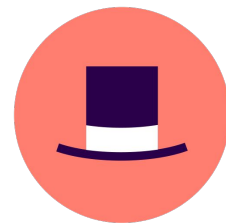


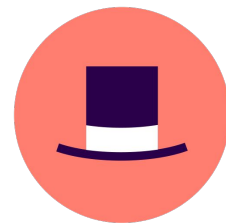
## TFIDF project



# Annif projects

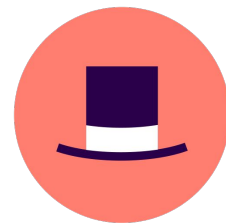


- A project is used to set a vocabulary, a backend (i.e. algorithm), and other settings.



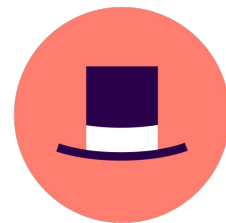
# Annif projects

- A project is used to set a vocabulary, a backend (i.e. algorithm), and other settings.
- Projects are defined in a file usually called `projects.cfg` located in the current directory where Annif is executed.
  - This default filename/location can be overridden using `ANNIF_PROJECTS` environment variable or `--projects` option after a command.



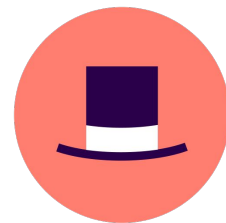
# Annif projects

- A project is used to set a vocabulary, a backend (i.e. algorithm), and other settings.
- Projects are defined in a file usually called `projects.cfg` located in the current directory where Annif is executed.
  - This default filename/location can be overridden using `ANNIF_PROJECTS` environment variable or `--projects` option after a command.
- A project is identified by a project id, which is typically a short string such as `yso-tfidf-en`.



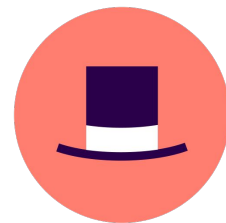
# Annif projects

- A project is used to set a vocabulary, a backend (i.e. algorithm), and other settings.
- Projects are defined in a file usually called `projects.cfg` located in the current directory where Annif is executed.
  - This default filename/location can be overridden using `ANNIF_PROJECTS` environment variable or `--projects` option after a command.
- A project is identified by a project id, which is typically a short string such as `yso-tfidf-en`.
- `annif list-projects` command shows the configured projects.



## Exercise 2: TFIDF project

- The “Hello World” algorithm of automated subject indexing: quick to set up, train and test, but not the final say!

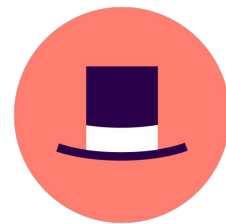


# Example projects.cfg file for TFIDF project

```
[yso-tfidf-en]  
name=YSO TFIDF project  
language=en  
backend=tfidf  
vocab=yso-en  
analyzer=snowball(english)
```

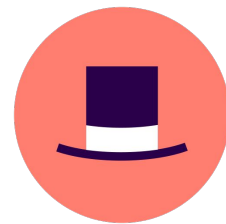
```
[stw-tfidf-en]  
name=STW TFIDF project  
language=en  
backend=tfidf  
vocab=stw-en  
analyzer=snowball(english)
```

Setting up, training, and testing TFIDF project



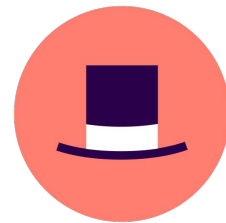


# Setting up, training, and testing TFIDF project



1. Add project configuration to `projects.cfg`; verify using **`annif list-projects`**

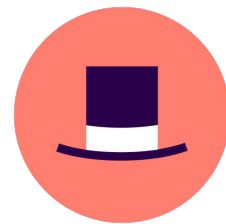
# Setting up, training, and testing TFIDF project



1. Add project configuration to `projects.cfg`; verify using **`annif list-projects`**
2. Load vocabulary: **`annif loadvoc PROJECT_ID SUBJECT_FILE`**

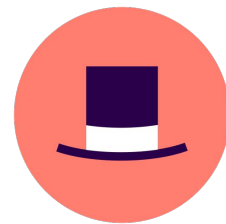
TSV or SKOS/RDF

# Setting up, training, and testing TFIDF project



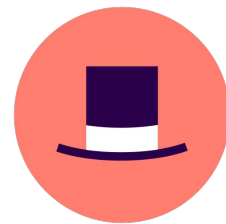
1. Add project configuration to `projects.cfg`; verify using **`annif list-projects`**
  2. Load vocabulary: **`annif loadvoc PROJECT_ID SUBJECT_FILE`**
  3. Train: **`annif train PROJECT_ID TRAINING_DATA`**
- TSV or SKOS/RDF
- (gzipped) TSV file  
or directory

# Setting up, training, and testing TFIDF project



1. Add project configuration to `projects.cfg`; verify using **`annif list-projects`**
  2. Load vocabulary: **`annif loadvoc PROJECT_ID SUBJECT_FILE`**
  3. Train: **`annif train PROJECT_ID TRAINING_DATA`**
- TSV or SKOS/RDF
- (gzipped) TSV file  
or directory

# Setting up, training, and testing TFIDF project



1. Add project configuration to `projects.cfg`; verify using **`annif list-projects`**
2. Load vocabulary: **`annif loadvoc PROJECT_ID SUBJECT_FILE`**
3. Train: **`annif train PROJECT_ID TRAINING_DATA`**
4. Test:

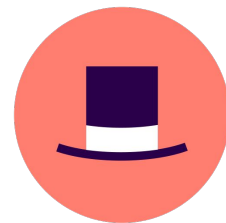
TSV or SKOS/RDF

(gzipped) TSV file  
or directory

- a. Using one sentence:

```
echo "This is an example." | annif suggest PROJECT_ID
```





# Setting up, training, and testing TFIDF project

1. Add project configuration to `projects.cfg`; verify using **annif list-projects**
2. Load vocabulary: **annif loadvoc PROJECT\_ID SUBJECT\_FILE**
3. Train: **annif train PROJECT\_ID TRAINING\_DATA**

TSV or SKOS/RDF

4. Test:

- a. Using one sentence:

```
echo "This is an example." | annif suggest PROJECT_ID
```

(gzipped) TSV file  
or directory

- b. Using a text file:

```
annif suggest PROJECT_ID <FILE.TXT
```