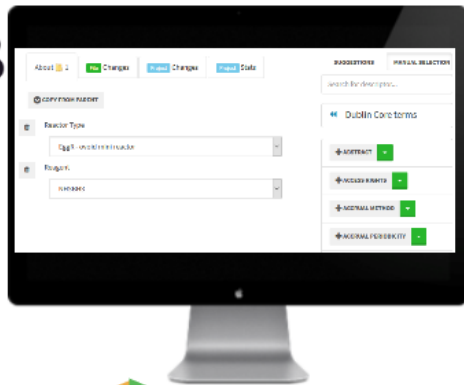


STEP 2

Deposit



Description



STEP 3

Publication



STEP 4

Annotation





Folder



Selection



My Projects

open-nlp.zip

stanford-corenlp.zip

nltk.zip

 spacy.zip **2**

About HAREM%20NER%20MODELS

Changes

Changes

Stats

COPY FROM PARENT

IN MANUAL MODE ▾

Extent

106MB

License

Creative Commons Attribution Share-Alike

Description

B**I****H****3**

Pre-trained models for named entity recognition in Portuguese, using the categories, types and subtypes of the Second HAREM dataset as entity classes.

SUGGESTIONS

MANUAL SELECTION

Search for descriptor...

Dublin Core terms

ABSTRACT ▾

ACCESS RIGHTS ▾

ACCRUAL METHOD ▾

ACCRUAL PERIODICITY ▾

 ACCRUAL POLICY ▾ **4**

ALTERNATIVE TITLE ▾

My repository bookmarks

SELECT A DESTINATION REPOSITORY ▾



Connect to a new repository

CHOOSE REPOSITORY TYPE ▾

Connect to B2Share

Label

trngB2share



This will be used to label these credentials and address as one of your favorite repositories

External repository address

trng-b2share.eudat.eu



The EUDAT Test deposit server. Use this to test your deposit prior to publishing. It will be deleted after a while, so do not use this for permanent data deposit.

Access Token

MmGKBzjpdIT382lag38zxhsKttZDw9e7u6zZmzucVFUu1aYM5i55WpeUSgFE|



Your B2Share token

CANCEL

SAVE NEW BOOKMARK

RECORDS - A4906773DC1F42F882BD03BE0C9846C3

HAREM NER Models for OpenNLP, Stanford CoreNLP, spaCy, NLTK

by [André Pires](#);

Jun 22, 2017

Keywords: [named entity recognition](#); [models](#); [text mining](#); [portuguese](#);

DOI: [XXXX/b2share.a4906773dc1f42f882bd03be0c9846c3](#)

PID: [0000/a4906773dc1f42f882bd03be0c9846c3](#)

Files

Name	Size
> HAREM NER MODELS.json	7.01KB
> HAREM NER MODELS.rdf	2.04KB
> HAREM NER MODELS.txt	2.10KB
> HAREM NER MODELS.zip	111.61MB
> nltk.zip	5.76MB
> open-nlp.zip	2.41MB

1

```
dcterms:http://purl.org/dc/terms/
```

```
./HAREM NER MODELS
```

```
+Extent(dcterms:extent) : 106MB
```

```
+License(dcterms:license) : Creative Commons Attribution Share-
```

```
+Description(dcterms:description) : Pre-trained models for name
```

```
+Publisher(dcterms:publisher) : INESC TEC
```

```
+Title(dcterms:title) : HAREM NER Models for OpenNLP, Stanford
```

```
+Creator(dcterms:creator) : André Pires
```

```
+Format(dcterms:format) : *.zip
```

```
+Language(dcterms:language) : PT
```

```
+Relation(dcterms:relation) : Master's thesis: PIRES, André (20
```

```
+Subject(dcterms:subject) : portuguese
```

```
+Subject(dcterms:subject) : models
```

```
+Subject(dcterms:subject) : text mining
```

```
+Subject(dcterms:subject) : named entity recognition
```

```
+Type(dcterms:type) : Serialized Models for OpenNLP, Stanford C
```

```
--.spacy.zip
```

```
--.open-nlp.zip
```

```
--.stanford-corenlp.zip
```

```
--.nltk.zip
```

```
+Description(dcterms:description) : NER models for NLTK, for  
classifier: DecisionTree entropy_cutoff (0.08, 0.09), DecisionTre
```

```
dcterms:
```

2

HAREM NER Models for OpenNLP, Stanford CoreNLP, spaCy,

by André Pires;

Jun 22, 2017

Keywords: named entity recognition; models; text mining; portuguese;

DOI: [XXXX/b2share.a4906773dc1f42f882bd03be0c9846c3](https://doi.org/10.21203/b2share.a4906773dc1f42f882bd03be0c9846c3)

PID: [0000/a4906773dc1f42f882bd03be0c9846c3](https://purl.org/urn:uuid:0000/a4906773dc1f42f882bd03be0c9846c3)

Name	Size
> HAREM NER MODELS.json	7.01KB
> HAREM NER MODELS.rdf	2.04KB
> HAREM NER MODELS.txt	2.10KB
> HAREM NER MODELS.zip	111.61MB
> nltk.zip	5.76MB
▼ open-nlp.zip	2.41MB
Checksum: md5:677f1c21c83c7fbf5420626f5370a447	
PID: 0000/open-nlp.zip <input type="button" value="Copy"/>	
<input type="button" value="Annotate in B2Note"/> 1	
> spacy.zip	15.72MB
> stanford-corenlp.zip	87.75MB

Basic metadata	
Open Access	True <input checked="" type="checkbox"/>
Publisher	INESC
Language	PT

The screenshot shows the B2NOTE interface with a modal window open. The modal has a dark blue header with the B2NOTE logo and a 'PILOT' badge. Below the header, there are navigation icons for user profile (Yulia), search, download, share, and help. The main content area of the modal has two tabs: 'Semantic tag' (selected) and 'Free-text keyword'. Below the tabs is a 'Comment' section with a text input field containing 'Type-in to select a tag.' and a 'Create' button. There are two expandable sections: 'All my annotations' and 'All annotations about this file'. At the bottom of the modal is a blue button that says 'Let us know what you think'.

open-nlp.zip
Checksum: md5:677f1c21c83c7fbf5420626f5370a447

PID: 0000/open-nlp.zip

Annotate in BzNote

1

B2NOTE
Yulia

Semantic tag | Free-text keyword

Comment

Type-in to select a tag. Create

All my annotations

All annotations about this file

Let us know what you think

2

B2NOTE
Yulia

The comment annotation was created. OK

NER models for OpenNLP, for the categories, types and subtypes of HAREM, with the best

All my annotations

All annotations about this file

3

B2NOTE
Yulia

Type-in to select a tag. Create

All my annotations

- Semantic tag 2
- Free-text keyword 3
- Comment 2

All annotations about this file

Semantic tag	0	0
Free-text keyword	0	0
Comment	0	1

Let us know what you think

4

B2NOTE
Yulia

No semantic tag annotation exists about this file.

Free-text keyword ^top
No free-text keyword annotation exists about this file.

Comment ^top

Label	Actions
NER models for OpenN...	

5

B2NOTE
Yulia

Search

Free-text keyword

OpenNLP

+

Q

6

B2NOTE
Yulia

Export search results

Query: OpenNLP

New search Generate JSON export

Exact match

- Files (select all)
- <https://trmg-b2share.eudat.eu/api/files/aa3b2653-00b5-44a3-a540-2787802a4d25/open-nlp.zip>

Related match (synonyms)

No file retrieved on synonym match.

Download this as JSON-LD Download as RDF/XML

7

B2NOTE
nelsonpereira1991

Comment

This is a really useful dataset. I recommend it.

By: nelsonpereira1991

Used by me to annotate this file:

<https://trmg-b2share.eudat.eu/api/files/aa3b2653-00b5-44a3-a540-2787802a4d25/open-nlp.zip>

9

B2NOTE
nelsonpereira1991

Annotations

Semantic tag	1	0
Free-text keyword	1	1
Comment	1	1

Semantic tag ^{^top}

Label	Nb Files	Actions
Models, Genetic	1	

Free-text keyword ^{^top}

Label	N Files	Actions
CoreNLP	1	
OpenNLP	1	

Comment ^{^top}

Label	Actions
This is a really use..10	
NER models for OpenNLP	

Edit is not possible. This annotation was not created by me.

HAREM NER Models for OpenNLP, Stanford CoreNLP, spaCy, NLTK

by [André Pires](#);

Jun 22, 2017

Keywords: [named entity recognition](#); [models](#); [text mining](#); [portuguese](#);

DOI: [XXXX/b2share.a4906773dc1f42f882bd03be0c9846c3](https://doi.org/10.5281/zenodo.1611111)

PID: [0000/a4906773dc1f42f882bd03be0c9846c3](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-63862-p0000-1-7)

Files

Name	Size
HAREM NER MODELS.json	7.01KB
HAREM NER MODELS.rdf	2.04KB
HAREM NER MODELS.txt	2.10KB
HAREM NER MODELS.zip	111.61MB
nltk.zip	5.76MB
open-nlp.zip Checksum: md5: 677f1c21c83c7fbf5420626f5370a447 PID: 0000/open-nlp.zip <input type="button" value="Copy"/> <input type="button" value="Annotate in BzNote"/>	2.41MB
spacy.zip	15.72MB
stanford-corenlp.zip	87.75MB

Basic metadata

Open Access	True <input checked="" type="checkbox"/>
Publisher	INESC TEC
Language	PT

B2NOTE
nelsonpereira1991

Annotations

Semantic tag	Nb Files	Actions
Models, Genetic	1	<input type="button" value="edit"/> <input type="button" value="delete"/>

Free-text keyword	N Files	Actions
CoreNLP	1	<input type="button" value="edit"/> <input type="button" value="delete"/>
OpenNLP	1	<input type="button" value="edit"/> <input type="button" value="delete"/>

Comment

Label	Actions
This is a really use... NER models for OpenNLP	5 <input type="button" value="edit"/> <input type="button" value="delete"/>

Edit is not possible. This annotation was not created by you.

Ligações web úteis

Muito obrigada!

1. Dendro lightweight ontologies in EUDAT B2SHARE: <https://b2share.eudat.eu/records/7c429c0ec56d4db590faa7aeda7d7c22>
1. Yulia Karimova et al. “Promoting semantic annotation of research data by their creators: A use case with B2NOTE at the end of the RDM workflows”. In: *Metadata and Semantics Research: 11th International Conference, MTSR 2017, Tallinn, Estonia*. <http://www.mtsr-conf.org/>
1. Training instance of B2SHARE for testing the B2NOTE service <https://trng-b2share.eudat.eu>
1. Annotated dataset used on experience <https://trng-b2share.eudat.eu/records/a4906773dc1f42f882bd03be0c9846c3>