

Coreference Visualization

Christoph Wessarges, Moritz Wegner, Swetlana Shaban

Universität Hamburg

29 Sept 2022

Outline

- Goal
- What is a coreference?
- The model
- General project description
- Backend
- Frontend
- Demo
- Possible features

Goal

Goal: building a web application for multi-user coreference annotation in German.

Coreference

- Coreference resolution is the task of resolving text spans in documents that refer to the same entities.¹
- Used for question answering, text summarization, literary analysis.
- Example:

[Alice]₁ was not a bit hurt, and [she]₁ jumped up on to
[her]₁ feet in a moment: [she]₁ looked up, but it was
all dark overhead; before [her]₁ was [another long
passage]₂, and [the White Rabbit]₃ was still in sight,
hurrying down [it]₂.

Figure 1: Coreference gold annotations for “Alice’s Adventures in Wonderland” (annotations from Bamman et al., 2020)

¹Schröder, Hatzel, and Biemann, “Neural End-to-end Coreference Resolution for German in Different Domains”.

The model

- Incremental Coreference Model variant
- Based on pretrained BERT neural network (RNN)
- Trained on DROC (Deutsches ROman Corpus)

General project description

Users (annotators) are able to:

- Access & edit anywhere

Details

Access & edit annotated documents from anywhere through login.

General project description

Users (annotators) are able to:

- Access & edit anywhere
- Create projects

Details

Create projects for easier management of annotated documents.

General project description

Users (annotators) are able to:

- Access & edit anywhere
- Create projects
- Load in texts

Details

Load text documents or user-input text (automatic annotation).

General project description

Users (annotators) are able to:

- Access & edit anywhere
- Create projects
- Load in texts
- Coreference information and statistics

Details

Easy access to detailed information about each coreference.

General project description

Users (annotators) are able to:

- Access & edit anywhere
- Create projects
- Load in texts
- Coreference information and statistics
- Correct errors

Details

Correct errors made by the model.

General project description

Users (annotators) are able to:

- Access & edit anywhere
- Create projects
- Load in texts
- Coreference information and statistics
- Correct errors
- Edit coreferences

Details

Simple interface to add and edit coreferences persistently.

General project description

Users (annotators) are able to:

- Access & edit anywhere
- Create projects
- Load in texts
- Coreference information and statistics
- Correct errors
- Edit coreferences
- Download documents

Details

Download annotated documents in common file formats.

- NoSQL MongoDB database

Details

JSON can natively handle the nested list structure of annotations. Therefore a MongoDB database was employed. It allows efficient and intuitive transformation from Javascript objects to JSON.

Backend

- NoSQL MongoDB database
- Python Flask backend with PyMongo

Details

We use basic Flask as our Web framework and the library PyMongo as the interface to the MongoDB instance.

Backend

- NoSQL MongoDB database
- Python Flask backend with PyMongo
- Coreference Resolution model integrated into backend

Details

The neural network model is fully integrated into the backend. Booting up the server automatically initializes the trained model.

Backend

- NoSQL MongoDB database
- Python Flask backend with PyMongo
- Coreference Resolution model integrated into backend
- Server exposes RESTful API

Details

The frontend only communicates over the REST API with the backend, including the model inference. Only JSON is allowed in request and response bodies.

Backend

- NoSQL MongoDB database
- Python Flask backend with PyMongo
- Coreference Resolution model integrated into backend
- Server exposes RESTful API
- Login functionality (User management)

Details

Login functionality fully implemented with session cookies (Flask Login Manager).

Backend

- NoSQL MongoDB database
- Python Flask backend with PyMongo
- Coreference Resolution model integrated into backend
- Server exposes RESTful API
- Login functionality (User management)
- Jinja HTML templates for Login.

Details

For the registration and login screens, the backend exposes endpoints using HTML templates.

Backend

- NoSQL MongoDB database
- Python Flask backend with PyMongo
- Coreference Resolution model integrated into backend
- Server exposes RESTful API
- Login functionality (User management)
- Jinja HTML templates for Login.
- Statistics saved in database and presented with plotly

Details

The probability distributions corresponding to the model's decision confidences (softmax output) are saved with each document and plotted with plotly JS.

Application Entities

- CRUD operations defined for all entities (reflected in REST API).
- Model inference applied automatically when adding a new document.
- MongoDB is schema-less, but we defined a schema to limit undesired usage/effects.

Application Entities

- CRUD operations defined for all entities (reflected in REST API).
- Model inference applied automatically when adding a new document.
- MongoDB is schema-less, but we defined a schema to limit undesired usage/effects.

Database models

- ① User: E-mail (unique, for login), password, name, roles
- ② Project: Name (unique per user), documentIds, createdBy, createdAt
- ③ Document: Name (unique per project), tokens, clusters, probs, annotatedBy, createdBy, createdAt

Database models (pydantic)

- User model:

```
schema_extra = {
    "example": {
        "_id": "066de609-b04a-4b30-b46c-32537c7f1f6e",
        "name": "Max Mustermann",
        "email": "max.mustermann@email.de",
        "password": "...",
        "role": 1,
        "active": True
    }
}
```

Database models (pydantic)

- Project model:

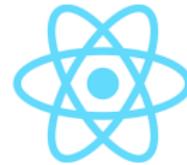
```
schema_extra = {
    "example": {
        "_id": "123e4567-e89b-12d3-a456-426655440000",
        "name": "My Project",
        "createdBy": "066de609-b04a-4b30-b46c-32537c7f1f6e",
        "sharedWith": [],
        "docIds": ["123e4567-e89b-12d3-a456-42665544321"],
        "createdAt": datetime.now()
    }
}
```

Database models (pydantic)

- Document model:

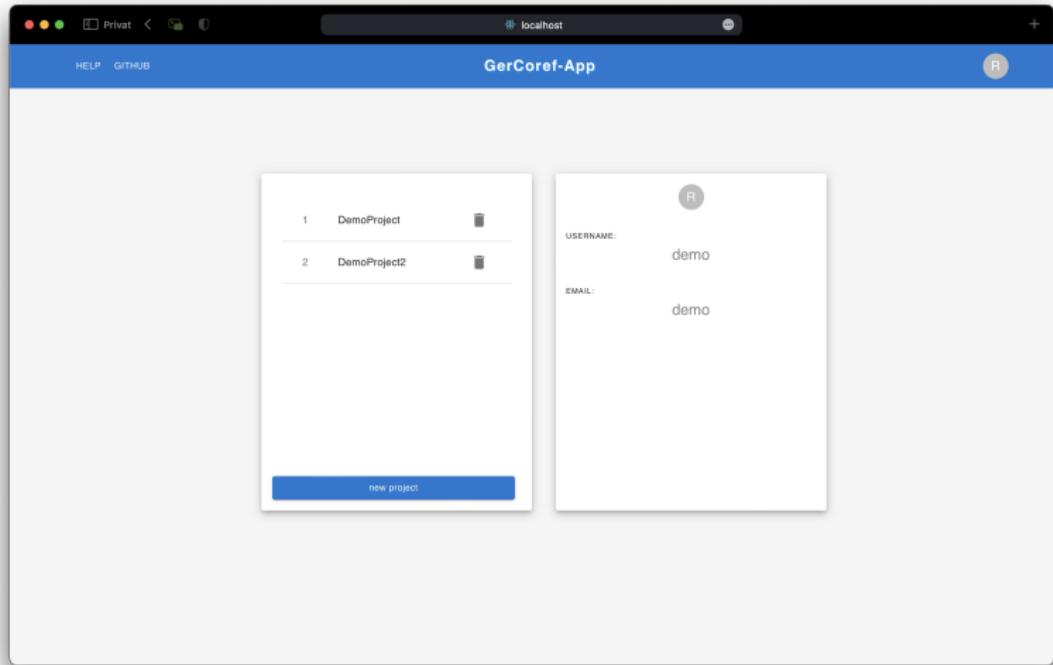
```
schema_extra = {
    "example": {
        "_id": "123e4567-e89b-12d3-a456-426655440000",
        "name": "The Hitchhiker's Guide to the Galaxy",
        "createdBy": "066de609-b04a-4b30-b46c-32537c7f1f6e",
        "sharedWith": [],
        "tokens": ["Time", "is", "an", "illusion", ".", "Lunchtime", "doubly", "so", "."],
        "clust": [[[0, 0], [3, 3]]],
        "annotatedBy": [["066de609-b04a-4b30-b46c-32537c7f1f6e", "066de609-b04a-4b30-b46c-32537c7f1f6e"]],
        "probs": [[[1, 0], [.1, 0, .9]]],
        "createdAt": datetime.now(),
        "projectFk": "066de609-b04a-4b30-b46c-32537c7f1eee"
    }
}
```

- Made with React
- two pages are **Dashboard** and the **Mainpage**



Frontend

- Dashboard: Hub-Area for all projects



Frontend

- Our initial mock-up design

German CoRef Annotator

 Help  Login

Text Selection:

Lorem ipsum

Coreferences:

- Vulpitate
- neque**
- commodo

Add new Coreference

Delete selected Coreference

Auto-Annotate

1 Lorem ipsum dolor sit amet, consectetur adipiscing elit.

2 Nulla quam velit, **vulpitate** eu pharetra nec, mattis ac **neque**.

3 Duis vulputate **commodo** lectus, ac blandit elit tincidunt id.

4 Sed rhoncus, tortor sed eleifend tristique, tortor mauris molestie elit, et lacinia ipsum quam nec dui.

⋮

View of the Document

Annotated by:
GerCoRef

Settings **Statistics**

- Documents
- Document 1
- Document 2 (shared)
- Document 3

Load Documents

Share Documents

Download Annotated Document

Submit Annotation
(Submit for online learning)

A project of the University of Hamburg (UHH)
[About Page](#)



Frontend

- Mainpage: consists of 3 parts
- Corefmanager-panel, Mainview-panel and Input-Search-Statistics-panel

The screenshot shows the GerCoref-App interface on a Mac OS X desktop. The window title is "GerCoref-App" and the tab is "DemoProject".

Corefmanager-panel (Left):

- Current Text Selection: "Schwester Kathleen"
- Coreferences:
 - Schwester Kathleen
- Buttons: "Add new Coreference", "Delete Coreference", "Save Changes (Ctrl + S)"
- Switches: "Model Inference" (checked), "Hover Information" (checked)

Mainview-panel (Center):

- Text file: "demotext.txt"
- Text content:
 - Nachdem [er], 1935 die Schule abgeschlossen hatte, reiste [er] mit [seiner], [Schwester Kathleen] und [seinen], [Eltern], nach England, um [sich], dort an der London School of Economics für ein Studium der Volkswirtschaftslehre einzuschreiben.
 - Eine Krankheit zwang [ihn], jedoch, das Vorhaben aufzugeben und [sich], in ärztliche Behandlung zu begeben.
 - Zurück in den USA, meldete [er], [sich], an der Princeton University an,
 - musste diese aber wieder verlassen, nachdem [er], an Malaria erkrankt war.
 - Ein Jahr später, 1936, schaffte [Kennedy], schließlich den Einstieg ins Studium.
 - Von 1936 bis 1940 studierte [er], Politik an der Harvard University.
 - [Seinem], [Freund Lem Billings], [er], über [seine], Anziehungskraft auf [Frauen].
 - [Ich], werde jetzt hier [Playboy], genannt.
 - Während [seiner], Studienzeit reiste [er], viel, auch durch das kurz vor dem Krieg stehende Europa.
- Pagination: < 1 2 3 >

Input-Search-Statistics-panel (Right):

- Text input field: "Tyr"
- Buttons: "File", "Search", "Help", "Graph Dashboard"
- Section: "Model's Coreference Confidence" with a bar chart.
 - Cluster 1: New Cluster (blue bar)
 - No Mention: No barA legend indicates that blue bars represent "New Cluster" and no bars represent "No Mention".
- Buttons: "Open Graph Dashboard", "Close"

Coreference Manager Panel

Current Text Selection:

Die Organisation

Coreferences:

Die Organisation

sie

sie

sie

- 1 **[Die Organisation]**, gab bekannt**[sie]**, habe Spenden veruntreut.
- 2 Außerdem haben **[sie]**₁ betont dass **[sie]**₁ mit schwerwiegenden Konsequenzen rechnen müssen.
- 3 Erfahren Sie mehr von unserem Interview mit dem Vorstandschef Klaus Ehrlich.

Coreference Manager Panel

Current Text Selection:

Sie

Coreferences:

N/A

Add new Coreference

New Cluster

Cluster 1

- 1 **[Die Organisation]**, gab bekannt**[sie]₁**, habe Spenden veruntreut.
- 2 Außerdem haben **[sie]₁** betont dass **[sie]₁** mit schwerwiegenden Konsequenzen rechnen müssen.
- 3 Erfahren **Sie** mehr von unserem Interview mit dem Vorstandschef Klaus Ehrlich.

Coreference Manager Panel

Current Text Selection:

Sie

Coreferences:

Sie

- 1 **[Die Organisation]**₁ gab bekannt**[sie]**₁, habe Spenden veruntreut.
- 2 Außerdem haben**[sie]**₁ betont dass**[sie]**₁ mit schwerwiegenden Konsequenzen rechnen müssen.
- 3 Erfahren**[Sie]**₂ mehr von unserem Interview mit dem Vorstandschef Klaus Ehrlich.

Coreference Manager Panel

Current Text Selection:

Vorstandschef Klaus Ehrlich

Coreferences:

N/A

Add new Coreference

New Cluster

Cluster 1

Cluster 2

- 1 **[Die Organisation]**₁ gab bekannt**[sie]**₁ habe Spenden veruntreut.
- 2 Außerdem haben**[sie]**₁ betont dass**[sie]**₁ mit schwerwiegenden Konsequenzen rechnen müssen.
- 3 Erfahren**[Sie]**₂ mehr von unserem Interview mit dem **Vorstandschef Klaus Ehrlich**.

Text manual input

HELP GITHUB

GerCoref-App

P

Current Text Selection:

No Selection

Coreferences:

N/A

Add new Coreference

Delete Coreference

Save Changes (Ctrl + S)

Model Inference

Hover Information

text.txt

1 [Es], war einmal ein kleines süßes [Mädchen], [das], hatte jedermann lieb, [der], [sie], nur ansah, am allerliebsten aber [ihre], [Großmutter], [die], wusste gar nicht, was [sie] alles dem [Kinde], geben sollte.

2 Einmal schenkte [sie], [ihm], ein Käppchen von rotem Samt, und weil [ihm], das so wohl stand, es nichts anders mehr tragen wollte, hieß [es], nur das [Rotkäppchen].

3 Eines Tages sprach [seine], [Mutter], zu [ihm], "Komm, [Rotkäppchen], da hast [du], ein Stück Kuchen und eine Flasche Wein, bring das der [Großmutter], hinaus;

4 [sie], ist krank und schwach und wird [sich], daran laben.

5 Mach [dich], auf, bevor es heiß wird, und wenn [du], hinauskommst, so geh hübsch sittsam und lauf nicht vom Wege ab, sonst fällt [du], und zerbricht das Glas, und die [Großmutter], hat nichts.

6 Und wenn [du], in [ihre], Stube kommst, so vergiss nicht guten Morgen zu sagen und guck nicht erst in allen Ecken herum!"

text

Es war einmal ein kleines süßes Mädchen, das hatte jedermann lieb, der sie nur ansah, am allerliebsten aber ihre Großmutter, die wusste gar nicht, was sie alles dem Kind geben sollte. Einmal schenkte sie ihm ein Käppchen von rotem Samt, und weil ihm das so wohl stand, es nichts anders mehr tragen wollte, hieß er das Rotkäppchen. Eines Tages sprach seine Mutter zu ihm: "Komm, Rotkäppchen, da hast du ein Stück Kuchen und eine Flasche Wein, bring das der Großmutter hinaus; sie ist krank und schwach und wird sich daran laben. Mach dich auf, bevor es heiß wird, und wenn du hinauskommst, so geh hübsch sittsam und lauf nicht vom Wege ab, sonst fällt du und zerbricht das Glas, und die Großmutter hat nichts. Und wenn du in ihre Stube kommst, so vergiss nicht guten Morgen zu sagen und guck nicht erst in allen Ecken herum!"

1

SUBMIT

Upload documents

HELP GITHUB R

Current Text Selection:
No Selection

Coreferences:
N/A

Add new Coreference
Delete Coreference
Save Changes (Ctrl + S)

Model Inference
Hover Information

GerCoref-App

p

gross2.txt

1 [Es], war einmal ein kleines süßes [Mädchen], [das], hatte jedermann lieb,
[der], [sie], nur ansah, am allerliebsten aber [ihre], [Großmutter].
[die], wusste gar nicht, was [sie], alles dem [Kinde], geben sollte.

2 Einmal schenkte [sie], [ihm], ein Käppchen von rotem Samt, und weil
[ihm], das so wohl stand, und es nichts anders mehr tragen wollte, hieß [es],
nur das [Rotkäppchen].

3 Eines Tages sprach [seine], [Mutter], zu [ihm]: " Komm,
[Rotkäppchen], da hast [du], ein Stück Kuchen und eine Flasche Wein, bring
das der [Großmutter], hinaus;

4 [sie], ist krank und schwach und wird [sich], daran laben.

5 Mach [dich], auf, bevor es heiß wird, und wenn [du], hinauskommst, so geh
hübsch sittsam und lauf nicht vom Wege ab, sonst fällst [du], und zerbricht das
Glas, und die [Großmutter], hat nichts.

6 Und wenn [du], in [Ihre], Stube kommst, so vergiss nicht guten Morgen zu
sagen und guck nicht erst in allen Ecken herum!

7 " " [ich], will schon alles richtig machen, " sagte [Rotkäppchen], zur

Choose file No file chosen Upload

Files

gross2.txt (1)
gross2.txt
G.txt
text.txt

New Document Name Rename

Download annotated document

XML
CoNLL-2012
plaintext

Search

HELP GITHUB

R

GerCoref-App

p

Current Text Selection:

No Selection

Coreferences:

N/A

Add new Coreference

Delete Coreference

Save Changes (Ctrl + S)

Model Inference

Hover Information

33 [Es] wunderte [sich], dass die Tür aufstand, und wie [es], in die Stube trat, so kam es [ihm], so seltsam darin vor, dass [es], dachte: [Ei], [du], mein Gott, wie ängstlich wird mir's heute zumut, und bin sonst so gerne bei der [Großmutter].

34 [Es] rief: "Guten Morgen," bekam aber keine Antwort.

35 Darauf ging [es] zum Bett und zog die Vorhänge zurück.

36 Da lag die [Großmutter], und hatte die Haube tief ins Gesicht gesetzt und sah so wunderlich aus.

37 " Ei, [Großmutter], was hast [du] für große Ohren! " - " Dass [ich],

[dich], besser hören kann! " - " Ei, [Großmutter], was hast [du] für große Augen! "

38 - " Dass [ich], [dich], besser sehen kann! " - " Ei, [Großmutter], was hast [du] für große Hände! " - " Dass [ich], [dich], besser packen kann! " - " Aber, [Großmutter], was hast [du] für ein entsetzlich großes Maul! " - " Dass [ich], [dich], besser fressen kann! "

39 Kaum hatte der [Wolf], das gesagt, so tat [er] einen Satz aus dem Bette und verschlang das arme [Rotkäppchen].

Was der [Wolf] [gesagt] Apapapit apapit hatte, hörte er [Lach].

< 1 2 3 4 5 6 7 >

Tr	W	C
Search	du	X
Sent.	Words	Coref
3	du	1
5	du	1
5	du	1
6	du	1
14	du	1
17	du	1
20	Du	1
20	du	1
22	du	1
23	du	1
24	Du	1
24	du	1
25	durch	
33	du	1
37	du	3
37	du	3
38	du	3
38	du	3

Statistics

GerCoref-App

Current Text Selection:

Coreferences:

wir

beide

Delete Coreference

Save Changes (Ctrl + S)

Model Inference

Hover Information

p

12 " Schönen

13 Dank, [Wolf] „ "

14 - " Wo hinaus so früh, [Rotkäppchen], ? " - " Zur [Großmutter], . - " Was trägst [du], unter der Schürze? "

15 - " Kuchen und Wein.

16 Gestern haben [wir], gebacken, da soll [sich], die kranke und schwache [Großmutter], etwas zugut tun und [sich], damit stärken. "

17 - [Rotkäppchen], wo wohnt [deine], [Großmutter], ? " - " Noch eine gute Viertelstunde weiter im Wald, unter den drei großen Eichbäumen, da steht [Ihr], .

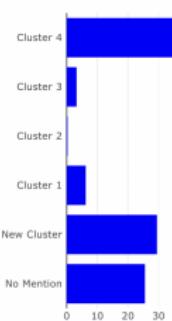
Haus, unten sind die Nusshecken, das wirst [du], ja wissen, " sagte [Rotkäppchen], .

18 Der [Wolf], dachte bei [sich], :

19 Das junge, zarte [Ding], . das ist ein fetter Bissen, der wird noch besser schmecken als die [Alte], .

20 [Du], musst es lustig anfangen, damit [du], [beide], schnappt.

Tracer



Christoph, Moritz, Swetlana (UHH)

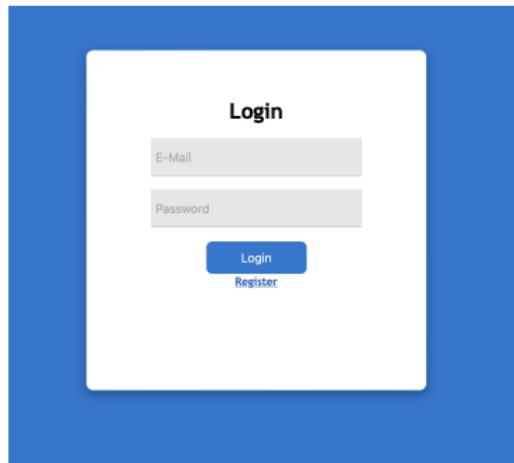
Coreference Visualization

29 Sept 2022

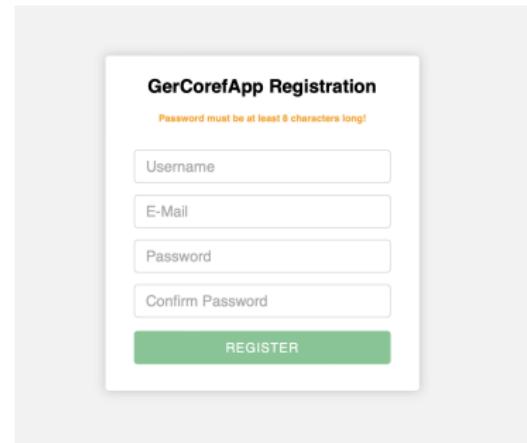
23 / 29

Extras

- Login and Registration page



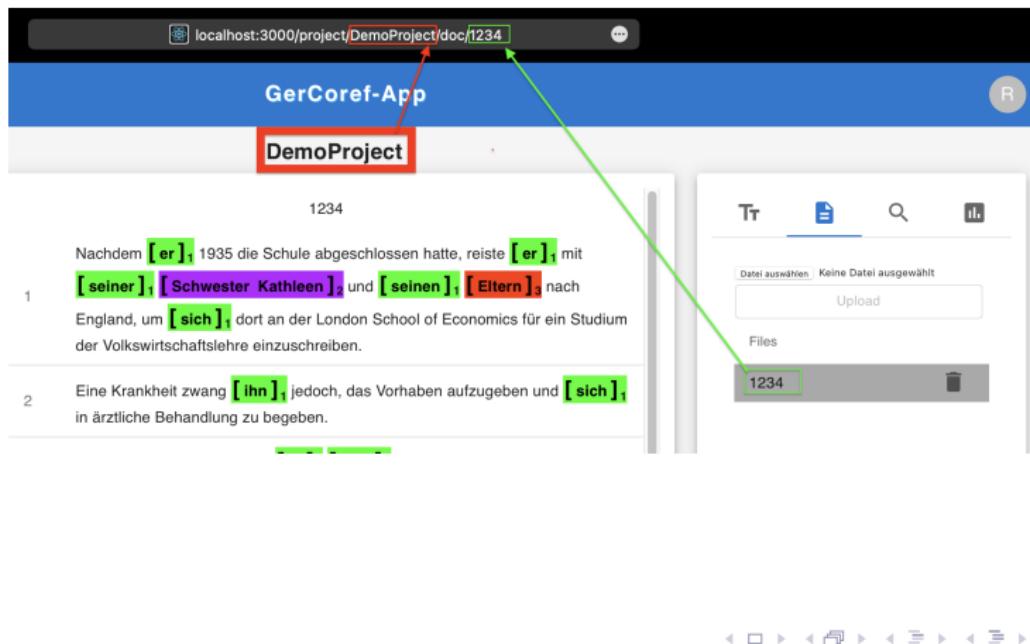
The login page features a white rectangular form on a blue background. At the top center is the word "Login". Below it are two input fields: "E-Mail" and "Password", both with placeholder text. A blue button labeled "Login" is positioned below the password field, and a smaller link "Register" is located directly beneath it.



The registration page has a light gray background and a white rectangular form in the center. At the top is the title "GerCorefApp Registration". Below it is a yellow error message: "Password must be at least 8 characters long!". There are four input fields: "Username", "E-Mail", "Password", and "Confirm Password". A large green button labeled "REGISTER" is at the bottom.

Urls

- paths are /project/projectName(/doc/docname)



Demo

- GitHub: <https://github.com/ChrisWess/GerCorefApp>

Possible features

- REST documentation (work in progress)
- Overlapping coreferences (work in progress)
- Sending information back to the model for online learning
- Simpler: collect a huge corpus of training data to improve the model
- Sharing projects between users
- More statistics: e.g. inter-annotator reliability, cluster graph(s)

Questions?

Thank you for your attention!