

# Group Project

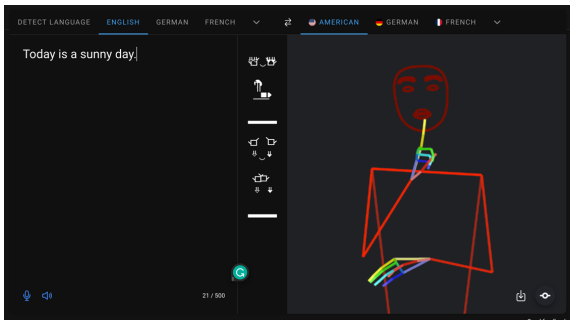
- 3 team members
- Create a web application related to language technology
- You can discuss ideas with us or tutors
- Send your project descriptions by  
Wednesday, **October 18, 2023**

# Structure of the Project

- Key components: database, frontend, backend
- Higher complexity on at least one component
  - Database: data processing
  - Frontend: visualization, innovative interface
  - Backend: business logic

# Idea 1: A Sign Language Browser Extension

- Translate selected text in a browser to sign language using API (<https://sign.mt>)
- <https://github.com/sign/translate>

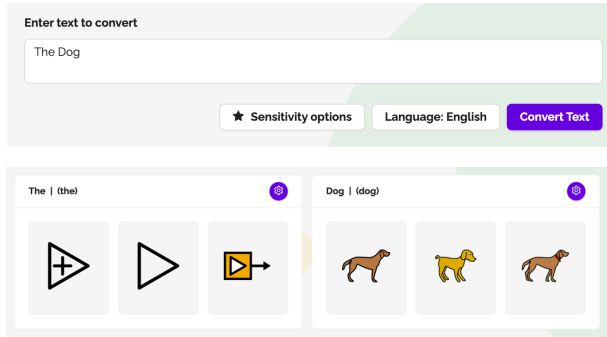


## Idea 2: Words Voting Application (Quiz)

- Create an engaging quiz-style application
- Present users with random words along with suggested part-of-speech tags
- Track user responses and display statistics
- Implement user authentication

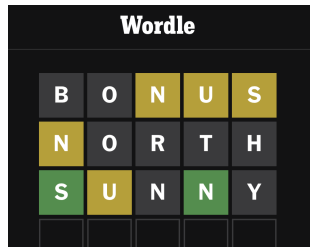
# Idea 3: Text to Pictograms

- Develop an application that translates text into pictograms
- When users input text into the frontend, the application will display relevant pictograms for each word



## Idea 4: Linguistic Game

- Create an interactive linguistic game
- Offer challenges and puzzles
- Track user progress and achievements
- Examples:
  - Language Identification Quiz
    - Players listen to a person speaking and guess the language ([potential dataset](#))
  - Der/die/das Correction
    - A web interface that inserts, removes, or corrects articles in given German or English text



- You can use your own dataset
- Swissdox@LiRI - A database containing over 24 million media articles
  - [swissdox.linguistik.uzh.ch](http://swissdox.linguistik.uzh.ch)
- LEGaM – Historical corpora of the Gallo-Romance territory
- Bullinger Digital – Collection of letters from/to Heinrich Bullinger
  - [www.bullinger-digital.ch](http://www.bullinger-digital.ch)

## Idea 5: Find the main topics of articles

- Analyze data from `Swissdiox@LiRI` to identify the main topics over the last 5 years
- Find the most important words in each article
- Present the analyzed data to users and allow them to compare topics and insights in a selected time spans

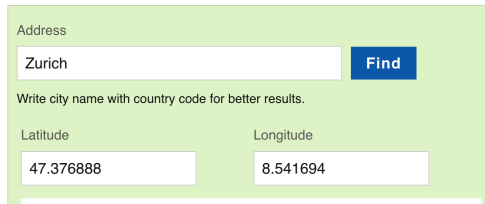


## Idea 6: The most frequent words

- Analyze data from `Swissdox@LiRI` to identify the most frequent word per media source over the last 5 years
- Group data per month and week to uncover temporal patterns and trends
- Present the distribution of words from the analyzed data to users, allowing them to explore

## Idea 7: Build a map with a LEGaM data

- Create an interactive map displaying the manuscript locations from LEGaM data
- Implement filtering options for users to customize map displays
- Utilize an API, such as [geocode.maps.co](https://geocode.maps.co/), to retrieve geographical coordinates for locations



A screenshot of a web form for geocoding. The form has a light green background. At the top, there is a label "Address" above a text input field containing "Zurich". To the right of the input field is a blue button with the text "Find". Below the input field, there is a line of text: "Write city name with country code for better results." At the bottom, there are two labels: "Latitude" and "Longitude". Below "Latitude" is a text input field containing "47.376888". Below "Longitude" is a text input field containing "8.541694".

## Idea 8: Display word information

- Bullinger Digital - Collection of 10,000+ letters from/to Heinrich Bullinger
- Display word information for Latin words in the Bullinger letters
- Display word information for Early New High German words in the Bullinger letters
- More information you can get from Martin Volk



Questions?