



---

Information Retrieval Lab  
Faculty of Computer Science  
University of Indonesia

---

# Language Computing Technology : HR Development from Indonesian Perspectives

Mirna Adriani

15 Januari 2009



---

# Faculty of Computer Science University of Indonesia

---

## ■ Research Labs

- Computational Intelligence
- Digital Library & Distance Learning
- Formal Method in Software Engineering
- High Performance & Grid Computing
- Image Processing & Geographic Information System
- **Information Retrieval**
- IT-Governance
- E-Government



# Research Topics in Bahasa Indonesia

- Electronic Dictionary (Kamus Besar Bahasa Indonesia Elektronik)
  - Collaboration work with Pusat Bahasa (Center for Bahasa Indonesia) (1994-1995)
- Spelling Checker for Bahasa Indonesia (1994-1995)
- Electronic Glossary for Computer Science (1994)





## Research Topics in Bahasa Indonesia

### ● 1993-1996 : Word Morphology

#### ■ *Stemmer*

● Example: dimainkan : **di – main – kan**

#### ■ word generator

● Misal : lari : **melarikan, dilarikan, berlari**

#### ■ Word segmentation

● Misal : seminar = **se.mi.nar**



# Research Topic : Information Retrieval

## ● Textual

- Cross-Language Information Retrieval (CLIR)
  - Indonesian ↔ English, German, Dutch, Italian etc.
- Summarization
- Geographic Information Retrieval (GIR)
- Question Answering (QA)
- Classification
- Web-IR

● <http://ir.cs.ui.ac.id>



# Research Topic : Information Retrieval

## ● Multimedia

- Image

- Music

- Speech

  - Speech Retrieval

  - Topic Segmentation

  - Speaker Identification

  - Speech Recognition

## ● Peer-to-Peer (P2P) Information retrieval

- IR system on Adaptive P2P system



---

# Research Topic : Computational Linguistics

---

## ● NLP Tools

- Semantic Analyzer
- <http://bahasa.cs.ui.ac.id/resources.php>
- Symbolic Parser
- Morphological Analyzer
- Indonesian Wordnet
  - <http://bahasa.cs.ui.ac.id/iwn>
- Named Entity Recognition
- Part of Speech Tagging (POSTAG)



**Welcome user!**



Edit profile



List of Synsets



Logout

*pengembangan wordnet untuk bahasa Indonesia*



FAKULTAS ILMU KOMPUTER  
UNIVERSITAS INDONESIA



Synset

**Synset ID:** s100543233

**Word(s):**

music

**Category:**

noun

**Definition:**

musical activity (singing or whistling etc.)

**Example:**

1. his music was his central interest

[Previous Synset](#)

[Next Synset](#)

**Synset and Candidate Mapping**

**Search Kamus Besar Bahasa Indonesia**

All checked & unchecked

**Definisi kata hasil usulan/saran**

	No.	Kata	Definisi	Contoh
<input type="checkbox"/>	1	<b>musik</b> kata sifat	ilmu atau seni menyusun nada atau suara di urutan, kombinasi, dan hubungan temporal untuk menghasilkan komposisi (suara) yg mempunyai kesatuan dan kesinambungan	-
<input type="checkbox"/>	2	<b>musik</b> kata sifat	nada atau suara yg disusun demikian rupa sehingga mengandung irama, lagu, dan keharmonisan (terutama yg menggunakan alat-alat yg dapat menghasilkan bunyi-bunyi itu)	-



---

## Research Topic : Knowledge Management

---

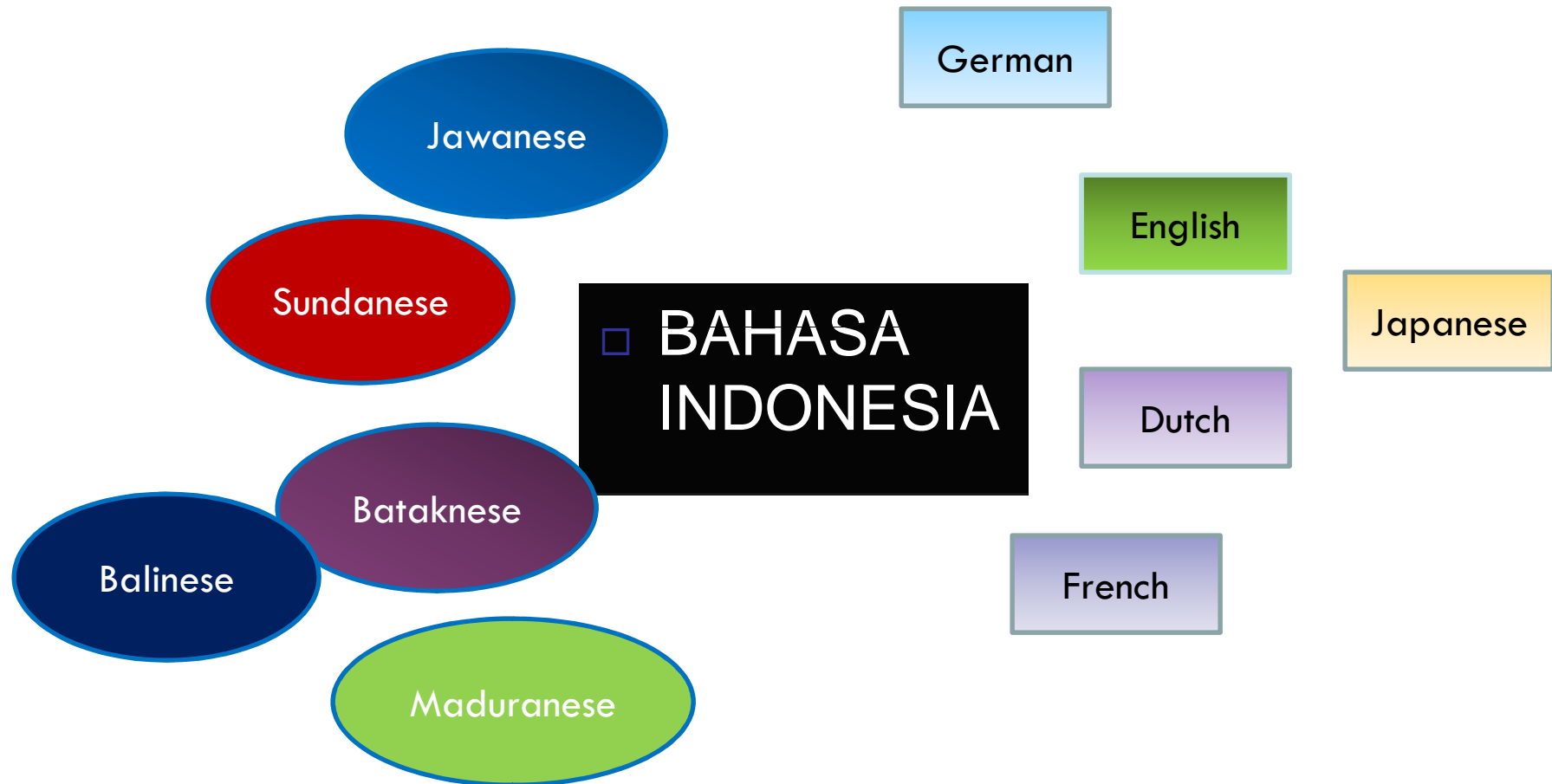
- Semantic Web
- Data Mining
  
- Research work done by undergraduate & graduate students



# LOCAL LANGUAGE RESOURCES IN INDONESIA



# Bahasa Indonesia's Position





# We Have Many Local Languages

- Indonesia has 742 local languages<sup>1</sup>
- We need to preserve our cultures
  - Local languages not only different languages but also different alphabets
- Information about local languages need to be available so it can be accessed easily

<sup>1</sup> <http://www.ethnologue.com> Fakultas Ilmu Komputer – Universitas Indonesia



# Culture Preservation

- Information about local languages is not available widely

မာဘာအိအိ မာဘာအိအိ

- Need an effort to preserve local languages

မာဘာအိ အိအိ

- If information is available in digital format, it can be accessed easily.



# CLIR for Local Languages

- We can search information written in local languages using Bahasa Indonesia or foreign languages
- How about developing machine translation for local languages?
  - Javanese-Indonesian: is there any resources?
- Lack of research funding for research related to language topics



# Research Collaboration

- University of Sydney
  - Corpus Repository Website
  - Finite-state Morphological Analyzer
  
- Kyoto University
  - Web-services for language resources : Language Grid Project
  
- PANL10n Project
  
- More collaboration with other universities



# Research Collaboration

## ● MALINDO Workshop

- Malaysian Indonesian Language Engineering Workshop
- NUS, Singapore (Stephane Bressan)
- MMU, Malaysia (Bali Ranaivo-Malancon)

## ● THIRD MALINDO WORKSHOP

- Collocated Workshop in ACL 2009 (Singapore)
- 12 August 2009

● <http://www.acl-ijcnlp-2009.org/main/collocatedevents.html>

● <http://www.comp.nus.edu.sg/~malindo/>



---

● Thank You