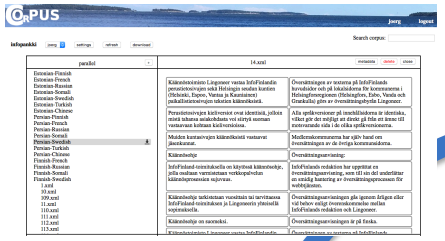
European Language Grid – META-FORUM 2020
Jörg Tiedemann (University of Helsinki, Finland)

ELG Pilot Project:
Open Translation Models, Tools and Services

OPUS-MT – Public NMT Models & Tools

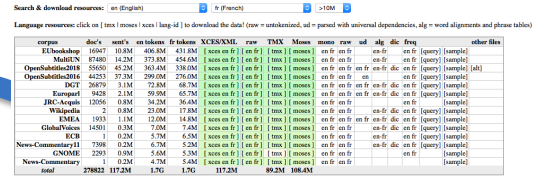
Language Technology - University of Helsinki

EUROPEAN LANGUAGE GRID



https://opus.nlpl.eu

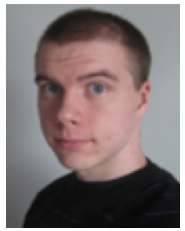
OPUS



https://opus-repository.ling.helsinki.fi

Jörg Tiedemann

Mikko Aulamo



Sam Hardwick

https://blogs.helsinki.fi/fiskmo-project

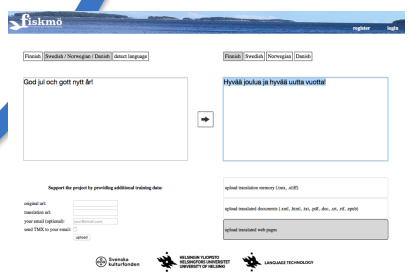
fiskmö

finsk-svensk korpus & maskinöversättning

https://github.com/Helsinki-NLP/OPUS-CAT

Tommi Nieminen

ELG-Project:
Focus on public NMT for European Minority Languages



https://translate.ling.helsinki.fi/

FOTRAN
Found in Translation

HELSINKI NLP
meaning
understanding
generating
the World's languages

WIKIMEDIA FOUNDATION

MeMAD
Methods for Managing Audiovisual Data

OPUS mt

https://github.com/Helsinki-NLP/Opus-MT

OPUS-MT - Public NMT Models & Tools

https://github.com/Helsinki-NLP/Opus-MT

Available software:

- MT server solution based on Marian NMT
- dockerised web-app, translation interface and API
- NMT training pipeline (OPUS-MT-train)
- CAT integration (OPUS-CAT)

Pre-trained translation models:

- number of bilingual models: 1,048
- number of multilingual models: 53
- number of supported source languages: 229
- number of supported target languages: 222
- number of supported language pairs: 1,715

OPUS-MT – Public NMT Models & Tools

https://github.com/Helsinki-NLP/Opus-MT

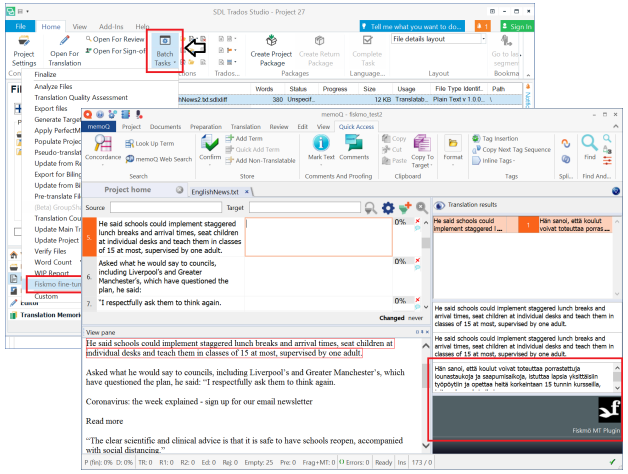
EUROPEAN LANGUAGE GRID

OPUS


```
{
    "alignment": [
        "0-0 0-2 1-1 2-3",
        "0-0 1-1 3-2 4-3 5-4"
    ],
    "result": "How are you? The translation is fun.",
    "server": "192.168.1.18:20001",
    "source": "fi",
    "source-segments": [
        "Mit\u00e4 kuuluu ?",
        "K\u00e4\u00e4@@ nn\u00f6@@ s on hauskaa ."
    ],
    "source-sentences": [
        "Mit\u00e4 kuuluu?",
        "K\u00e4\u00e4nn\u00f6s on hauskaa."
    ],
    "target": "en",
    "target-segments": [
        "How are you ?",
        "The translation is fun ."
    ],
    "target-sentences": [
        "How are you?",
        "The translation is fun."
    ]
}
```


Development demo:
https://translate.ling.helsinki.fi/ui/sami

OPUS-CAT - plugins and local MT engines
https://github.com/Helsinki-NLP/OPUS-CAT



OPUS-MT at huggingface
https://huggingface.co/Helsinki-NLP

HUGGING FACE

Back to all models

Model: `Helsinki-NLP/opus-mt-ROMANCE-en`

pytorch · rust · marian · lm-head · seq2seq · roa · en · translation

Hosted inference API

translation

Your sentence here...

Compute

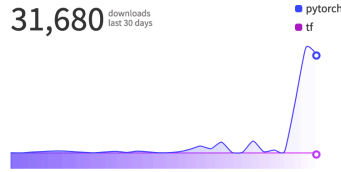
This model can be loaded on the Inference API on-demand.

Monthly model downloads

Helsinki-NLP/opus-mt-ROMANCE-en

31,680 downloads last 30 days

pytorch
tf



How to use this model directly from the `/transformers` library:

```
from transformers import AutoTokenizer,
AutoModelWithLMHead

tokenizer = AutoTokenizer.from_pretrained("Helsinki-
NLP/opus-mt-ROMANCE-en")

model = AutoModelWithLMHead.from_pretrained("Helsinki-
NLP/opus-mt-ROMANCE-en")
```

List all files in model · See raw config file