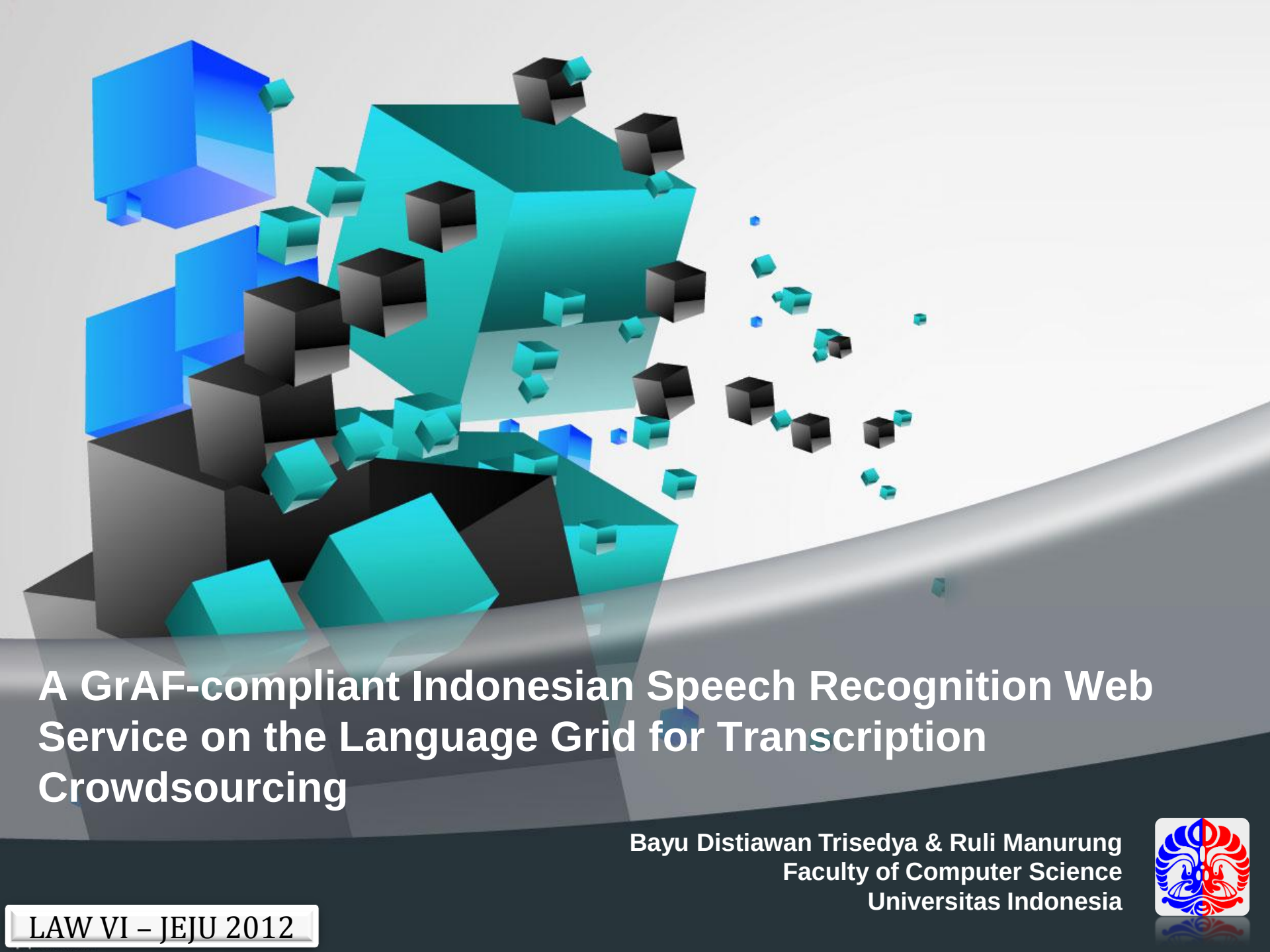


A GrAF-compliant Indonesian Speech Recognition Web Service on the Language Grid for Transcription Crowdsourcing

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

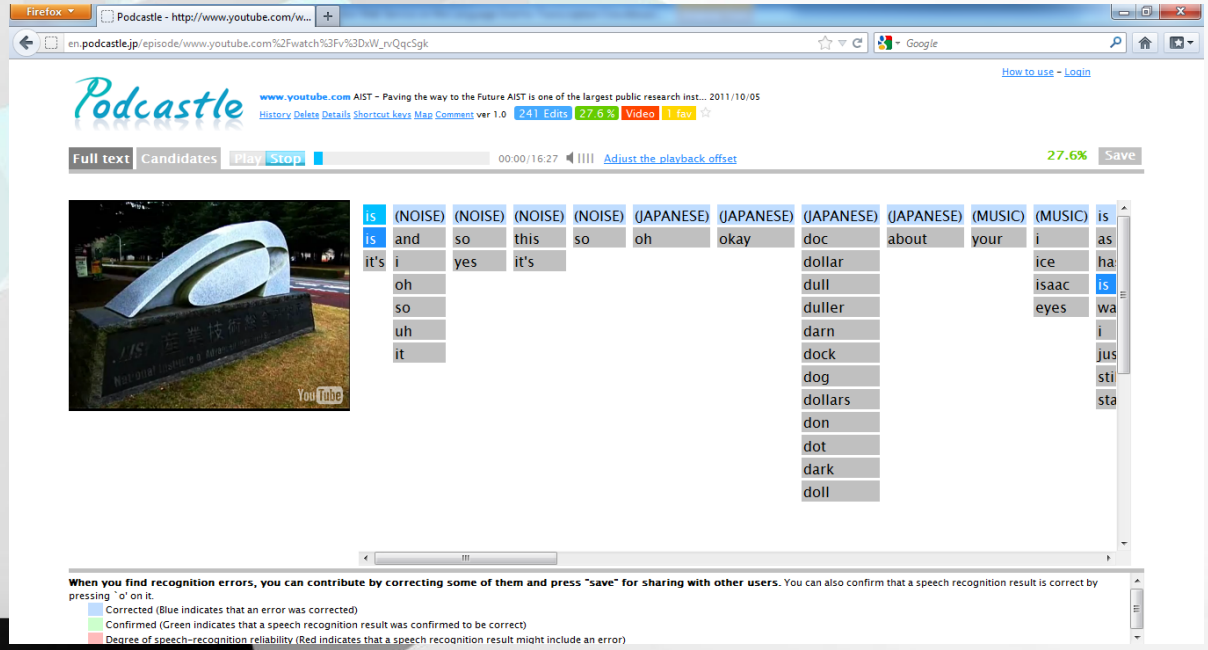
LAW VI – JEJU 2012

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 crowdsource project: PodCastle Project (Goto and Ogata, 2010).



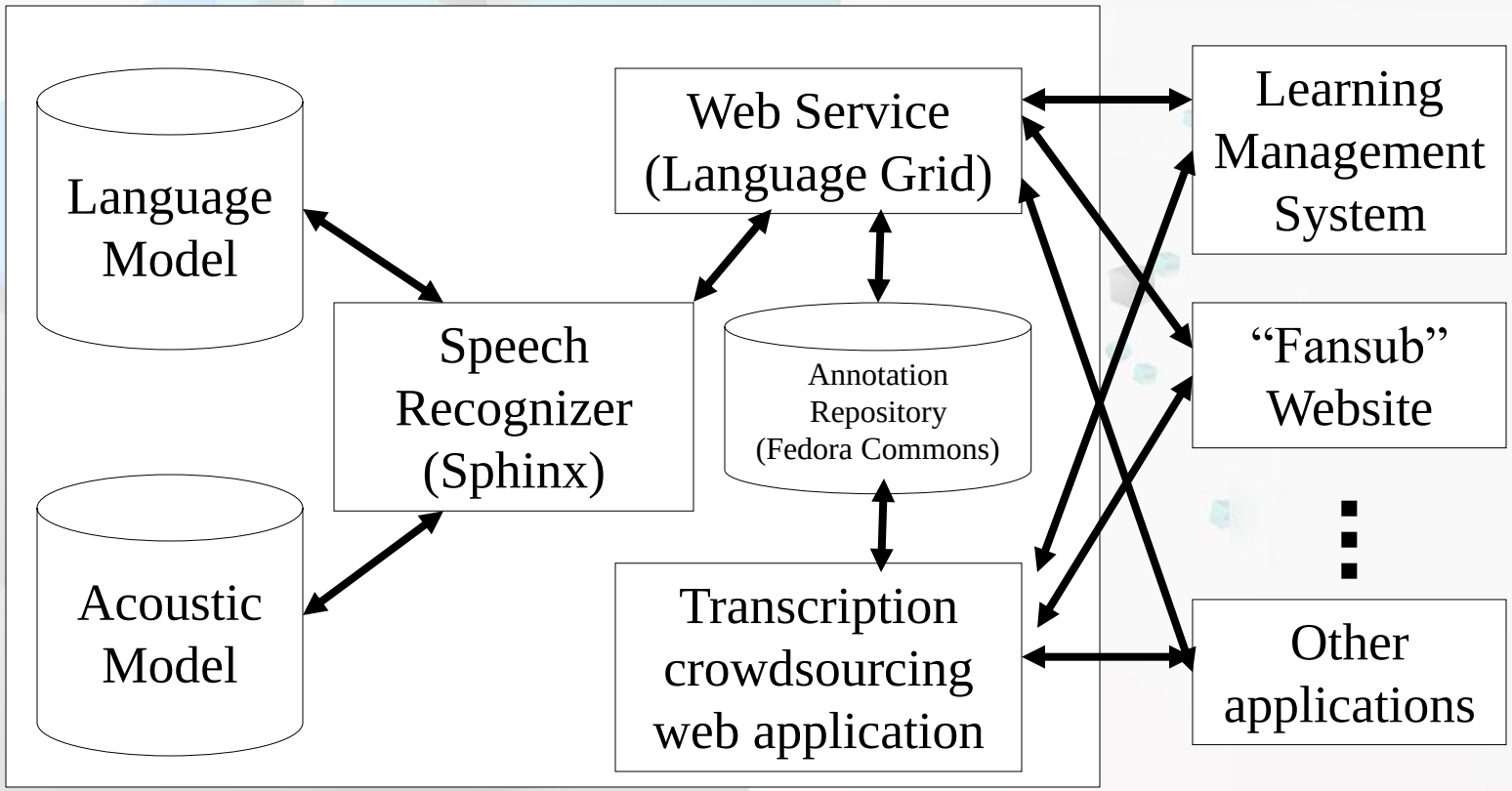
System Overview

- 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

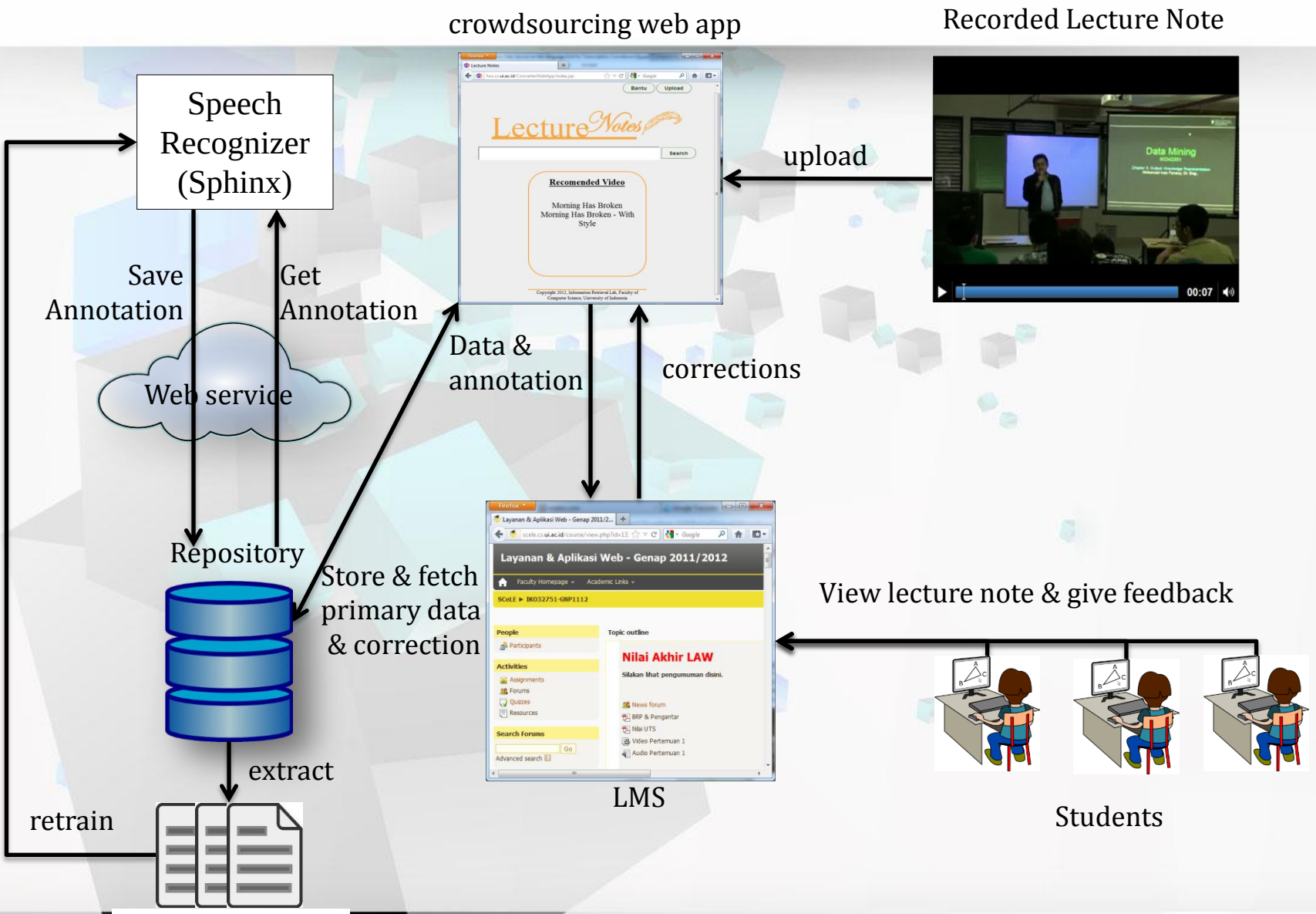
System Overview

- 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

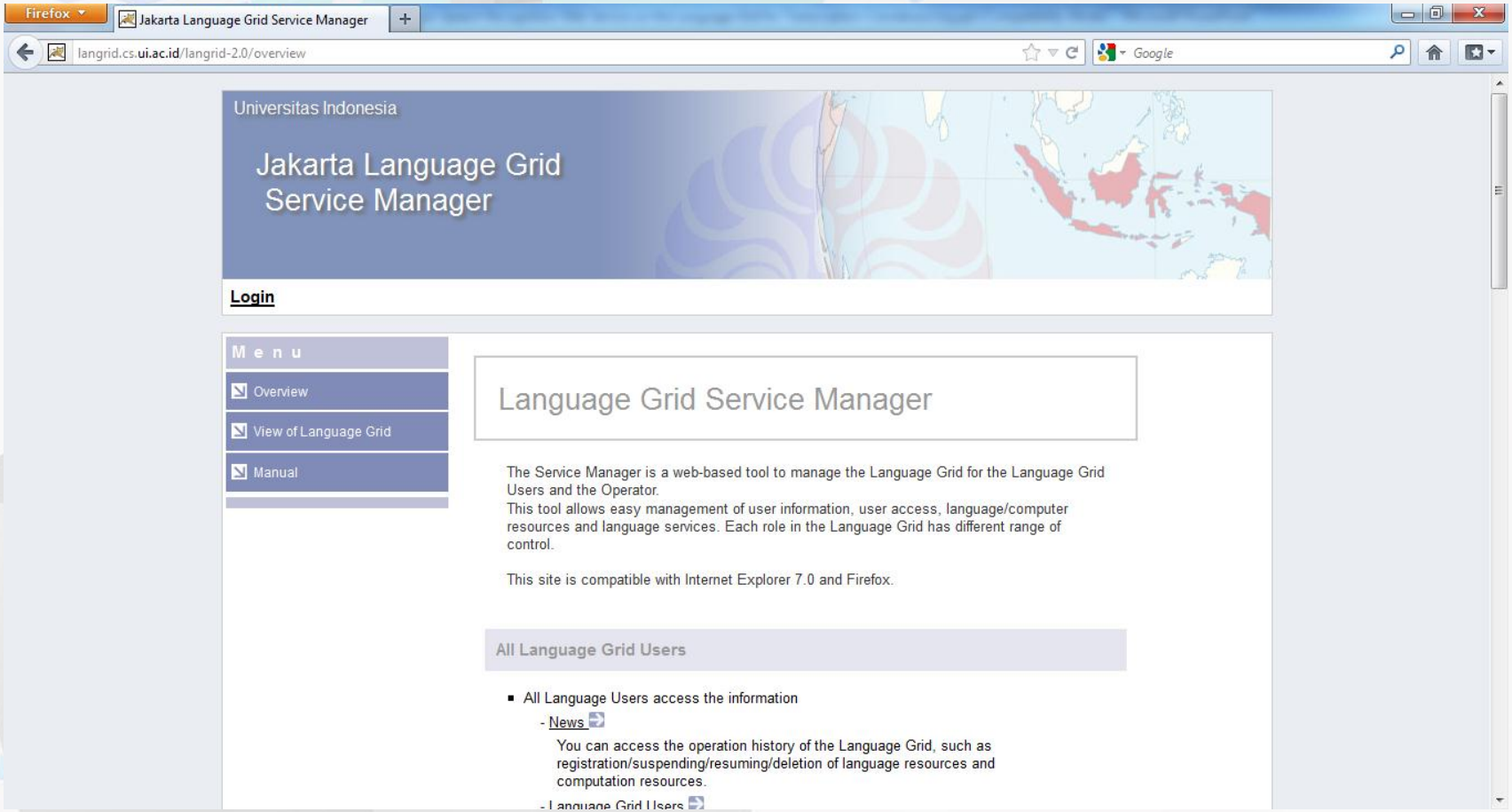
- Developed in early 2005
  - National Institute of Information and Communication Technology (NICT), universities and research institutes around Kyoto.
  - Machine translation which includes five languages: Chinese, Malaysian, Japanese, Korean, and English
- Purpose
  - To overcome the language barriers 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



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

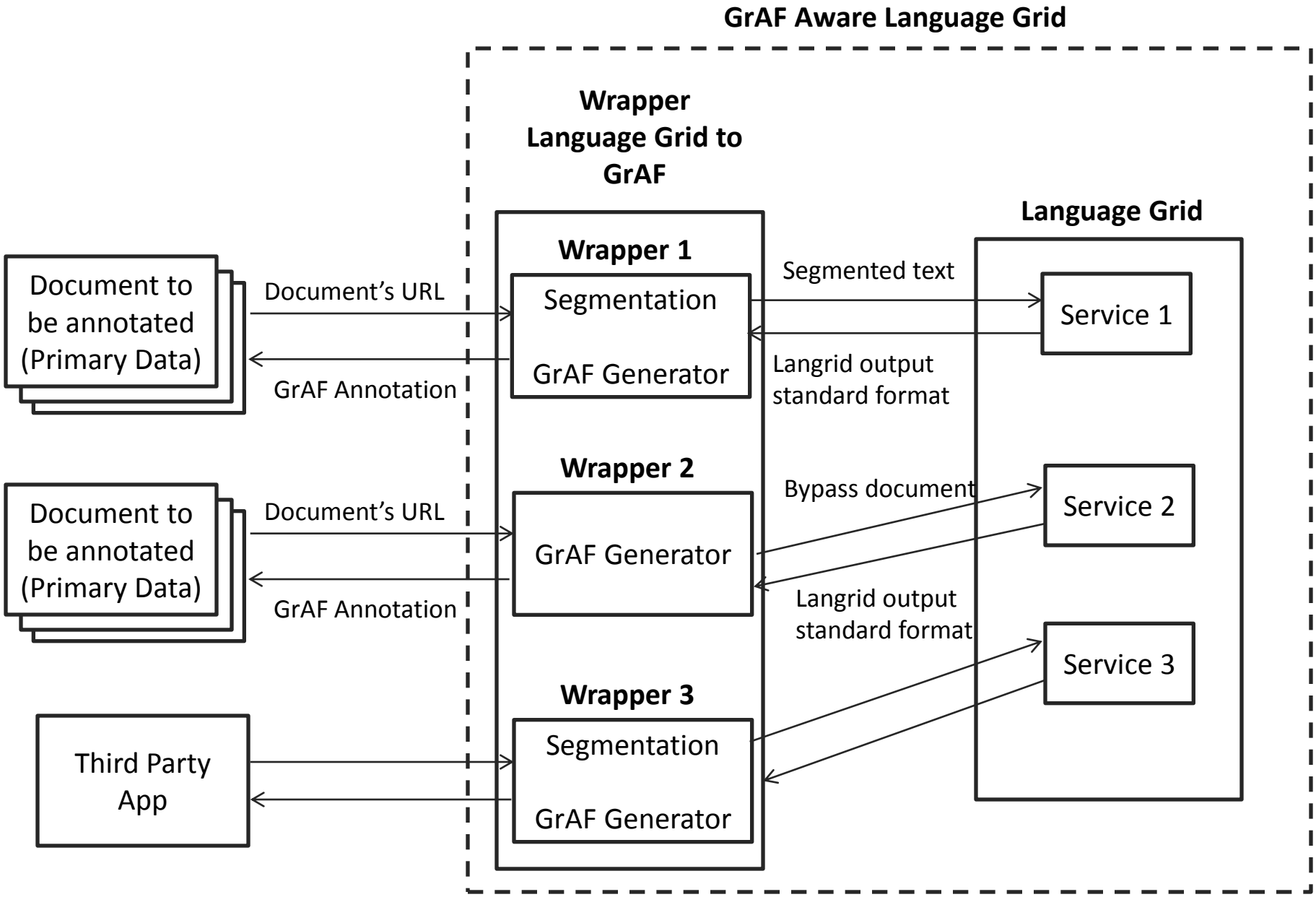
Data Mining

selamat sore assalamualaikum warahmatullahi wabarakatuh ee salam sejahtera buat teman teman semuanya jadi hari ini kita memasuki chapter ket taman gkin menyesuaikan dengan silabus kita a 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

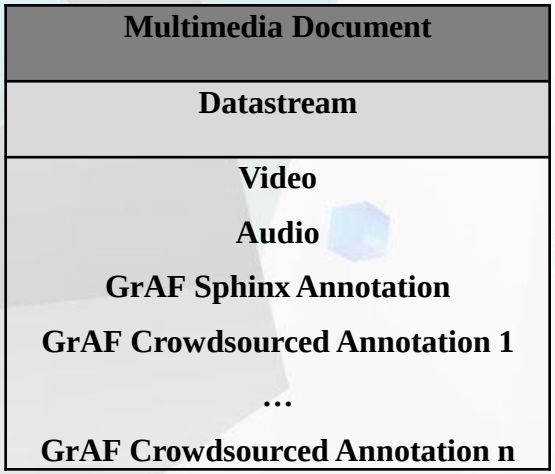

GrAF Aware Language Grid
Wrapper
Language Grid to
GrAF
Language Grid
Wrapper 1
Segmentation
GrAF Generator
Wrapper 2
GrAF Generator
Wrapper 3
Segmentation
GrAF Generator
Document to be annotated (Primary Data)
Document's URL
GrAF Annotation
Document to be annotated (Primary Data)
Document's URL
GrAF Annotation
Third Party App
Segmented text
Langrid output standard format
Bypass document
Langrid output standard format
Service 1
Service 2
Service 3

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



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 crowdsource annotation to retrain speech recognition system

THANK YOU