

WordNet (Bahasa)

Pangkalan data leksikal (bahasa Indonesia/Melayu)

David Moeljadi

28 Juli 2016

Jurusan Linguistik dan Kajian Multibahasa, Fakultas Humaniora dan Ilmu Sosial,
Universitas Teknologi Nanyang

–SEMINAR LEKSIKOGRAFI INDONESIA, HOTEL SANTIKA TAMAN MINI INDONESIA INDAH, JAKARTA–

Perkenalan diri

David Moeljadi

2014 - (2018?) S3 Linguistik di Universitas Teknologi Nanyang

1986 lahir - TK - SD - SMP - SMA(Jurusan Bahasa) di Malang, Jawa Timur

2004 - 2005 S1 Sastra Jepang Binus, Jakarta

2005 - 2006 Pusat Bahasa dan Kebudayaan Jepang, Universitas Kajian Asing Osaka

2006 - 2010 S1 Linguistik di Univ. Tokyo

2010 - 2012 S2 Linguistik di Univ. Tokyo

2012 - 2013 Rakuten Travel, Inc.

2013 - 2014 Mahasiswa peneliti di Lembaga Penelitian Bahasa dan Budaya Asia dan Afrika, Universitas Kajian Asing Tokyo

compling.hss.ntu.edu.sg/who/david/

Proyek saat ini

1. Indonesian Resource Grammar (INDRA)
github.com/davidmoeljadi/INDRA
chimpanzee.ling.washington.edu/demophon/indra/
delph-in.github.io/delphin-viz/demo/
sweaglesw.org/linguistics/objects-demo
2. Wordnet Bahasa
3. NTU Multilingual Corpus
compling.hss.ntu.edu.sg/ntumc/
4. Indonesian Loanword Search Engine
getnocms.com/loanword_searchengine
5. Kamus Pemelajar Kanji Jepang-Indonesia
6. KBBI IV
7. Analisis sentimen teks bahasa Indonesia

WordNet (Bahasa)

1. WordNet: Apa itu?
2. WordNet: Apa saja kegunaannya?
3. Wordnet Bahasa dan berbagai Wordnet lainnya
4. Wordnet Bahasa: Sejarah
5. Wordnet bahasa Abui
6. Pemecahan struktur data KBBI Edisi Keempat
7. Bengkel

WordNet: Apa itu?

- Laboratorium Ilmu Kognitif Universitas Princeton
- Jaringan semantik leksikal untuk bahasa Inggris (Amerika)
- Entri leksikal disusun berdasarkan makna

Bagi pengguna: cara baru menggunakan kamus untuk mencari dan mempelajari kata

Bagi komputer: sumber untuk melakukan analisis semantik pada teks bahasa alami

SYNSET “SYNONYM SET” / HIMPUNAN SINONIM

- Satuan dasar dalam WordNet
- Tiap *synset* mewakili sebuah konsep
- Makna kata diwakili oleh:
 - sinonim
 - kalimat definisi
 - hubungannya dengan *synset* lainnya
- Anggota *synset* adalah kata-kata berkonsep sama
- Kata yang bermakna lebih dari satu menjadi anggota lebih dari satu *synset*
- 5 kategori *synset*:
n (nomina), *v* (verba), *a* (adjektiva), *r* (adverbia),
x (kata penggolong, kata seru, gelar/sebutan)

Contoh kata bermakna banyak dan memiliki *synset* lebih dari satu

<noun.group> **court**, tribunal, judicature – (an assembly (including one or more judges) to conduct judicial business)

<noun.group> **court**, royal_court – (the sovereign and his advisers who are the governing power of a state)

<noun.artifact> **court** – (a specially marked horizontal area within which a game is played; "players had to reserve a **court** in advance")

...

HUBUNGAN ANTAR-SYNSET NOMINA

hipernim istana *adalah* tempat tinggal, kediaman, rumah

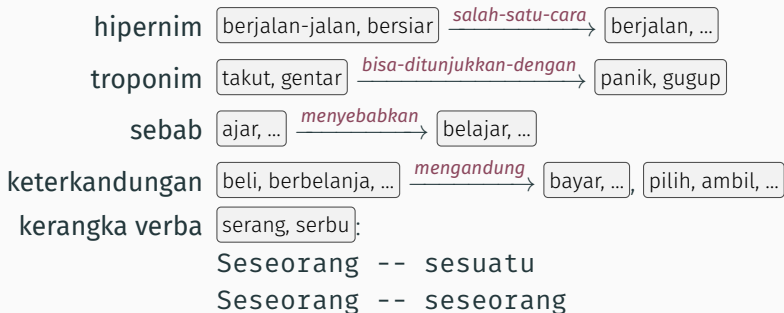
holonim jari *bagian-dari* tangan

tepung *bahan-untuk* roti, adonan, kue

juri *anggota-dari* mahkamah, pengadilan, tribunal

contoh Mozart, Wolfgang Amadeus Mozart *contoh* komponis, komposer, ...

HUBUNGAN ANTAR-SYNSET VERBA

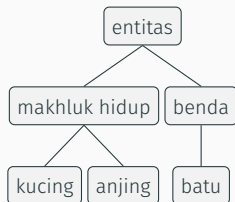


HUBUNGAN LEKSIKAL

- antonim** «keburukan» $\xrightarrow{\times}$ «keindahan», «tarik» $\xrightarrow{\times}$ «dorong»,
«sulit, sukar» $\xrightarrow{\times}$ «mudah, gampang», «segera» $\xrightarrow{\times}$ «perlahan»
- atribut** «kekuatan» $\xrightarrow{\text{beratribut}}$ «rapuh», «tahan lasak», «lemah», «kuat»
- derivasi** «maintain» $\xrightarrow{\text{berderivasi-menjadi}}$ «maintainable», «maintenance»,
«maintainer»
- domain** «linguistik» $\xrightarrow{\text{memiliki-istilah}}$ «sintaksis», «analitis», «sintetis», ...
«Jepang» $\xrightarrow{\text{memiliki-istilah}}$ «sumo», «karate», «origami», ...
«bahasa percakapan» $\xrightarrow{\text{memiliki-istilah}}$ «kongkalikong», «nyerocos», ...
- pertainim** «biannual» $\xrightarrow{\text{berkaitan-dengan}}$ «year»,
«ancestral» $\xrightarrow{\text{berkaitan-dengan}}$ «ancestor»,
«Liverpudlian» $\xrightarrow{\text{berkaitan-dengan}}$ «Liverpool»

WordNet: Apa saja kegunaannya?

- nomina di WordNet membentuk hierarki hipernim 'adalah'
- verba di WordNet membentuk hierarki hipernim 'salah-satu-cara'
- Skor kesamaan: mengukur kesamaan dua *synset*



- $Sim(kucing, anjing) = 0.8667$
- $Sim(kucing, batu) = 0.4545$

`maraca.d.umn.edu/cgi-bin/similarity/similarity.cgi`

- Apa makna suatu kata dalam konteks tertentu?

'stars and moons'

- **star#n#1** = (astronomy) a celestial body of hot gases that radiates energy derived from thermonuclear reactions in the interior
- **star#n#2** = someone who is dazzlingly skilled in any field
- **star#n#3** = any celestial body visible (as a point of light) from the Earth at night
- ...
- **moon#n#1** = the natural satellite of the Earth
- ...
- **moon#n#6** = any natural satellite of a planet

PEMROSESAN TEKS MENGGUNAKAN UKURAN KESAMAAN (2)

'stars and moons'

$\text{Sim}(\text{star}\#\text{n}\#1, \text{moon}\#\text{n}\#1) = 0.8235$

$\text{Sim}(\text{star}\#\text{n}\#1, \text{moon}\#\text{n}\#6) = 0.8235$

$\text{Sim}(\text{star}\#\text{n}\#3, \text{moon}\#\text{n}\#1) = 0.8235$

$\text{Sim}(\text{star}\#\text{n}\#3, \text{moon}\#\text{n}\#6) = 0.8235$

$\text{Sim}(\text{star}\#\text{n}\#2, \text{moon}\#\text{n}\#5) = 0.6667$

$\text{Sim}(\text{star}\#\text{n}\#6, \text{moon}\#\text{n}\#5) = 0.625$

...

star#n#1 = (astronomy) a celestial body of hot gases that radiates energy derived from thermonuclear reactions in the interior

'stars and celebrities'

- **star#n#1** = (astronomy) a celestial body of hot gases that radiates energy derived from thermonuclear reactions in the interior
- **star#n#2** = someone who is dazzlingly skilled in any field
- **star#n#3** = any celestial body visible (as a point of light) from the Earth at night
- ...
- **celebrity#n#1** = a widely known person
- **celebrity#n#2** = the state or quality of being widely honored and acclaimed

'stars and celebrities'

$\text{Sim}(\text{star}\#n\#2, \text{celebrity}\#n\#1) = 0.6667$

$\text{Sim}(\text{star}\#n\#6, \text{celebrity}\#n\#1) = 0.625$

$\text{Sim}(\text{star}\#n\#4, \text{celebrity}\#n\#1) = 0.5882$

$\text{Sim}(\text{star}\#n\#1, \text{celebrity}\#n\#1) = 0.5263$

$\text{Sim}(\text{star}\#n\#3, \text{celebrity}\#n\#1) = 0.5263$

$\text{Sim}(\text{star}\#n\#5, \text{celebrity}\#n\#2) = 0.4706$

...

$\text{star}\#n\#2$ = someone who is dazzlingly skilled in any field

MASALAH UKURAN KESAMAAN BERDASARKAN HIERARKI

- Beberapa konsep yang berhubungan ada di bagian hierarki yang berbeda

• pistol $\xrightarrow{\text{adalah}}$ benda

• batu $\xrightarrow{\text{adalah}}$ benda

• perang $\xrightarrow{\text{adalah}}$ peristiwa

$$\text{Sim}(\text{pistol}, \text{batu}) = 0.6667$$

$$\text{Sim}(\text{pistol}, \text{perang}) = 0.6316$$

- Atau memiliki kelas kata yang berbeda, di hierarki yang sama sekali berbeda

UKURAN KEDEKATAN MENGGUNAKAN KALIMAT MAKNA

Menghitung **seberapa dekat kalimat makna synset** (dan hipernim synset tersebut)

Kalimat makna

- **gun** a weapon that discharges a missile at high velocity (especially from a metal tube or barrel)
 - $\xrightarrow{\text{is-a}}$ **weapon, arm, weapon system** any instrument or instrumentality used in fighting or hunting
- **rock** a lump or mass of hard consolidated mineral matter
- **war, warfare** the waging of **armed** conflict against an enemy

$$\text{Rel}(\text{gun}, \text{rock}) = 15$$

$$\text{Rel}(\text{gun}, \text{war}) = 17$$

`translate.google.com/about/intl/id_ALL/license.html`

.

Mengelompokkan kata-kata terjemahan alternatif

- Gratis
 - Kendala: sambungan internet
 - Solusi: aplikasi luring ponsel
- Multibahasa
- Informasi semantik, sintaks, definisi, contoh, domain, dll.
- Tautan derivasi kata
- Pembuatan kamus multilingual dengan \LaTeX
www.overleaf.com

Wordnet Bahasa dan berbagai Wordnet lainnya

- Banyak WordNet untuk berbagai bahasa telah dikembangkan!
- Pendekatan perluasan:
 - Menerjemahkan *synset* di Princeton WordNet ke bahasa lain
 - Mengambil hubungan makna dari Princeton dan merevisinya
 - Menghubungkan *synset* di wordnet dengan ID di Princeton WordNet

CONTOH: SYNSET BAHASA INGGRIS-BAHASA INDONESIA

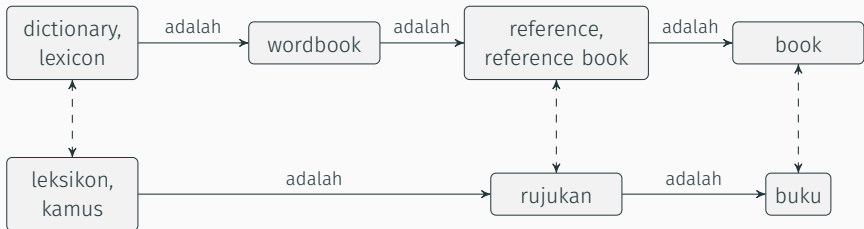
bahasa Inggris

106418901

106418693

106417598

106410904



06418901-n

06417598-n

06410904-n

bahasa Indonesia

- Proyek WordNet awal yang penting:
 - EuroWordNet** (Vossen, 2004) untuk bahasa-bahasa Eropa Barat
 - BalkaNet** (Tufiş, Cristea & Stamou, 2004) untuk bahasa-bahasa Eropa Timur
- Berbagai proyek WordNet yang lain, dengan lisensi yang berbeda
<http://globalwordnet.org/wordnets-in-the-world/>

OPEN MULTILINGUAL WORDNET (BOND & PAIK, 2012)

- <http://compling.hss.ntu.edu.sg/omw/>
- Kumpulan berbagai WordNet dengan lisensi sumber terbuka

Wordnet	Lang	Synsets	Words	Senses	Core	Licence	Data	Citation
Albanet	als	4,676	5,990	9,602	31%	CC BY 3.0	als.zip (+xml)	cite:als ; (bib)
Arabic WordNet (AWN)	arb	10,165	14,595	21,751	48%	CC BY SA 3.0	arb.zip (+xml)	cite:arb ; (bib)
BulTreeBank Wordnet (BTB-WN)	bul	4,999	6,783	9,056	100%	CC BY 3.0	bul.zip (+xml)	cite:bul ; (bib)
Chinese Open Wordnet	cmn	42,312	61,533	79,809	100%	wordnet	cmn.zip (+xml)	cite:cmn ; (bib)
Chinese Wordnet (Taiwan)	qcn	4,913	3,206	8,069	28%	wordnet	qcn.zip (+xml)	cite:qcn ; (bib)
DanNet	dan	4,476	4,468	5,859	81%	wordnet	dan.zip (+xml)	cite:dan ; (bib)
Greek Wordnet	ell	18,049	18,227	24,106	57%	Apache 2.0	ell.zip (+xml)	cite:ell ; (bib)
Princeton WordNet	eng	117,659	148,750	206,978	100%	wordnet	eng.zip (+xml)	cite:eng ; (bib)
Persian Wordnet	fas	17,759	17,560	30,461	41%	Free to use	fas.zip (+xml)	cite:fas ; (bib)
FinnWordNet	fin	116,763	129,859	189,227	100%	CC BY 3.0	fin.zip (+xml)	cite:fin ; (bib)
WOLF (Wordnet Libre du Français)	fra	59,091	55,373	102,671	92%	CeCILL-C	fra.zip (+xml)	cite:fra ; (bib)
Hebrew Wordnet	heb	5,448	5,325	6,872	27%	wordnet	heb.zip (+xml)	cite:heb ; (bib)
MultiWordNet	ita	34,728	40,343	61,558	83%	CC BY 3.0	ita.zip (+xml)	cite:ita ; (bib)
Japanese Wordnet	jpn	57,184	91,964	158,069	95%	wordnet	jpn.zip (+xml)	cite:jpn ; (bib)
Multilingual Central Repository	cat	45,826	46,551	70,622	81%	CC BY 3.0	cat.zip (+xml)	cite:cat ; (bib)
Multilingual Central Repository	eus	29,413	26,240	48,934	71%	CC BY 3.0	eus.zip (+xml)	cite:eus ; (bib)

02121620-n 0 feline mammal usually having thick soft fur and no ability to rear; domestic cats;

Search WN

English

(Malaysian)

Arabic	قط, قطة, كلب, كلبه
Bulgarian	котец
Catalan	gat, gata
Chinese (simplified)	猫
Danish	kat, <i>miskat</i> , <i>mås</i>
Greek	γάτι, γάτα
English	cat, <i>true cat</i>
Basque	kata
Finnish	kissa, <i>kieraeläin</i>
French	chat
Galician	gato
Hebrew	חתול
Indonesian	kucing
Italian	gato
Japanese	猫, ネコ, ネコ, にゃんこやん
Nynorsk	katt, <i>puskatt</i> , <i>pus</i>
Bokmål	katt, <i>puskatt</i> , <i>pus</i>
Polish	kot
Portuguese	gato, <i>gato doméstico</i> , <i>Gato doméstico</i> , <i>bichano</i> , <i>gata</i> , <i>Gato doméstico</i>
Chinese (traditional)	貓
Slovene	mačka
Spanish	<i>felis silvestris canis</i> , <i>gato</i> , <i>gato</i>
Swedish	katt
Thai	แมว
Malaysian	kucing

[Definition](#)

Japanese

猫科、厚く柔らかい毛皮を持ち、吠えることのできないネコ科の哺乳類：家ネコ、ヤマネコ

Bulgarian

Животно от семейство Котки, което със специална козина. — *Котките представляват биологично класификация си, дива котка: домашна котка*

Greek

αἰλουροειδές με εὐέλικτο και ελαστικό σώμα, στρογγυλό κεφάλι, μακριά ουρά και γαμψά νύχια, σχήμα για την αναρρόληση και τη σάλπιγγη του θηλασματος, οικόδοτο και άγριο

English

feline mammal usually having thick soft fur and no ability to rear; domestic cats; wildcats

[Definition](#)

Hyponym: [domestic_cat](#) [wildcat](#)

Hypernym: [feline](#)

Semantic Field: [animal](#),

[animal](#)

SUMO: [c](#) [Feline](#)

TempoWN: ◀ ◻ ▶ (Past: 0.000; Present: 0.000; Future: 0.226)

SentiWN: ▲ ▼ (+0.00 -0.00) ML-SentiCon: ▲ ▼ (+0.00 -0.00)

Search WN

Langs: English

: Catalan

Wordnet Bahasa: Sejarah

- Makna kata diambil dari Kamus Inggeris-Melayu Dewan (KIMD)
- Dihubungkan ke *synset* Princeton WordNet secara manual oleh ahli bahasa dari USM (Siti Khaotijah, Tan Ewe Hoe, Jumiya, Nour Azimah, Norliza Hani, Noor Azlina)
- Tanpa kalimat definisi dalam bahasa Melayu
- 12.429 *synset* untuk nomina dan 5.805 *synset* untuk verba
- Digunakan secara internal untuk penelitian

(*dot*, noun, 1, [small round spot, small circular shape], ⟨titik, bintik⟩). (KIMD)

(10025218-n, *dot*, noun, 1, [a very small circular shape]). (PWN)

- Pendekatan 'Perluasan'
- 25.755 *synsets* (data dari Asian Wordnet)
- 4 penerjemah menggunakan kamus daring dan sistem penerjemahan mesin

- Diperluas secara otomatis menggunakan multi-pivot dan semantic code
- Menghubungkan melalui:
 - Kamus French-English-Malay (FEM)
 - KAMus Melayu Inggeris China (KAMI)
- ke tiga wordnet:
 - Inggris (PWN)
 - Prancis
 - Chinese
- 19.207 *synset*
- Termasuk di dalamnya 5.000 *synset* yang paling banyak digunakan di British National Corpus; diperiksa secara manual

Entry in **FEM***lexical entry*

Malaysian busur

English bow

French arc

Part-of-Speech noun

Hypernym arme 'weapon'

Entry in **KAMI***lexical entry*

Malaysian busur

English bow

Chinese 弓

Part-of-Speech noun

Hypernym ⟨940 : *worktool*⟩

Wordnet candidates (only two of many)

a.

<i>synset</i>	
Lexemes	[English bow Chinese 弓 French arc]
Part-of-Speech	noun
Relations	[Hypernym <i>weapon</i> ...]
Definition	a weapon for shooting arrows, ...

b.

<i>synset</i>	
Lexemes	[English bowing, obeisance, bow Chinese 鞠躬, 弯腰, 运弓法 ¹⁰ French révérence]
Part-of-Speech	noun
Relations	[Hypernym <i>reverence, motion</i> ...]
Definition	bending the head or body or knee as a sign of reverence ...

- <http://wn-msa.sourceforge.net>
- Sumber terbuka
- Menggabungkan tiga wordnet
- Menandai lema yang digunakan dalam:
 - bahasa Malaysia (*zsm*)
 - bahasa Indonesia (*ind*)
 - keduanya (*msa*)
- Anotasi korpus NTU-MC

- Dr Francis Bond (Universitas Teknologi Nanyang, Singapura)
- David Moeljadi (Universitas Teknologi Nanyang, Singapura)
- Dr Hammam Riza (Badan Pengkajian dan Penerapan Teknologi, Indonesia)
- Dr Tang Enya Kong (Linton University College, Malaysia)
- Dr Lim Lian Tze (Malaysia)

KONEKSI DENGAN PRINCETON WORDNET DAN TIM WORDNET GLOBAL



Konferensi Wordnet Global ke-8 di Rumania

- Prof Christiane Fellbaum (Universitas Princeton)
- Prof Piek Vossen (VU Amsterdam)
- ...

BENGKEL WORDNET BAHASA PERTAMA DAN KEDUA



Universitas Teknologi Nanyang, Singapura, 26–27 Oktober 2014



Universitas Teknologi Nanyang, Singapura, 15–16 Januari 2016

BENKEL WORDNET BAHASA KETIGA



Badan Bahasa



Universitas Indonesia

Tahun 2017?

15032376-n **o** 'any substance that causes injury or illness or death of a living organism';

Search WN

English

English

English *poison₄ (n n n) , toxicant , poisonous substance*

Indonesian *racun , bisa*

Malaysian *racun*

[Edit Synset \(Experimental\)](#)

Definitions

English

any substance that causes injury or illness or death of a living organism

Relations

Hyponym: [atropine hyoscyamine toxin](#)

Hypernym: [substance](#)

Semantic Field: substance_n

Universal Links

SUMO: [c BiologicallyActiveSubstance](#)

TempoWN: ◀ ◁ ▷ (Past: 0.000; Present: 0.010; Future: 0.000)

SentiWN: ◁ ▼ (+0.00 -0.12) **MLSentiCon:** ◁ ▼ (+0.00 -0.38)

Search WN

Langs: English

English

[More detail about the Wordnet Bahasa \(1.0\)](#)

This project is now integrated in the [Extended Open Multilingual Wordnet \(1.0\)](#)

Maintainer: [Francis Bond](#) <bond@ieee.org>

01/07/2017 10:00:00

WORDNET BAHASA DALAM ANGKA (BOND ET AL., 2014)

Wordnet	Bhs	Synset	Kata	Makna
Wordnet B.Indonesia	ind	27 506	30 358	57 560
Wordnet B.Malaysia	zsm	23 953	23 833	48 996
Wordnet Bahasa	ind	19 316	19 522	48 111
	zsm	19 347	19 572	48 181
Gabungan	ind	48 689	58 541	133 005
	zsm	38 736	45 664	114 025

- Menambah data menggunakan:
 - Kamus Besar Bahasa Indonesia (Indonesia)
 - Penerjemahan judul artikel Wikipedia dan Wikidata API (Malaysia)
 - Data bahasa Melayu Brunei (Brunei?)

- Cakupan: beberapa konsep yang biasa/sering dipakai masih belum ada
- Sedikit kalimat definisi (masalah hak cipta dengan data kamus yang ada sekarang)
- Kekeliruan (terutama bahasa Malaysia vs bahasa Indonesia)
- Kesalahan ketik dan konsep yang tidak ada padanannya

Wordnet bahasa Abui

- Wordlist
- Korpus teks
- Rapid Words
- WordNet

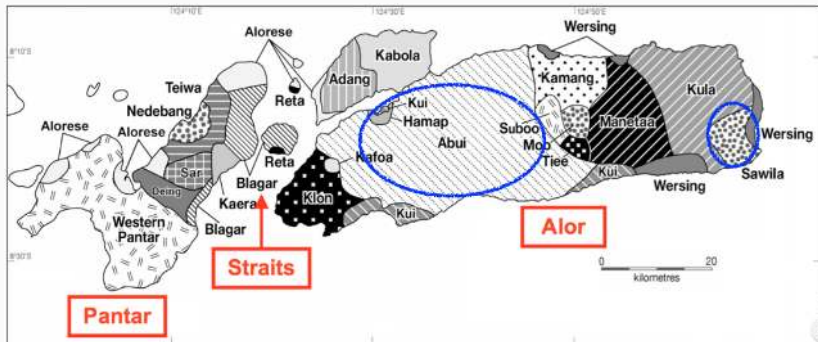
- Komunitas penutur bahasa
- Pendidikan bahasa untuk penutur asli
- Bagi linguis:
 - Linguistik komparatif (Austronesian Basic Vocabulary Database, ACD)
 - Tipologi
 - Kontak bahasa
 - Leksikografi (PanLex)

LOKASI GEOGRAFIS ALOR-PANTAR



Base 802958AI (C00429) 11-02

SITUASI LINGUISTIK DI KEPULAUAN ALOR-PANTAR



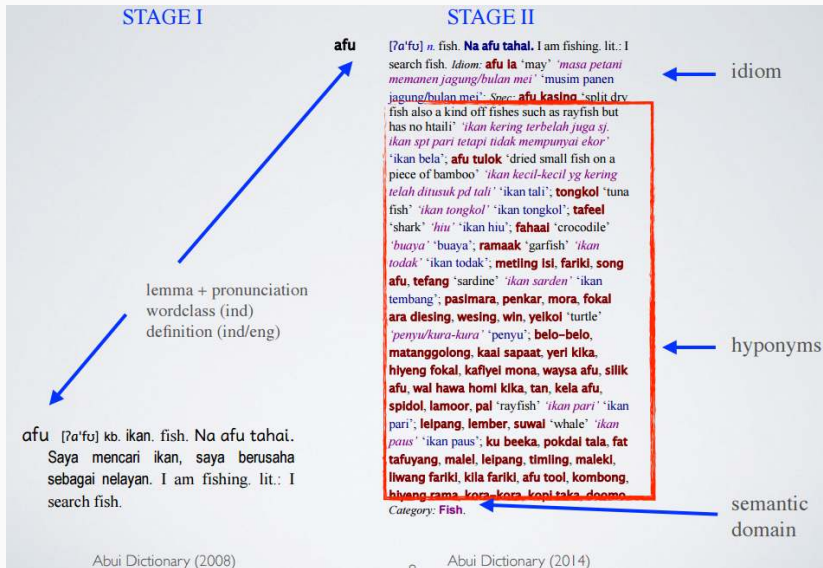
Jumlah penutur kira-kira 16.000 orang

PENELITIAN LINGUISTIK DI ALOR-PANTAR

- Dokumentasi bahasa di Alor-Pantar berjalan lambat dan hanya bergantung pada satu atau dua ahli bahasa
- Ada lebih dari 20 bahasa Alor-Pantar. Hanya 5 bahasa yang memiliki kamus kecil (Abui, Kamang, Sawila, Teiwa, dan Pantar Barat) yang disusun oleh 4 orang ahli bahasa. Selebihnya hanya ada daftar kosakata untuk 15 bahasa lainnya (daftar Holle, daftar Stokhof, daftar kosakata Pusat Bahasa).
- Semua kamus kecil tersebut hanya berisi tidak lebih dari 4.000 kata meskipun untuk menyusunnya masing-masing diperlukan beberapa tahun
- Disusun dengan metode pengumpulan kata yang berbeda-beda, bergantung pada masing-masing penyusun
- Hanya 4 yang pernah dicetak dan dibagikan di komunitas penutur
- Tidak ada sumber dalam bahasa Belanda

- František Kratochvíl: 2003-2008 (5 tahun, 3 tahun secara intensif)
- Benidiktus Delpada (1/2 tahun secara intensif)
- Dicitak dan dibagikan ke komunitas penutur pada tahun 2008
- Dukungan dana dari Universitas Leiden dan Kabupaten Alor
- Menggunakan *Multi-Dictionary Formatter* di SIL Toolbox:
 - Tahap 1: 2.700 lema, definisi dalam bahasa Indonesia dan Inggris, contoh kalimat, sinonim, paradigma infleksi, sketsa tata bahasa
 - Tahap 2: 3.000 lema, ditambah *Semantic Domain ID*
 - Tahap 3 (sedang dikerjakan): 15.000 lema, dihubungkan ke WordNet

PENYUSUNAN KAMUS BAHASA ABUI (TAHAP 1 DAN 2)



PENYUSUNAN KAMUS BAHASA ABUI (TAHAP 3)

lemma & pronunciation >	\lx pok \ph 'pok \ps v.0		
part of speech >	\pn kki \ge split \ve broken \ve smashed		
gloss, reversal, and definition (ENG) >	\ve split \ide split, burst, hatch, broken, smashed	←	used for mapping
gloss, reversal, and definition (IND) >	\gn pecah \vn retak \vn menetas \dn pecah, menetas, retak	←	used for mapping
gloss, reversal, and definition (AM) >	\gr pecah \vr menetas \dr pecah, menetas	←	used for mapping
example sentence & translations >	\vref Poku.001 \xv Pingai nu hayei poku. \xe A plate fell down and broke. \xn Piring itu jatuh dan pecah. \xr Piring tu jatu pecah. \ue \un \pd I (Perf) \vs 7.8		
semantic domain ID & name >	\sd Divide into pieces \sdn Membagi sesuatu menjadi banyak bagian \vs 7.8.2 \sd Crack \sdn Retak \nt \dt 03/Feb/2015	←	not yet used for mapping

Menghubungkan ke Wordnet

HASIL PENYUSUNAN KAMUS BAHASA ABUI (TAHAP 3)

full match: 10pt

neighbour: 1pt

triangulation score =
full + neighbour match

triangulation matrix
[1,1,1]

\lx pok
\ph 'pok
\ps v.0
\pn kki
\ge split
\re broken
\re smashed
\re split
\de split, burst, hatch, broken, smashed ← lookup in English WN
\gn pecah
\vn retak
\vn menetas
\vn pecah
\dn pecah, menetas, retak
\gr peca
\vr menetas
\vr peca
\dr peca, menetas
\ref Poku.001
\xv Pingai nu hayei poku.
\xe A plate fell down and broke.
\xn Piring itu jatuh dan pecah.
\xr Piring tu jatu peca.
\ue
\un
\pd 1 (Perf)
\vs 7.8
\sd Divide into pieces
\sdn Membagi sesuatu menjadi banyak bagian ← lookup in WN-SemDom linking
\vs 7.8.2
\sd Crack
\sdn Retak
\vt
\vt 03/Feb/2015

lookup in Bahasa WN

lookup in WN-SemDom linking
(Muhammad Zulhelmy et al. 2014)

HASIL PENCOCOKAN OTOMATIS

Triangulation type	Triangulation Score	Generated links	Sample	Match
Full (min 30; 1,1,1)	80~40	2368	100	83
Full (min 30; 1,1,1)	32~30	2368	100	73
ENG+IND (min 20; 1,1,0)	61~31	3032	100	61
ENG+IND (min 20; 1,1,0)	20	3032	100	34
ENG+ZSM (min 20; 1,0,1)	31-20	206	100	56
ENG+ZSM (min 20; 1,0,1)	20	206	100	45
IND+ZSM (min 20; 0,1,1)	63-31	4877	100	61
IND+ZSM (min 20; 0,1,1)	20	4877	100	42
ENG (min 10; 1,0,0)	32~20	9716	100	67
ENG (min 10; 1,0,0)	10	9716	100	7
IND (min 10; 0,1,0)	32~21	17380	100	55
IND (min 10; 0,1,0)	10	17380	100	11
ZSM (min 10; 0,0,1)	21~11	1582	100	44
ZSM (min 10; 0,0,1)	10	1582	100	21

Sekitar 39.000 tautan ke 3 wordnet dihasilkan secara otomatis

- Gabungan kata, idiom, peribahasa, metafora
- Penyusunan dan struktur kamus

PERMASALAHAN (1) - SALAH KETIK/TULIS

Alor Malay: tuju 'seven'

<i>yeting-ayoku</i> 'seven'	num	11467996-n	10	[0, 0, 1]	the striking of a light beam on a surface	incidence	tuju
--------------------------------	-----	------------	----	-----------	---	-----------	------

tuju I (Wik3237758) ṭṭṭṭ

1. Pointing at; making straight for; aiming at; = ([ETY: KEDAH MALAY.]) pērëmbang.

Also the name for a kind of witchcraft (described below).

Tujuⁱ and mēnuju: to aim at; to make for; to assail by sorcery.

Pēnuju: rows of stakes (supporting screens) for guiding fish into large traps (bélat Bétawi, bélat këmbang).

Sa-tu[^]ju: in unison; in harmony; with the same aims. Mēnuju tuding: to aim so as to hit the side.

Tiap-tiap përkataan itu mēnuju ia kapada tujuan-nya: every word he used went straight to its mark.

The form of sorcery described as tuju is one of those most feared by Malays. In essence it means "pointing out" your victim to a spirit of evil or setting the spirit of evil at him. It is regarded as potentially deadly; cf.: mēnuju orang sampai mati (aim spells at a man till he dies), [Lr: *Ht. Abd.* 283]; and adinda sudah këna (di-tuju) nyaris tiada mati (I was struck by a spell and nearly died), [Lr: *Mal. Dewa* 115.] In another passage ([Lr: *Mal. Dewa* 116]) the procedure is described; the wizard-king called up the evil spirit known as jin panah Ranjuna (Arjuna's arrow) by means of a magic formula (man-tëra), and then directed it at his victim, whom it struck and slew.

But protective magic could be used against this tuju, cf.

Bahasa Melayu Alor: tuju "tujuh"

PERMASALAHAN (2) - TERLALU BANYAK KATA DALAM WN BAHASA

Abui hayik 'carry on one's head'

10	lift or elevate	heft; heave up; heft up; heave	<i>beralun; mengangkat; membinjat; mengalun</i>
10	raise in rank or condition	raise; lift; elevate	<i>memperbaiki; meningkatkan; memulihkan keadaan; naikkan; membinjat; memiara; mencaduk; menyondang; meninggikan; menyingsinnng; mengemukakan; membesarkan; memeliharaakan; menggembirakan; menimbulkan; mengumpil; menaikkan; meroket; mengangkat; memelihara; menganjungkan; membubungkan</i>
10	form an arch or curve	curve; arc; arch	<i>melaur; mengangkat; melengkung; melentur; terangkat</i>
10	give a promotion to or assign to a higher position	upgrade; kick upstairs; raise; advance; elevate; promote	<i>memperbaiki; dinaikkan pangkat; meningkatkan; membangkit; menyorongan; naikkan; pendahuluan; memiara; menaikkan pangkat; maju; mencaduk; meninggikan; membesarkan; memeliharaakan; menaikkan; tingkat; meroket; bergerak ke hadapan; mara; mengangkat; naik; memelihara; berkembang maju; membubungkan</i>

Bahasa Abui: hayik “menjunjung”

PERMASALAHAN (3) - PERLUASAN MAKNA

Abui faala 'wooden tray on house posts or for carrying food'

faala	n	21	used to reproduce music by rotating while a phonograph needle tracks in the groove; sound recording consisting of a disk with a continuous groove	disc; disk; phonograph record; phonograph recording; record; platter
faala	n	21	a memory device consisting of a flat disk covered with a magnetic coating on which information is stored	disc; magnetic disc; magnetic disk; disk

Bahasa Abui: faala “nampan kayu”

- Semantic Domains (SD) dari SIL International
- Wordnet dan Rapid Words saling melengkapi
 - Rapid Words memiliki ranah semantik yang lebih luas (bdk. “masalah tenis” di Wordnet)
 - Wordnet memberikan hubungan semantik yang eksplisit

PENCOCOKAN KE RANAH SEMANTIK

No.	Domain	No.	Domain	No.	Domain
1	Universe, creation	4	Social behavior (cont)	7	Physical actions
1.1	Sky	4.5	Authority	7.1	Posture
1.2	World	4.6	Government	7.2	Move
1.3	Water	4.7	Law	7.3	Move something
1.4	Living things	4.8	Conflict	7.4	Have, be with
1.5	Plant	4.9	Religion	7.5	Arrange
1.6	Animal	5	Daily life	7.6	Hide
1.7	Nature, environment	5.1	Household equipment	7.7	Physical impact
2	Person	5.2	Food	7.8	Divide into pieces
2.1	Body	5.3	Clothing	7.9	Break, wear out
2.2	Body functions	5.4	Adornment	8	States
2.3	Sense, perceive	5.5	Fire	8.1	Quantity
2.4	Body condition	5.6	Cleaning	8.2	Big
2.5	Healthy	5.7	Sleep	8.3	Quality
2.6	Life	5.8	Manage a house	8.4	Time
3	Language and thought	5.9	Live, stay	8.5	Location
3.1	Soul, spirit	6	Work and occupation	8.6	Parts of things
3.2	Think	6.1	Work	9	Grammar
3.3	Want	6.2	Agriculture	9.1	General words
3.4	Emotion	6.3	Animal husbandry	9.2	Part of speech
3.5	Communication	6.4	Hunt and fish	9.3	Very
3.6	Teach	6.5	Working with buildings	9.4	Semantic constituents related to verbs
4	Social behavior	6.6	Occupation	9.5	Case
4.1	Relationships	6.7	Tool	9.6	Connected with, related
4.2	Social activity	6.8	Finance	9.7	Name
4.3	Behavior	6.9	Business organization		
4.4	Prosperity, trouble				

Figure 2: Top two levels of the Semantic Domains

- 1.6** **Binatang**
- 1.6.1 Jenis-jenis binatang
- 1.6.1.1 Binatang menyusui
- 1.6.1.1.1 Binatang Primata
- 1.6.1.1.2 Binatang pemakan daging
- 1.6.1.1.3 Binatang berkuku besar
- 1.6.1.1.4 Binatang pengerat
- 1.6.1.1.5 Binatang berkantong
-
- 1.6.1.2 Burung
- 1.6.1.3 Binatang melata / reptil
- 1.6.1.3.1 Ular
- 1.6.1.3.2 Kadal
-

1.6.1.5 Ikan

Gunakan bidang ini untuk ikan (keluarga Chordata, kelas Osteichthyes).

Kata-kata umum apa yang digunakan untuk menyebut binatang yang hidup di dalam air?

- *ikan, binatang air, binatang laut*

Apa saja jenis ikan air tawar?

- *mujair, gurame, lele*

Apa saja jenis ikan laut?

- *haring, salmon, kuda laut*

Kata-kata apa yang digunakan untuk menyebut binatang yang sangat kecil di dalam laut?

- *plankton*

Kata-kata apa yang menjelaskan keadaan lingkungan hidup ikan?

- *air tawar, air asin*

Kata-kata apa yang menjelaskan sifat/bentuk ikan-ikan tertentu, yang membedakan ikan tersebut dari ikan lainnya?

- *bertulang, bersirip, bersisik*

dengan sinonim, hiponim, tempat, ciri, sumber makanan

1.1 Langit

Gunakan bidang ini untuk kata-kata yang berhubungan dengan langit.

Kata-kata apa yang digunakan untuk menyebut langit?

- *langit, cakrawala*

Kata-kata apa yang digunakan untuk menyebut udara di sekitar bumi?

- *udara, angkasa*

Kata-kata apa yang digunakan untuk menyebut tempat atau wilayah di atas langit (di atas udara)?

- *surga, luar angkasa*

Kata-kata apa yang menjelaskan sifat sesuatu yang ada di langit atau sesuatu yang terjadi di langit?

- *di udara, di angkasa*

Kata-kata apa yang menjelaskan rupa langit, atau tampak seperti apa langit itu?

- *biru, cerah, berawan, bertaburan bintang*

Kata-kata apa yang digunakan untuk menyebut batas antara langit bagian bawah dengan permukaan bumi?

- *kaki langit, batas pandangan, tepi langit*

Kata-kata umum apa yang digunakan untuk menyebut sesuatu yang ada di langit?

- *benda langit, benda angkasa*

Kata-kata apa yang digunakan untuk menyebut cahaya yang kadang-kadang tampak di sebelah utara (atau selatan) langit? (CATATAN: Kemungkinan cahaya ini tidak terlihat di negara tropis seperti Indonesia.)

- *cahaya di utara, cahaya di selatan*

Kata-kata apa yang digunakan untuk menjelaskan bahwa sesuatu ada di langit?

- *di langit, di udara, di atas, di atas langit*

dengan sinonim, bagian, tempat, keadaan











CONTOH DATA YANG DIDAPAT

3

Nomor Bidang Makna: _____

Nama Bidang Makna: _____ Diisi oleh _____

Penutur Asli (huruf inisial) _____

Awal Tanggal: _____ Jam: _____

120

Kata atau istilah

Terjemahan dalam Bahasa Indonesia

- | | | |
|---------------|---|--|
| 2.6.A.2.1. 1. | <u>Hot hawel</u> | <u>Membicarakan anak lewat makan minum</u> |
| 2. | <u>Hamuna haring</u> | <u>Halih sayang</u> |
| 3. | <u>Flea hemama</u> | <u>Menjadi ter, lepa</u> |
| 4. | <u>Huang homenis, Tang dei hotitaka</u> | <u>Terap memperbaiki / memedalikan</u> |
| 5. | <u>Hada amagra</u> | <u>Mengajar ^{dan mendidik} sesuatu untuk dimengerti</u> |
| 6. | <u>Hing ris</u> | <u>Melatih suatu untuk dikhawatirkan</u> |
| 7. | <u>Kaparing</u> | <u>Menasihati untuk diingat</u> |
| 8. | <u>Hon hupang laka hahansa</u> | <u>Membeberatkan / memperbaiki kesalahan</u> |
| 9. | <u>Hoku mahli / mahi mahli</u> | <u>Pandai mengasuh anak</u> |
| 10. | <u>Hoi halual habul</u> | <u>Anak yang selalu di manjakan / memanjai</u> |
| 11. | <u>He hlio</u> | <u>Mengabarkan anak / tidak mengisahkan</u> |
| 12. | <u>He halering</u> | <u>Memperlihatkan anak dengan kasar</u> |
| 13. | <u>He harkida</u> | <u>Mengadopsi / mengasuh anak orang ^{lain} untuk</u> |
| 14. | <u>Hok miki</u> | <u>Mengajar / mengasuh / mengasahi anak</u> |
| 15. | <u>Hakan hami</u> | <u>Seorang yg ditunjuk untuk mengasuh</u> |
| 2.6.A.3. 16. | <u>Ueng abet</u> | <u>Remuda</u> |
| 17. | <u>Ueng abet mayal abak</u> | <u>Remuda - remudi / remaja</u> |
| 18. | <u>Lasing tamel</u> | <u>Super dan anak balok</u> |
| 19. | <u>Hakan huyal / tanyal kang</u> | <u>Usia laranku / melahirkan</u> |
| 20. | <u>Dara tung pida maha</u> | <u>Salah satu umur sebelum masuk remaja</u> |

JUMLAH KATA YANG DIDAPAT

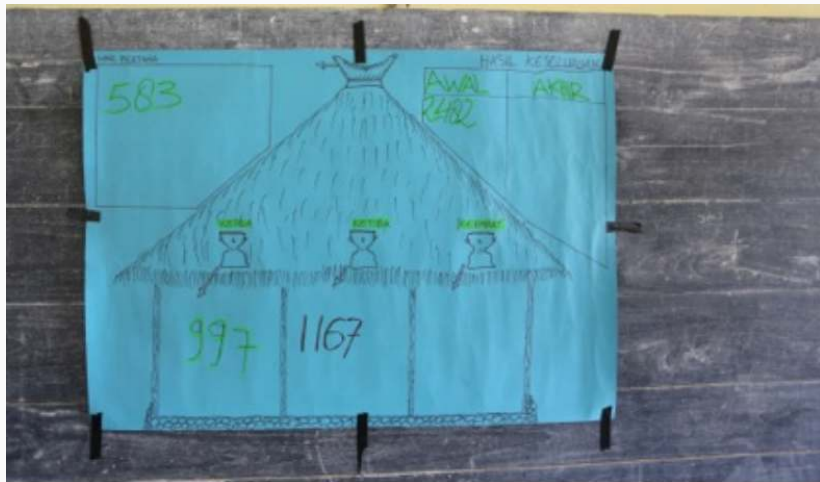




Foto proses perekaman tiap kata dengan contoh kalimat

- 15.000 kata diperoleh dalam 7 hari dengan bantuan lebih dari 20 narasumber
- Menambah dan memeriksa definisi dalam bahasa Indonesia
- Menambah definisi dalam bahasa Inggris > otomatis
- Menghubungkan ke WordNet > otomatis
- Menghubungkan ke sinonim dan antonim > otomatis

CONTOH DATA YANG SUDAH DIMASUKKAN DI KOMPUTER

Meibeka	Jelek	2.3.1.8.2	Ugly	Jelek
He on hepaneng meibeka	Perbuatan dan tingkah lakunya tidak bagus	2.3.1.8.2	Ugly	Jelek
Heng amakang	Yang penting manusia	2.3.1.8.2	Ugly	Jelek
Amakang pilasama	Manusia itu sama	2.3.1.8.2	Ugly	Jelek
Meibeka	Jelek, tidak baik	2.3.1.8.2	Ugly	Jelek
Tomi hatiknaha	Tidak dimengerti	2.3.1.8.2	Ugly	Jelek
Hapong aliking	Ganas, seram	2.3.1.8.2	Ugly	Jelek
Aliking naha	Tidak ganas	2.3.1.8.2	Ugly	Jelek
Kafering naha	Tidak ganas	2.3.1.8.2	Ugly	Jelek
Lakang meibeka	Sangat jelek	2.3.1.8.2	Ugly	Jelek
Tomi lahel diei	Merasa jengkel	2.3.1.8.2	Ugly	Jelek
Ama hapada	Manusia aneh	2.3.1.8.2	Ugly	Jelek
Meibeka	Jelek, buruk, tidak baik	2.3.1.8.2	Ugly	Jelek

PROSES MENGHUBUNGKAN RANAH SEMANTIK DAN WORDNET SECARA SEMI-OTOMATIS (MUHAMMAD ZULHELMI ET AL., 2014)

LVL	Example	# IDs	ID linked to WN		≥ 1 relation		≥ 2 relation		monosemous	
			eng	ind	eng	ind	eng	ind	eng	ind
1	1: universe	9	3	4	3	4	1	2	1	1
2	1.1: sky	68	54	46	27	27	6	13	7	7
3	1.1.1: sun	419	252	237	73	74	16	32	33	29
4	1.1.1.1: moon	985	702	605	90	69	8	8	86	65

PROSES MENGHUBUNGKAN RANAH SEMANTIK DAN WORDNET SECARA SEMI-OTOMATIS (MUHAMMAD ZULHELMI ET AL., 2014)

English			Indonesian		
8.4.1	15113229-n	period of time	8.4.1	15115926-n	jangka waktu
Hyponym			Hyponym		
8.4.1.2	14484516-n	day	8.4.1.2	14484516-n	hari
8.4.1.3	15135996-n	week	8.4.1.3	15135996-n	minggu
8.4.1.4	15206296-n	month	8.4.1.4	09358226-n	bulan
8.4.1.5	00294884-v	season	8.4.1.5	15239292-n	musim
8.4.1.6	15201505-n	year	8.4.1.6	15201505-n	tahun
8.4.1.7	15248564-n	era	8.4.1.7	15248564-n	zaman
not mapped			not mapped		
8.4.1.1	08266849-n; 06487395-n; 15173479-n	Calendar	8.4.1.1	15173479-n	Kalender
8.4.1.8	NIL	Special days	8.4.1.8	NIL	Hari-hari khusus

Table 4: Example of a good 2nd level mapping

PROSES MENGHUBUNGKAN RANAH SEMANTIK DAN WORDNET SECARA SEMI-OTOMATIS (MUHAMMAD ZULHELMI ET AL., 2014)

	English		Indonesian
8.3.3.3.4:	colors of the spectrum	8.3.3.3.4:	rentetan warna yang diuraikan oleh cahaya
3:	language and thought	3:	bahasa dan pikiran
9.4:	semantic constituents related to verbs	9.4:	konstituen atau unsur semantik yang berkaitan dengan
1.3.5:	solutions of water	1.3.5:	larutan air
2.3.1.9:	something used to see	2.3.1.9:	sesuatu yang digunakan untuk melihat

Table 2: **SD** main words not mapped to **WN**

- Keuntungan pemetaan: definisi dari PWN dan Wordnet Bahasa dapat diimpor
- Definisi tersebut dapat menjadi acuan untuk definisi dalam bahasa daerah

- Mengecek tautan ke Wordnet secara manual (dapat memperbaiki Wordnet Bahasa)
- Menulis definisi dalam bahasa Indonesia
- Memberi pelafalan dan contoh kalimat
- Merekam bentuk fonetis
- Menyediakan informasi tambahan tentang paradigma infleksi
- Merekam bentuk dialektal







Pemecahan struktur data KBBI Edisi Keempat

1. Excel dan Word

	A
1	A, a n 1 <u>huruf pertama abjad Indonesia</u> ; 2 <u>nama huruf a</u> ; 3 <u>penanda pertama dl urutan (mutu, nilai, dsb)</u>
2	à 1 <u>kira-kira; lebih kurang (antara dua angka untuk memperkirakan panjang, besar, dsb sesuatu): ular itu panjangnya 6 – 7 m; lama perjalanan 2 – 3 jam</u> ; 2 <u>harga tiap-tiap satuan: ia membeli bahan itu 5 m – Rp20.000,00</u>
3	a- <u>bentuk terikat</u> 1 <u>kekurangan: anemia</u> ; 2 <u>tidak atau bukan: aseksual</u> ; 3 <u>tanpa: anonim</u>
4	aa <u>Sd n akang</u>
5	¹ab <u>n wadah kecil dr timah untuk candu; hap</u>
6	²ab <u>ark n ayah</u>
7	ab- <u>bentuk terikat dari; jauh dr: abnormal</u>
8	aba <u>n ayah; bapak</u>

2. rtf

```
\trowd \trgaph30\trleft-30\trrh317\cellx1040\clmfg \cellx2351\clmrg \cellx18546\pard \intbl
\qc \f5\fs22 \cf55 1\cell \ql \f6\fs22 \b A\fs22 \b0 , \f6\fs22 \b a \f7\fs22 \i \b0 n\fs
\fs22 \i0 \f6\fs22 \b 1\fs22 \b0 huruf pertama abjad Indonesia; \f6\fs22 \b 2\fs22
\b0 nama huruf \f7\fs22 \i a\fs22 \i0 ; \f6\fs22 \b 3\fs22 \b0 penanda pertama dl
urutan (mutu, nilai, dsb) \cell \qr \f0\fs22 \cell
\pard \intbl \row\trowd \trgaph30\trleft-30\trrh317\cellx1040\clmfg \cellx2351\clmrg
\cellx18546\pard \intbl \qc \f5\fs22 2\cell \ql \f6\fs22 \b \u224\ 'e0\fs22 \b0 \f6\fs22
\b 1\fs22 \b0 kira-kira; lebih kurang (antara dua angka untuk memperkirakan panjang,
besar, dsb sesuatu): \f7\fs22 \i ular itu panjangnya 6 \u8212\ '97 7 m\fs22 \i0 ; \f7\fs22
\i lama perjalanan 2 \u8212\ '97 3 jam\fs22 \i0 ; \f6\fs22 \b 2\fs22 \b0 harga tiap-
tiap satuan: \f7\fs22 \i ia membeli bahan itu 5 m \u8212\ '97 Rp20.000,00\fs22 \i0 \cell
\qr \f0\fs22 \cell
```

3. html

```
<b>A, a</b><i>n</i><b>1</b> · huruf pertama abjad Indonesia; · <b>2</b> ·  
nama huruf <i>a</i>; · <b>3</b> · penanda pertama dl urutan (mutu,  
nilai, dsb)  
<b>à</b> · · <b>1</b> · kira-kira; lebih kurang (antara dua angka untuk  
memperkirakan panjang, besar, dsb sesuatu): <i>ular itu panjangnya 6  
\u8212\'97·7·m</i>; <i>lama perjalanan 2·\u8212\'97·3·jam</i>; · <b>2</b>  
<b> · harga tiap-tiap satuan: <i>ia membeli bahan itu 5·m·\u8212\'97·  
Rp20.000,00</i>  
<b>a- </b><i>bentuk terikat</i><b>1</b> · kekurangan: <i>anemia</i>; · <  
<b>2</b> · tidak atau bukan: <i>aseksual</i>; · <b>3</b> · tanpa: <i>anonim</i>  
<i>  
<b>aa</b><i>Sd</i><i>n</i> · akang  
<b>ab (1)</b><i>n</i> · wadah kecil dr timah untuk candu; hap  
<b>ab (2)</b><i>ark</i> · <i>n</i> · ayah  
<b>ab- </b><i>bentuk terikat</i> dari; jauh dr: <i>abnormal</i>  
<b>aba</b><i>n</i> · ayah; bapak
```

4. (Python)

```
for line in f.readlines():
    try:
        items = line.strip()
        #####
        ##### ENTRY WORDS #####
        #####
        # First table for entry words
        # search the entry words
        if bool(re.search(r'^<b>', items)):
            # extract the entry words
            lemma = re.findall(r'^<b>(.*?)</b>', items)
            word.append(lemma[0])
            master.append(lemma[0])
            ##### FOR WORD VARIANTS #####
            # change "alf(u)" to "alf" (variant 1) and "alfu" (variant 2)
            if bool(re.search(r'\s\(\D+\)', lemma[0])):
                lemma_var = [(re.findall(r'(.+)\(\D+\)', lemma[0])[0], re.findall(r'(.+)\s\(\D+\)', lemma[0])[0]+re.findall(r'.+\s\(\D+\)', lemma[0])[0])]
                allomorph.append(lemma_var[0][1])
                master.append(lemma_var[0][1])
                lemma_var_without_dots = re.sub(r'\.', '', lemma_var[0][1])
                outputLine = str(len(allomorph)) + '\t' + str(len(word)) + '\t' + ' ' + '\t'
                + lemma_var_without_dots
                outputVar.append(outputLine)
                masterLine = str(len(master)-1) + '\t' + lemma_var_without_dots + '\t' +
                'variant' + '\t' + str(len(allomorph))
                outputMaster.append(masterLine)
            # change "-anda (-nda, -da)" to "-anda" (variant 1) and "-nda, -da" (variant 2)
            elif bool(re.search(r'\s+\(\D+\)', lemma[0])):
```

5. txt

1	A	A			
2	à	à			
3	a-	a-			
4	aa	aa			
5	ab (1)	ab (1)	1	1	a
6	ab (2)	ab (2)	2	1051	alfu
7	ab-	ab-	3	1610	an-
8	aba	aba	4	1691	-nda
9	aba-aba	aba-aba	5	1691	-da

6. SQL



Pangkalan data KBBI IV terdiri dari 17 tabel yang saling terhubung:

1. Entri
2. Kata
3. Kata berimbuhan
4. Gabungan kata
5. Kata kiasan
6. Peribahasa
7. Varian
8. Rujuk
9. Makna
10. Contoh
11. Makna contoh
12. Nama ilmiah
13. Rumus kimia
14. Label ragam
15. Label kelas kata
16. Label bahasa
17. Label bidang ilmu

JUMLAH DATA YANG BERHASIL DIAMBIL DARI KBBI IV

- 92.011 lema, terdiri dari:
 - 41.472 kata dasar
 - 24.607 kata berimbuhan
 - 23.536 gabungan kata
 - 2.033 peribahasa
 - 272 idiom/kata kiasan
 - 91 varian
- 3.473 rujuk silang
- 109.005 makna
- 27.889 contoh
- 789 makna contoh
- 2.835 nama ilmiah
- 136 rumus kimia

- dari kata dasar
 - mencari semua entri dan subentri dari kata 'kacang' dengan maknanya
- secara ortografi:
 - mencari kata 'mereka' yang dapat berasal dari kata dasar yang berbeda
- per kategori:
 - mencari peribahasa dan idiom
 - mencari kata yang label bahasanya *Jw* dan ragamnya *kl*

TAHAP SELANJUTNYA (1)

- Memperbaiki pangkalan data KBBI IV
 - Memperbaiki struktur data: *likur*, ...
 - Mengganti singkatan dalam makna ke bentuk panjangnya
 - Memperbaiki penjelasan makna kata, terutama yang hanya berupa nama ilmiah: *bangan*, *tebal bibir*, ...
 - Mengecek secara manual subentri kata kiasan dan peribahasa yang bertanda kurung: *diambak*, *anak*, ...
 - Jumlah lema akan bertambah
 - Mengecek kesalahan tulis pada makna secara semi-otomatis: *gedung*, *tambahan pula*, *tonggos*, ...
 - Menambah kata-kata baru untuk KBBI V
- Memanfaatkan data KBBI IV untuk penelitian
 - kata-kata yang banyak digunakan sebagai *genus*
 - Penggolongan kata-kata berafiks “meN-, ber-” dsb.
 - Menambah leksikon untuk INDRA
- Mendokumentasikan proses pembuatan pangkalan data KBBI IV
 - Menulis makalah di jurnal internasional

TAHAP SELANJUTNYA (2)

- Menghubungkan data KBBI IV ke pranala luar
 - ke Open Multilingual Wordnet
 - Memperkaya penjelasan makna melalui Wordnet dan sebaliknya
 - ke Wikidata
- Membuat versi online KBBI IV
 - Laman web portal bahasa
 - Aplikasi ponsel
- Membuat antarmuka pengguna untuk menambah, menyunting, dan menghapus data KBBI
- Menambah data singkatan dan akronim, istilah-istilah dalam bahasa daerah dan bahasa asing
- Membuat aplikasi permainan dan pendidikan:
 - TTS, ...

- Resmi:
`kbbi4.portalbahasa.com/`
- (untuk bengkel 28 Juli 2016):
`compling.hss.ntu.edu.sg/omw/cgi-bin/kbbi.cgi`

- Mohon Bapak/Ibu membentuk kelompok semeja (3-4 orang)
- Paling banyak 24 kelompok

Bengkel

1. Menggunakan Wordnet dan KBBI IV Daring
2. Menghubungkan makna kata di KBBI IV dengan *synset* di Wordnet
 - setara dengan/sinonim (=)
 - tidak ada hubungan (X)
 - hipernim (>)
 - hiponim (<)
 - kurang lebih sama ()
 - berlawanan/antonim (><)
3. Memperbaiki KBBI IV dan Wordnet
 - Membandingkan penjelasan makna di KBBI IV dan Wordnet: menambah/menyunting/menghapus makna
 - Membandingkan lema di KBBI IV dan Wordnet: menambah/menyunting/menghapus lema, mengajukan penambahan lema daerah?
 - Contoh dan label dalam KBBI IV

1. Akses laman bengkel Wordnet Bahasa (bisa menggunakan ponsel):
`compling.hss.ntu.edu.sg/events/2016-ws-kbbi/`
2. Cek kelompok masing-masing
3. Buka Google Forms, Wordnet, dan KBBI IV Daring
4. Studi kasus:
 - 4.1 Grup 2 No. 4 - **kedai**
 - 4.2 Grup 5 No. 3 - **setempat**
 - 4.3 Grup 21 No. 2 - **menjual**

Ucapan terima kasih

- Terima kasih kepada Badan Bahasa (Ibu Dora Amalia) atas kesempatan yang diberikan untuk bengkel Wordnet Bahasa
- Terima kasih kepada Lim Lian Tze untuk slides presentasi tentang Wordnet
- Terima kasih kepada Francis Bond atas dukungan dan bantuannya untuk seminar dan bengkel Wordnet
- Terima kasih kepada Frantisek Kratochvíl untuk slides presentasi tentang Wordnet Abui
- Terima kasih kepada Hannah Choi atas bantuannya membuat Google Forms untuk bengkel
- Terima kasih kepada Luís Morgado da Costa atas bantuannya mengunggah antarmuka pengguna KBBI IV untuk bengkel
- Terima kasih kepada Ian Kamajaya dan Randy Sugianto atas bantuannya dalam pembuatan pangkalan data KBBI IV

Bond, F., Lim, L. T., Tang, E. K. and Riza, H. (2014). The combined Wordnet Bahasa. NUSA: Linguistic studies of languages in and around Indonesia, 57, 83–100. Retrieved from <http://hdl.handle.net/10108/79286>

Bond, F. and Paik, K. (2012). A survey of wordnets and their licenses. In Proceedings of the 6th Global WordNet Conference (GWC 2012) (pp. 64–71). Matsue, Japan.

Kratochvíl, F. and Delpada, B., Abui - Indonesian - English Dictionary. UBB Kupang. 2008.

Lim, L. T. and Hussein, N. (2006). Fast prototyping of a Malay WordNet system. In Proceedings of the Language, Artificial Intelligence and Computer Science for Natural Language Processing (LAICS-NLP) Summer School Workshop (pp. 13–16). Bangkok, Thailand.

Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38(11), 39–41.

Muhammad Zulhelmy bin Mohd Rosman, Bond, F., and Kratochvíl, F., “Bringing together over- and under-represented languages: Linking Wordnet to the SIL Semantic Domains,” In Orav, H., Fellbaum, C., and Vossen, P. (eds.), *Proceedings of the Seventh Global WordNet Conference*, pp. 40-48. Tartu, Estonia, January 25-29, 2014.
<http://gwc2014.ut.ee/>

Nurhil Hirfana, M. N., Suerya, S. and Bond, F. (2011). Creating the Open Wordnet Bahasa. In *Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation (PACLIC 25)* (pp. 255–264). Singapore.

Riza, H., Budiono and Hakim, C. (2010). Collaborative work on Indonesian Wordnet through Asian Wordnet (AWN). In Proceedings of the 8th Workshop on Asian language resources (pp. 9–13). Beijing.

Tufiş, D., Cristea, D. and Stamou, S. (2004). BalkaNet: aims, methods, results and perspectives – a general overview. Romanian Journal of Information Science and Technology Special Issue, 7(1), 9–43.

Vossen, P. (2004). EuroWordNet: a multilingual database of autonomous and language-specific wordnets connected via an Inter-Lingual-Index. Special Issue on Multilingual Databases, International Journal of Linguistics, 17(2).