

**30th Telecommunications Forum Telfor 2022**

Nov 15-16, 2022, Belgrade, Serbia



# **Towards the Speech-driven Database Design**

Drazen Brdjanin, **Goran Banjac**, Nebojsa Babic, Nenad Golubovic

**M-lab Research Group @ Faculty of Electrical Engineering**  
**University of Banja Luka, Bosnia & Herzegovina**

# Presentation outline

---

- Research context and motivation
- Research objectives and contributions
- Approach outline
- Implemented tool
- Illustrative example
- Conclusion and future work

# Research context & motivation



## Model-driven Software Engineering Laboratory

Faculty of Electrical Engineering • University of Banja Luka

<http://m-lab.etf.unibl.org>

M-lab research focus:

**Automatic database design**

**based on sources of different nature**

**(models, text, speech, ...)**

### Main M-lab achievements:

#### AMADEOS

<http://m-lab.etf.unibl.org:8080/amadeos>

- The first online web-based tool for automatic CDM derivation from collections of differently represented and differently serialized BPMs

#### TexToData

<http://m-lab.etf.unibl.org:8080/TexToData>

- The first online multilingual web-based tool for automatic CDM derivation from NL text

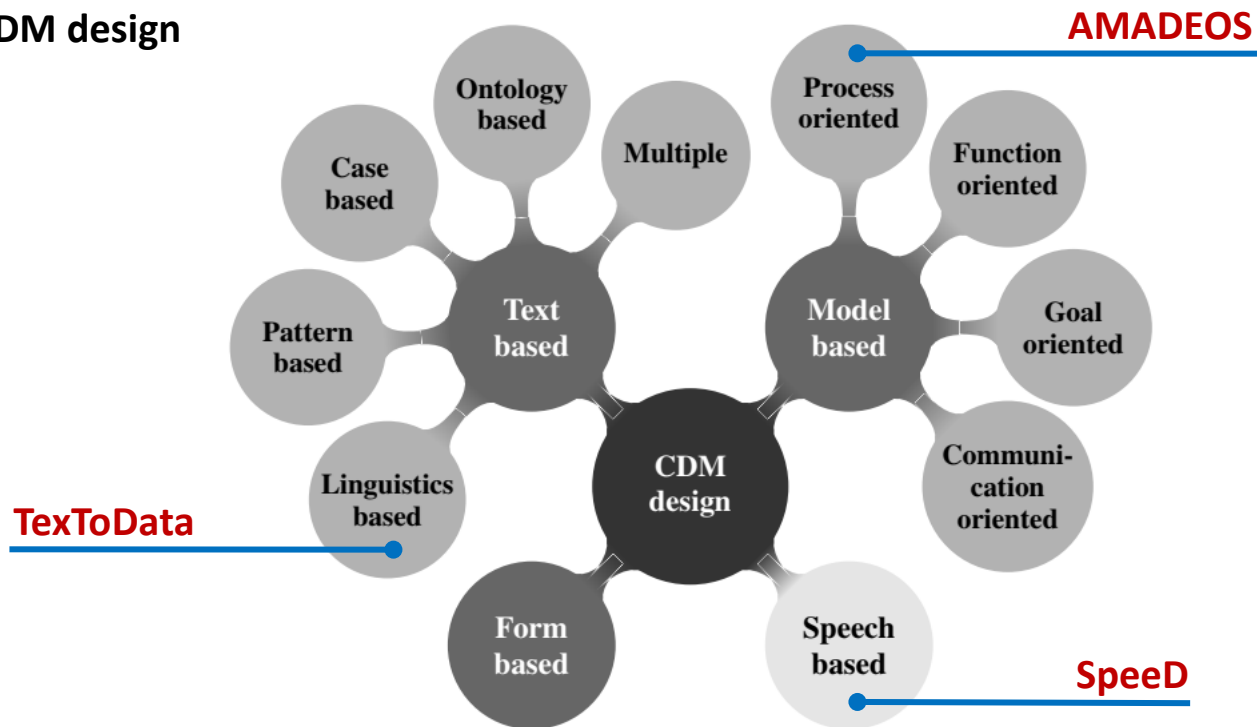
#### Speed

<http://m-lab.etf.unibl.org:8080/Speed>

- The first tool that provides functionality of CDM derivation from recorded speech

# Research context & motivation

Taxonomy of existing approaches to (semi-)automatic CDM design



# Research objectives & contributions

## Research objectives

- **Define an approach and implement an online tool / service able to automatically derive CDM from recorded speech**



(in order to be able to combine CDMs derived from different sources)

## Research contributions

- **Approach**
  - **Two-step CDM synthesis from recorded speech:**
    - Step #1: speech to text conversion**
    - Step #2: text analysis and CDM generation**
- **Implemented tool – Speed**
  - online web-based tool (set of online services)
  - supports English speech
  - automatic layouting and UML-based representation of generated CDM (editing and formatting functionalities, XMI-export to support model portability, ...)

<http://m-lab.etf.unibl.org:8080/Speed>

# Approach outline

## What do we want?

- Define an approach and implement an **online web-based tool/service** able to **automatically derive CDM from recorded speech**



## What do we need?

- Speech to text convertor
- NLP service(s) to analyze textual specifications
- CDM generator service
- Diagram layouting & editing services

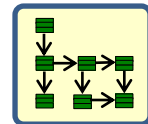


## What do we have?

- TexToData (NLP-based online tool that enables automatic derivation of conceptual database models from natural language text)

## What is missing?

- Speech to text convertor



CDM

# SpeedD – System architecture

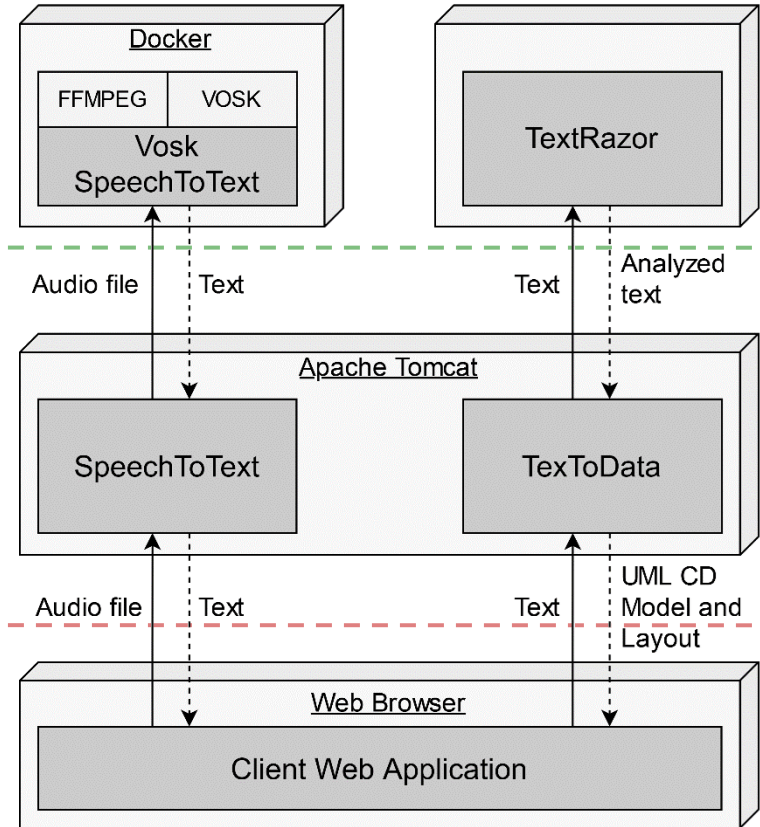
External publicly available services  
and other internal services

## Server side

- Service-oriented Architecture
- Consists of two components:
  - **SpeechToText**
  - **TextToData**

## Client side

- GUI, file upload, text modification, model manipulations, ...

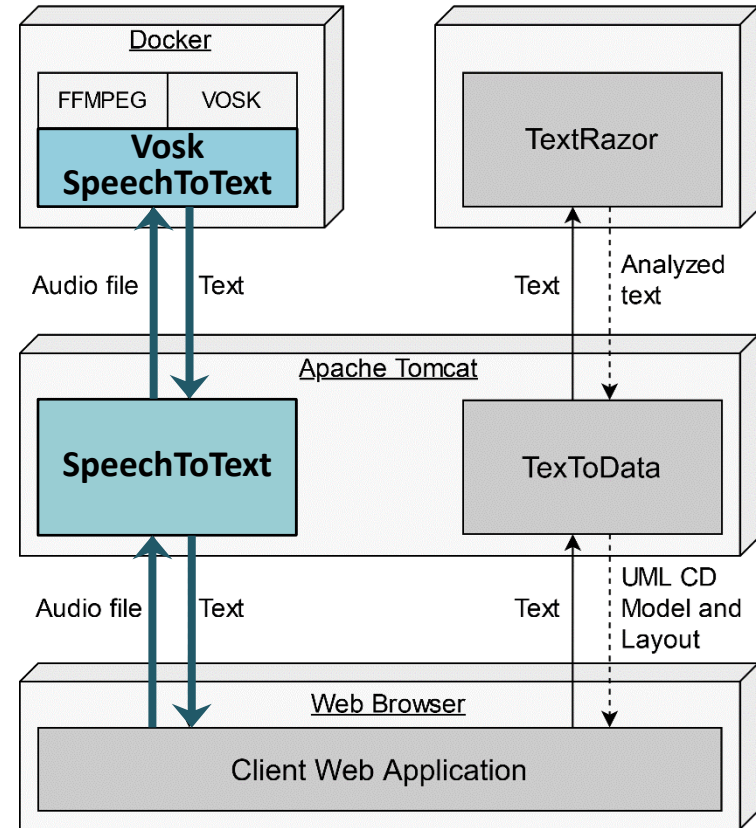


# Speed – CDM synthesis process

The process of the CDM synthesis consists of two steps:

## 1 Recorded speech to text conversion

- The **recorded speech** is sent to the **SpeechToText** service, plays the role of an adapter that employs another service
- Currently we employ the **Vosk SpeechToText** service, but we also plan to develop other speech-to-text services
- The **Vosk SpeechToText** service is based on Vosk (Python implementation)
- Vosk requires monochannel *wav* audio files with 16k sampling rate
- When the speech-to-text conversion is finished, the client web application receives the response from the **SpeechToText** service (**text**) and populates the corresponding text input field



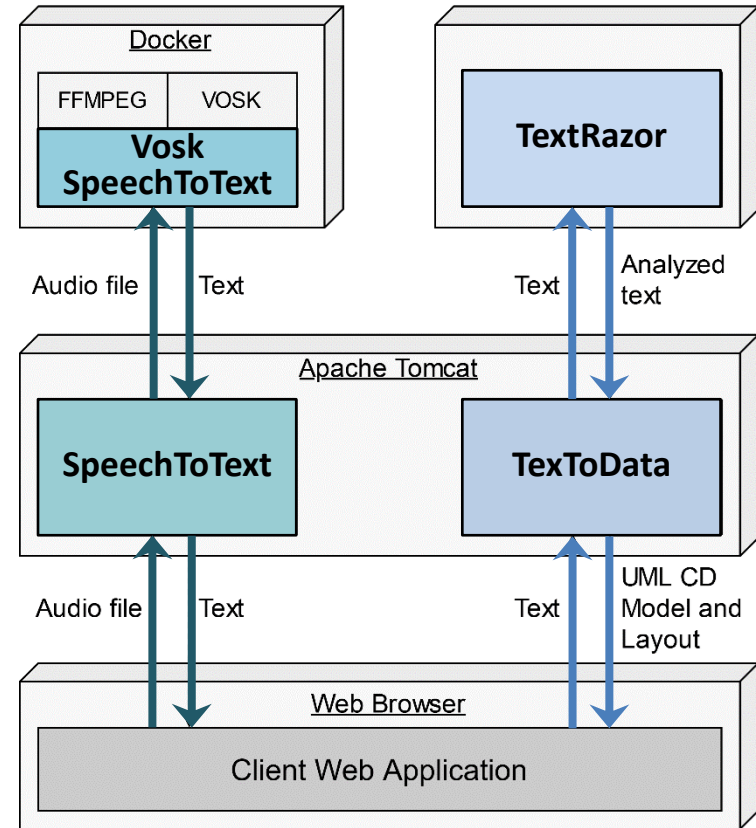


# Speed – CDM synthesis process

The process of the CDM synthesis consists of two steps:

## ② Text analysis and CDM generation

- The **text** is sent to the **TextToData** service which orchestrates the whole process
- Firstly, the NLP analysis of the source text (English) is performed by the external **TextRazor** service
- After NLP is finished, the analyzed text is sent to the internal **ModelingContext** service which generates an internal representation of the CDM
- The CDM is further sent to the **UMLmodelSerializer** service which serializes the generated model in the XMI format
- After the serialization, the model is sent to the AMADEOS **layouting** service which returns a layout of the class diagram
- Finally, the **model and the diagram** are merged into a single JSON object, and returned to the client



# Speed – Client side

## Client web application

- The client web application allows users to upload an *audio file* (*m4a, mp3, ogg, wav, wma, etc.*) with the recorded speech
- When the speech-to-text conversion is finished, the *client web application* receives the response (text) and populates the corresponding text input field
- If necessary, users are able to additionally modify the generated text

<http://m-lab.etf.unibl.org:8080/Speed>

Speed (SpeechToText+NLP-based System for Automated Database Design)

library member has id and name. library members borrow library units. library unit has name year and author. publisher has name and residence. publisher publishes library units.

✓ Load text from audio files    ✓ Analyze text

If the process takes too long, you can get your result here:  
<http://m-lab.etf.unibl.org:8080/Speed/?requestId=1668032077028>

✓ Add Class    ✓ Add Generalization    ✓ Add Association    ✓ Add Aggregation    ✓ Add Composition    ✓ Delete Element    ✓ Clear All    ✓ Export (XMI)

Class diagram

# Speed – Client side

## Client web application

- Then, user can send the (modified or unmodified) text for analysis
- When the entire process is finished, the client application receives the JSON response and visualizes the class diagram in the browser
- The visualized diagram is editable so users can additionally improve it
- It is also possible to export the model in the XMI format, and further use it in some other platform

<http://m-lab.etf.unibl.org:8080/Speed>

Speed (SpeechToText+NLP-based System for Automated Database Design)

library member has id and name. library members borrow library units. library unit has name year and author. publisher has name and residence. publisher publishes library units.

✓ Load text from audio files    ✓ Analyze text

Input text (en)	English text
library member has id and name. library members borrow library units. library unit has name year and author. publisher has name and residence. publisher publishes library units.	library member has id and name. library members borrow library units. library unit has name year and author. publisher has name and residence. publisher publishes library units.

Powered by Yandex.Translate <http://translate.yandex.com>

✓ Add Class    ✓ Add Generalization    ✓ Add Association    ✓ Add Aggregation    ✓ Add Composition    ✓ Delete Element    ✓ Clear All    ✓ Export (XMI)

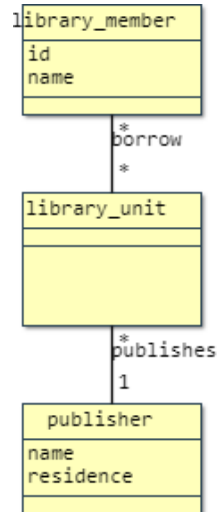
```
classDiagram
    class library_member {
        id
        name
    }
    class library_unit {
        name
        year
        author
    }
    class publisher {
        name
        residence
    }
    library_member "*" -- "*" library_unit : borrow
    library_unit "*" -- "1" publisher : publishes
```

# Illustrative examples of CDM generation



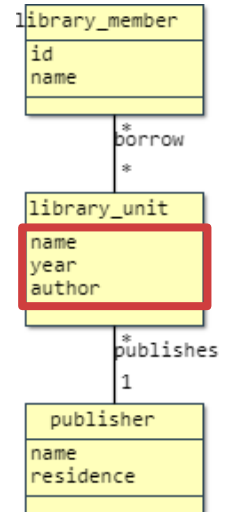
*library member has id and name. library members borrow library units. library unit has name year and author. publisher has name and residence. publisher publishes library units.*

**W/O text modification**



*Library member has id and name. Library members borrow library units. Library unit has name, year and author. Publisher has name and residence. Publisher publishes library units.*

**After text modification**



# Conclusion and future work

- In this paper, we presented the first online web-based tool which enables **conversion of a recorded speech into the CDM** represented by a UML class diagram
- CDM derivation is executed in two steps:
  - Step #1: the recorded speech is converted to the corresponding text
  - Step #2: the text is converted to the corresponding CDM by an NLP-based tool which uses external services
- Initial results show that the implemented tool has **drawbacks** compared to the existing tools, but also show that the approach has a huge **potential**

For example, the approach has great potential for practical use in the future since the tool enables CDM extraction directly from the recorded user stories, which could be a very beneficiary in **agile software development**
- The presented prototype constitutes a very **pioneering achievement in the field of speech-driven database design**, so a plethora of open issues should be resolved in the future:
  - Further improvement of the implemented tool
  - Online speech recognition
  - Multilingual support
  - ...