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]

[a]INESC TEC, Faculdade de Engenharia, Universidade do Porto, {ylaleo, joaoaguiarcastro, joaorosilva} @gmail.com

Findable   Accessible   Interoperable   Reusable

Plataforma Dendro

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

https://rdm.inesctec.pt/

https://b2share.eudat.eu/

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

Folder

Selection

My Projects

open-nlp.zip

stanford-corenlp.zip

nltk.zip

spacy.zip

2

About 📁 HAREM%20NER%20MODELS

File Changes   Project Changes

Project Stats

COPY FROM PARENT   ℹ 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   ▾

➕ ALTERNATIVE TITLE   ▾

4

My repository bookmarks   Connect to a new repository

My Projects > models

[SELECT A DESTINATION REPOSITORY ▾] [⟳]   [CHOOSE REPOSITORY TYPE ▾]   ⊘ SHARE   ⊕ SAVE

Folder   MANUAL SELECTION

[+] [⤴] [⤓] [🗄]   ...tor...

Connect to B2Share

Selection   ...re terms

[⤓] [🗄]

Label

My Projects   trngB2share   ✔

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

open-nlp.zip
                              External repository address

stanford-corenlp.zi   trng-b2share.eudat.eu   ✔

nltk.zip   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.

spacy.zip   **Access Token**

                              MmGKBzjpdlT382lag38zxhsKttZDw9e7u6zZmzucVFUu1aYM5i55WpeUSgFE   ✔

                              Your B2Share token

                              [CANCEL]   [🔖 SAVE NEW BOOKMARK]

Search records for...

Q SEARCH

HELP   COMMUNITIES   UPLOAD   CONTACT

nelsonpereira1991@gmail.com ▾

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   Copy

PID: 0000/a4906773dc1f42f882bd03be0c9846c3   Copy

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

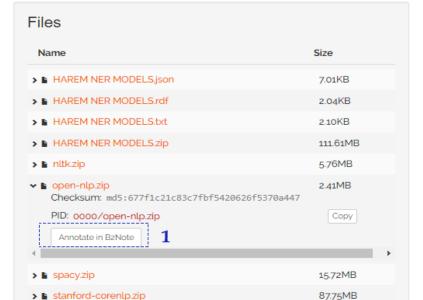# HAREM NER Models for OpenNLP, Stanford CoreNLP, spaCy,
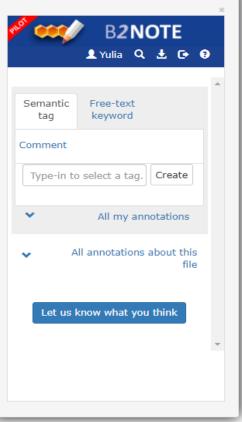
by André Pires;

Jun 22, 2017

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

DOI: XXXX/b2share.a4906773dc1f42f882bd03be0c9846c3   Copy

PID: 0000/a4906773dc1f42f882bd03be0c9846c3   Copy

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

Checksum: md5:677f1c21c83c7fbf5420626f5370a447

PID: 0000/open-nlp.zip   Copy

Annotate in B2Note   **1**

spacy.zip	15.72MB
stanford-corenlp.zip	87.75MB

Basic metadata

Open Access	True ✔
Publisher	INESC
Language	PT

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

Edit Record   Report Abuse

open-nlp.zip
Checksum: md5:677f1c21c83c7fbf5420626f5370a447
PID: 0000/open-nlp.zip

Annotate in B2Note

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

Panel 6

Search

Free-text keyw... ▾

OpenNLP

\+

🔍

6

Panel 7

Export search results

Query: OpenNLP

New search | Generate JSON expo

Exact match

☑ **Files (select all)**

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

Related match (synonyms)

No file retrieved on synonym
match.

7

Download this JSON-LD | Download as RDF/XML **8**

Panel 9

Comment

This is a really useful dataset.
I recommend it.

By: nelsonpereira1991

✏️ 🗑️

Used by me to annotate
this file:

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

9

Panel 10

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..	✏️ 🗑️
NER models for OpenN	🚫 🗑️

10

Edit is not possible,
This annotation was not crated by me

B2SHARE EUDAT

❓ Harem 🔍 SEARCH

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 `Copy`

PID: 0000/a4906773dc1f42f882bd03be0c9846c3 `Copy`

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
Checksum: md5:677f1c21c83c7fbf5420626f5370a447	
PID: 0000/open-nlp.zip `Copy`	
`Annotate in B2Note`	
⟩ 📄 spacy.zip	15.72MB
⟩ 📄 stanford-corenlp.zip	87.75MB

Basic metadata

Open Access	True ✔
Publisher	INESC TEC
Language	PT

B2NOTE

PILOT

👤 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...**5**	✏ 🗑
NER models for OpenN...	🚫 🗑

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

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