

# A GrAF-compliant Indonesian Speech Recognition Web Service on the Language Grid for Transcription Crowdsourcing

Bayu Distiawan Trisedya & Ruli Manurung  
Faculty of Computer Science  
Universitas Indonesia



# Outline

- Background
- System Overview
- System Scenario
- Tools & Standard
- Another Issue
- Future Work

# Background

- Initial groundwork for developing Indonesian speech recognition systems have been done, but still using small corpus
- Inspired by another crowdsourcing project: PodCastle Project (Goto and Ogata, 2010).



The screenshot shows a Firefox browser window displaying the PodCastle website. The address bar shows the URL: [http://www.youtube.com/watch?v=3DxW\\_rvQcSgk](http://www.youtube.com/watch?v=3DxW_rvQcSgk). The page title is "Podcastle" and the subtitle is "www.youtube.com AIST - Paving the way to the Future AIST is one of the largest public research inst... 2011/10/05". The page includes a navigation menu with "Full text", "Candidates", "Play", and "Stop" buttons. A video player is visible on the left side of the page, showing a thumbnail of a large, curved, metallic sculpture. The main content area displays a list of words and phrases, each with a color-coded background indicating its status: blue for corrected, green for confirmed, and red for potential errors. The words are: (NOISE) (NOISE) (NOISE) (NOISE) (JAPANESE) (JAPANESE) (JAPANESE) (JAPANESE) (MUSIC) (MUSIC) is, and so this so oh okay doc about your i as, it's i yes it's dollar ice ha, oh dull isaac is, so duller eyes wa, uh darn i, it dock dog jus, dollars dog sti, don dot doll sta.

**When you find recognition errors, you can contribute by correcting some of them and press "save" for sharing with other users.** You can also confirm that a speech recognition result is correct by pressing "o" on it.

- Corrected (Blue indicates that an error was corrected)
- Confirmed (Green indicates that a speech recognition result was confirmed to be correct)
- Degree of speech-recognition reliability (Red indicates that a speech recognition result might include an error)

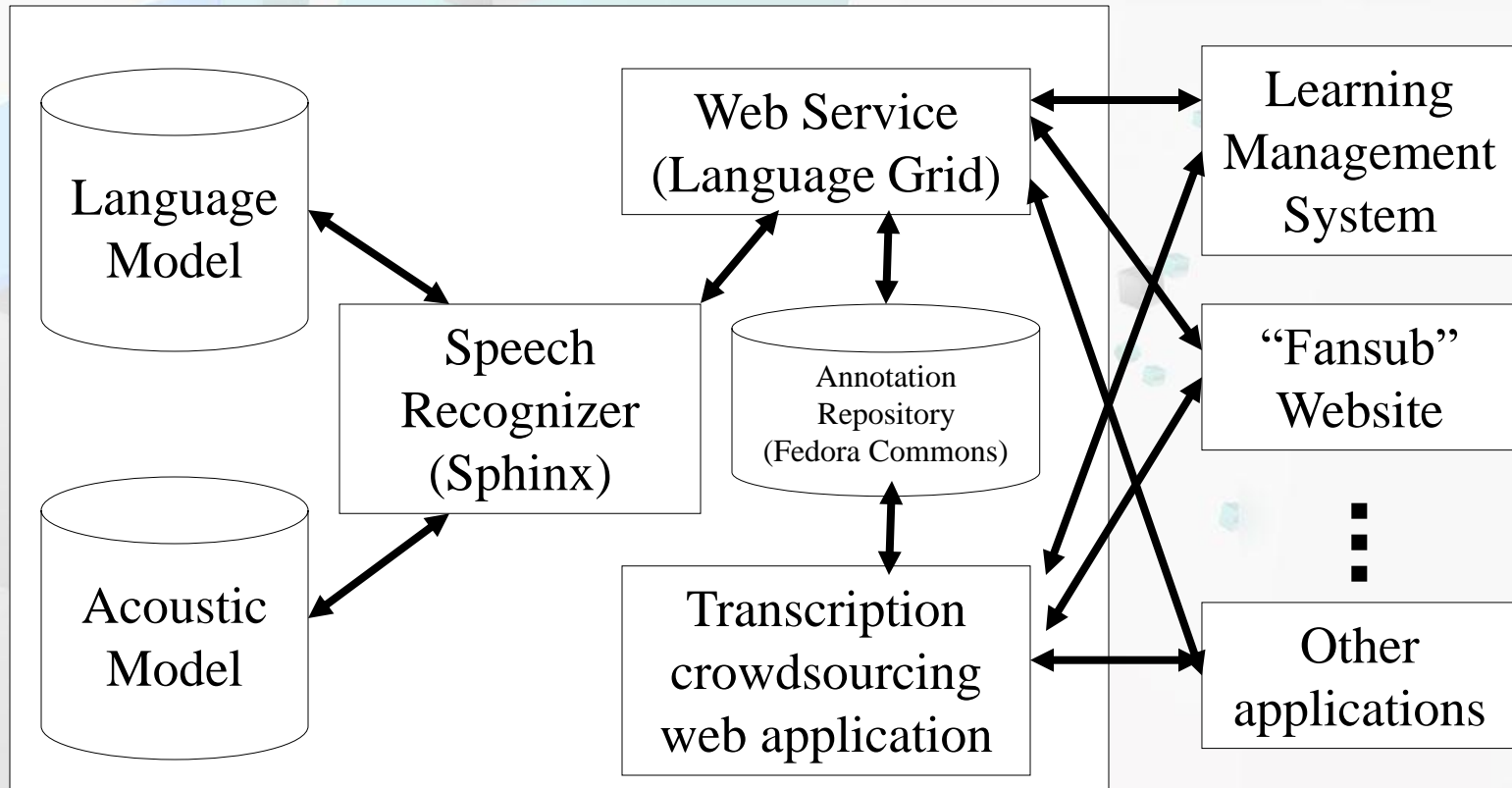
# System Overview

- How to get a large collection of annotated training data in the form of spoken audio data along with validated speech transcriptions? Manual?
- Our solution: provides a valuable service to users, whilst allowing the construction of a large speech corpus
- User can give correction from any arising speech recognition errors

# System Overview

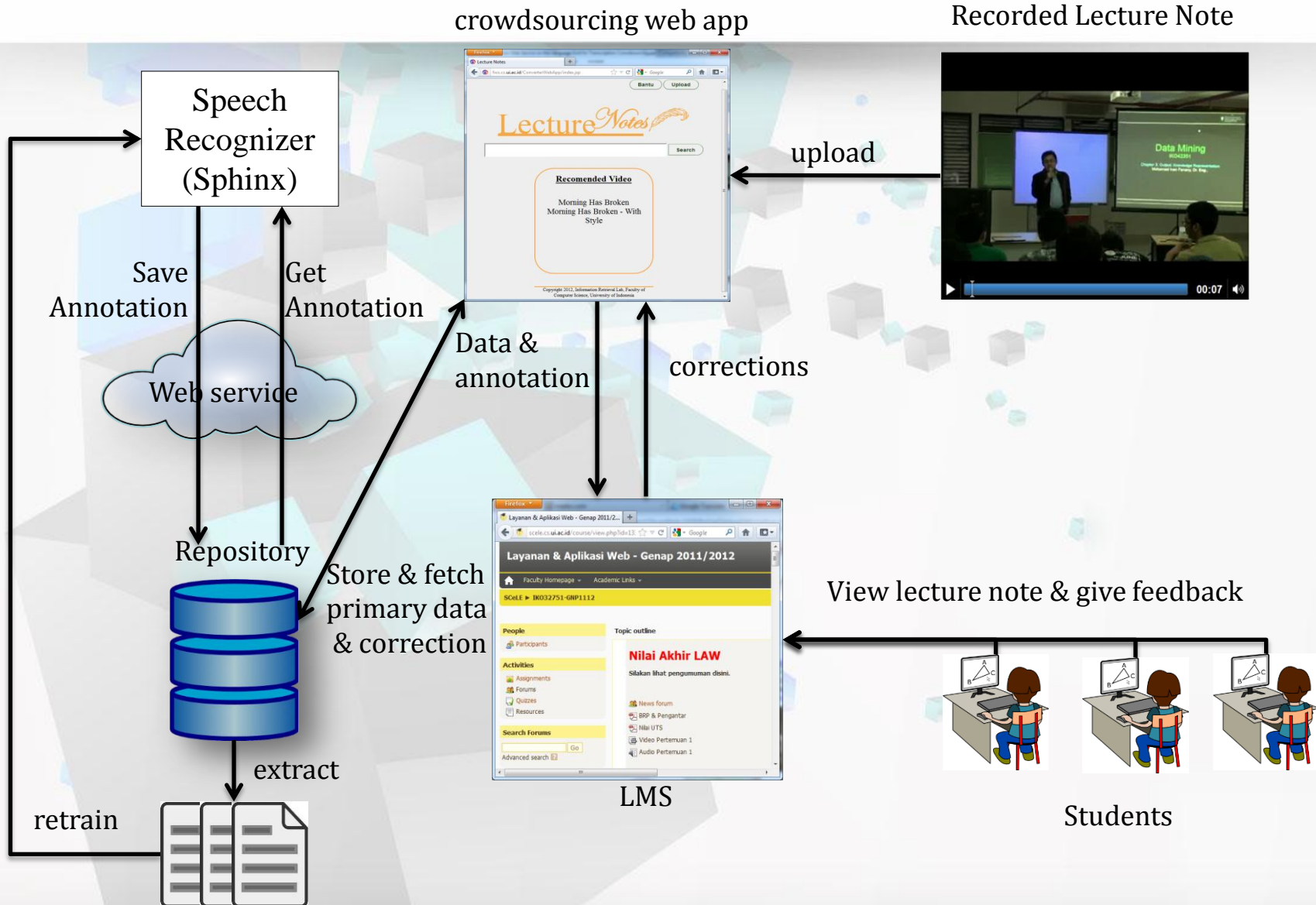
- Using CMU Sphinx to build speech recognition system
- Using Language Grid (Ishida, 2005) to provide interoperable service
- Using Linguistic Annotation Framework (Ide, Romary, 2006) to incorporate data, automatic annotation, & user annotation
- Using Fedora Commons to store data and its annotation (future work)

# System Overview





# System Scenario



# Language Grid Project

- Developed in early 2005
  - National Institute of Information and Communication Technology (NICT), universities and research institutes around Kyoto.
  - Machine translation which includes five languages: Chinese, Malaysian, Japanese, Korean, and English
- Purpose
  - To overcome the language barriers that often inhibit communication between people who have different languages
  - Make NLP services can be accessed by public
  - Built an integrated (composite) service



# Language Grid Resources

Resource Name	Resource Type	Provider
An Introduction to Schools in Japan: School Guidance for Foreign Guardians	Parallel Text	Language Grid Operation Center
Atsugi City School Life Starts Here	Parallel Text	Language Grid Operation Center
Aya and Musashi's Textbook: Japanese-Learning Aid	Parallel Text	Language Grid Operation Center
Bahasa Indonesia Morphological Analysis	Morphological Analyzer	Information Retrieval Lab, Faculty of Computer Science, University of Indonesia
Bilingual Dictionary With Longest Matching Cross Search Service Java	Bilingual Dictionary	Ishida and Matsubara Laboratory
BLEU	Similarity Calculator	Language Grid Operation Center

# Langrid Jakarta Operation Center

<http://langrid.cs.ui.ac.id/langrid-2.0/overview>

The screenshot shows a web browser window with the following content:

- Browser Tab:** Jakarta Language Grid Service Manager
- Address Bar:** langrid.cs.ui.ac.id/langrid-2.0/overview
- Header:** Universitas Indonesia  
Jakarta Language Grid Service Manager
- Navigation:** Login
- Menu:** Overview, View of Language Grid, Manual
- Main Content:**
  - Language Grid Service Manager**
  - The Service Manager is a web-based tool to manage the Language Grid for the Language Grid Users and the Operator. This tool allows easy management of user information, user access, language/computer resources and language services. Each role in the Language Grid has different range of control.
  - This site is compatible with Internet Explorer 7.0 and Firefox.
  - All Language Grid Users**
  - All Language Users access the information
      - [News](#)
      - You can access the operation history of the Language Grid, such as registration/suspending/resuming/deletion of language resources and computation resources.
      - [Language Grid Users](#)

# GrAF (Ide and Suderman, 2007)

- One of the formats that implement the conceptual standard annotation of the Language Annotation Framework (LAF)
- Used in our system to represent annotation for speech recognition
- Incorporate automatic annotation & crowdsourced user annotation

# GrAF

- Multiple annotation

```
<graph>
  <edgeSet id="Speech Segmentation">
    <instant id="e1" from="3" to="5"/>
    ...
  </edgeSet>
  <edge id="t1" ref="e1">
    <fs type="token">
      <f name="word" sVal="sedikitnya"/>
    </fs>
  </edge>
  <edge id="t2" ref="e1">
    <fs type="token">
      <f name="word" sVal="sedikit"/>
    </fs>
  ...
</graph>
```

# GrAF

- Overlapping segmentation

```
<graph>
  <edgeSet id="Speech Segmentation">
    <instant id="e1" from="4.02" to="4.3"/>
    <instant id="e2" from="4.02" to="4.2"/>
    <instant id="e3" from="4.2" to="4.3"/>
    ...
  </edgeSet>
  <edge id="t1" ref="e1">
    <fs type="token">
      <f name="word" sVal="sedikitnya"/>
    </fs>
  </edge>
  <edge id="t2" ref="e2">
    <fs type="token">
      <f name="word" sVal="sedih"/>
    </fs>
  </edge>
  <edge id="t3" ref="e3">
    <fs type="token">
      <f name="word" sVal="kita"/>
    </fs>
  </edge>
  ...
</graph>
```

# Example of User Interface Design

## Lecture Notes

Telusuri

\*click pada kata untuk memilih kandidat kata \*double click pada kata untuk edit secara manual

### Data Mining



selamat sore assalamualaikum warahmatullahi wabarakatuh ee salam sejahtera buat teman teman semuanya jadi hari ini kita memasuki chapter ketertaman teman teman mungkin menyesuaikan dengan silabus kita a teman teman gitu ya jadi target saya hari ini chapter tiga selesai jadi kita bisa masuk algoritma untuk mulai dari chapter empat untuk pekan depan begitu oke ee kita langsung aja ya jadi machine learning itu kita bisa lihat adalah sebuah sistem ya yang dia itu ada input ada outputnya begitu ya inputnya apa itu sudah dipelajari waktu kita di ee chapter dua kemarin ya sepekan lalu kemudian outputnya itu kita akan pelajari hari ini ya sekedar review kita ingat lagi input dari machine learning itu adalah apa aja itu konsep dalam artian

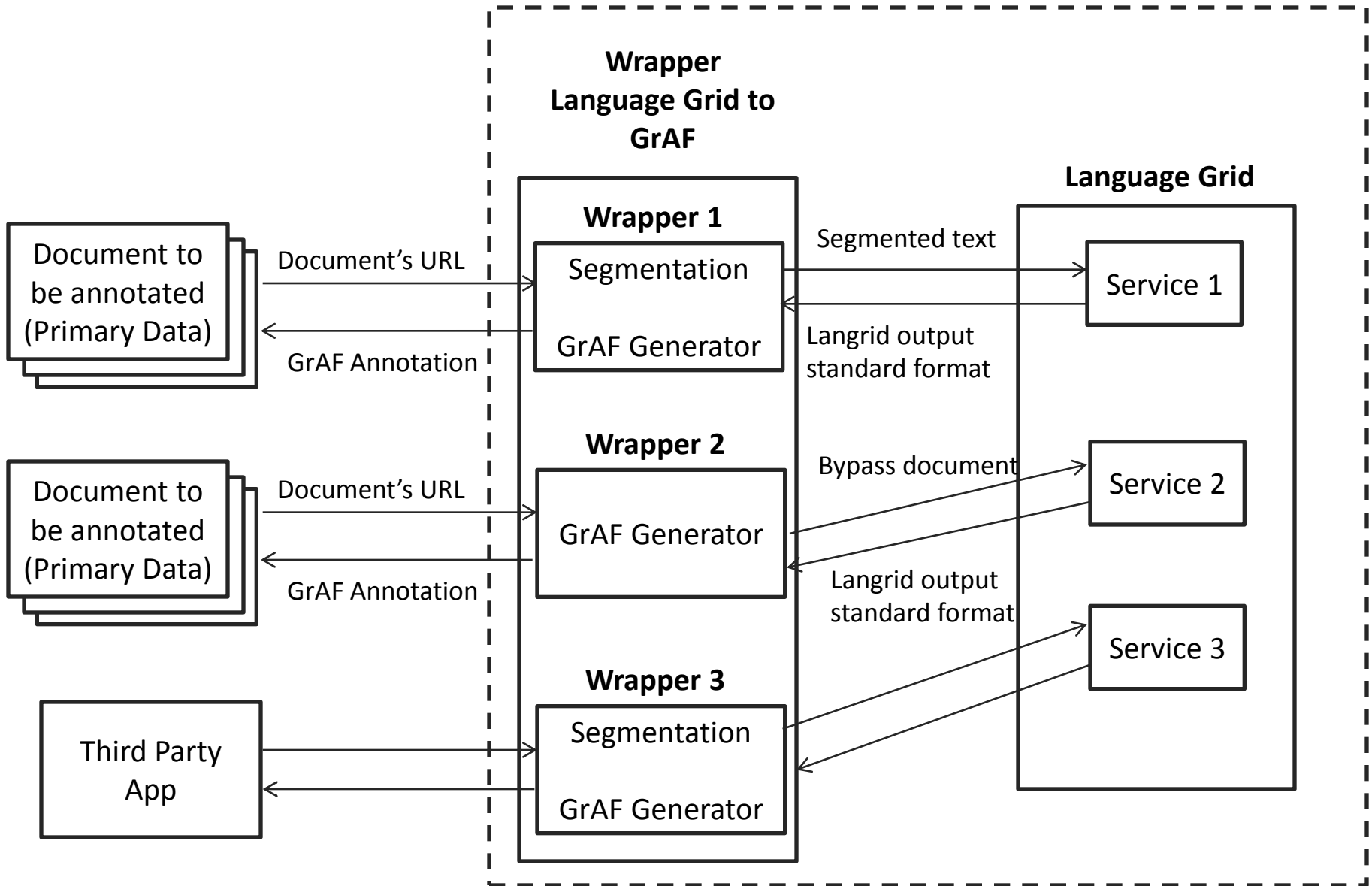
Save



# Integration of Language Grid Web Service and GrAF-based Annotation for Speech Recognition

- Develop GrAF-aware Language Grid framework (Distiawan and Manurung, 2010)
  - Segmentation: For speech data, segmentation will be defined in terms of the timestamps when utterances occur in the primary media file. Thus, an utterance token is marked with an edge tag, and contains information about the beginning and end timestamps.
  - Communication with the web services on the Language Grid. Invoke the existing speech recognition system in Language Grid Infrastructure
  - Mapping of the Language Grid service output to the initial segmentation ->Consistency in document segmentations

# GrAF Aware Language Grid



# Sample GrAF segmentation and annotation from the speech recognizer

```
<container xmlns:graf="http://www.tc37sc4.org/graf/v1.0.6b">
  <header>
    <primaryData loc="http://fws.cs.ui.ac.id/fedora/objects/Speech:1/datastreams/FILE/content" type="audio/wav"/>
  </header>
  <graph>
    <edgeSet id="Speech Segmentation">
      <instant id="e1" from="0.35" to="0.7"/>
      <instant id="e2" from="0.7" to="1.15"/>
      <instant id="e3" from="1.15" to="1.57"/>
      ...
    </edgeSet>
    <edge id="t1" ref="e1">
      <fs type="token">
        <f name="word" sVal="lima"/>
      </fs>
    </edge>
    <edge id="t2" ref="e2">
      <fs type="token">
        <f name="word" sVal="empat"/>
      </fs>
    </edge>
    ...
  </graph>
</container>
```

# Store data & annotation

- Use our previously developed corpus repository (Manurung et al., 2010) -> CORE
- Store all audio or video data along with its automatic or crowdsourced GrAF annotation

<b>Multimedia Document</b>
<b>Datastream</b>
<b>Video</b>
<b>Audio</b>
<b>GrAF Sphinx Annotation</b>
<b>GrAF Crowdsourced Annotation 1</b>
<b>...</b>
<b>GrAF Crowdsourced Annotation n</b>

# Crowdsourcing audio transcriptions

- User interface design: displaying the transcriptions
- Crowdsourcing incentive scheme: learning management system
- Utilizing user corrections: retraining the acoustic and language models of speech recognition system

# Further Work and Summary

- Tune up the speech recognition system -> speed up processing time
- Update to the latest GrAF standard
- Determine procedure to choose the best crowdsourced annotation to retrain speech recognition system





**THANK YOU**