

# Wordnet

Jejaring kata / Pangkalan data leksikal

---

David Moeljadi

23 Agustus 2016

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

MUSYAWARAH SEKRETARIAT KE-22 MAJELIS BAHASA BRUNEI DARUSSALAM-INDONESIA-MALAYSIA (MABBIM),  
MILLENNIUM HOTEL SIRIH, JAKARTA PUSAT, INDONESIA

# 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 di Univ. Bina Nusantara, 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. di Tokyo

**2013 - 2014** Lembaga Penelitian Bahasa dan Budaya Asia dan Afrika, Universitas Kajian Asing Tokyo

[compling.hss.ntu.edu.sg/who/david/](http://compling.hss.ntu.edu.sg/who/david/)

# Proyek saat ini

1. Indonesian Resource Grammar (INDRA)  
[github.com/davidmoeljadi/INDRA](https://github.com/davidmoeljadi/INDRA)  
[chimpanzee.ling.washington.edu/demophon/indra/](http://chimpanzee.ling.washington.edu/demophon/indra/)  
[delph-in.github.io/delphin-viz/demo/](https://delph-in.github.io/delphin-viz/demo/)
2. Wordnet Bahasa
3. NTU Multilingual Corpus  
[compling.hss.ntu.edu.sg/ntumc/](http://compling.hss.ntu.edu.sg/ntumc/)
4. Indonesian Loanword Search Engine  
[david.blogid.me/loanword\\_searchengine](http://david.blogid.me/loanword_searchengine)
5. Kamus Pemelajar Kanji Jepang-Indonesia
6. KBBI IV
7. Analisis sentimen teks bahasa Indonesia

1. Wordnet: Apa itu?
2. Wordnet: Apa saja kegunaannya?
3. Berbagai macam Wordnet
4. Wordnet Bahasa

– Tanya Jawab –

Wordnet: Apa itu?

---

- Laboratorium Ilmu **Kognitif** Universitas Princeton

- Princeton WordNet (PWN):

`wordnetweb.princeton.edu/perl/webwn`

- 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 (analisis teks), aplikasi kecerdasan buatan



# SYNSET “SYNONYM SET” / HIMPUNAN SINONIM

- Satuan dasar dalam Wordnet
- Tiap *synset* mewakili sebuah konsep (bdk. “buku” dan “kitab” di KBBI IV)
- Makna kata diwakili oleh:
  - sinonim
  - kalimat definisi
  - hubungannya dengan *synset* lainnya
- Anggota *synset* adalah kata-kata berkonsep sama
- Wordnet = Kamus + Tesaurus
- 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:

<nomina.tubuh> **kepala**, hulu – (bagian tubuh yang di atas leher (pada manusia dan beberapa jenis hewan merupakan tempat otak, pusat jaringan saraf, dan beberapa pusat indra))

<nomina.persona> **kepala**, ketua, pemimpin, penghulu – (orang yang mengepalai atau memimpin (rapat, dewan, perkumpulan, dsb); “**kepala** pasukan”)

<nomina.kognisi> **kepala**, kepintaran, kecerdikan, akal – (kemampuan mental; “jika kita masih punya **kepala**, kita harus diberi jatah satu kilogram gula”)

...

# HUBUNGAN ANTA-SYNSET NOMINA

**hipernim** istana  $\xrightarrow{\text{adalah}}$  tempat tinggal, kediaman, rumah

**holonim** jari  $\xrightarrow{\text{bagian-dari}}$  tangan

tepung  $\xrightarrow{\text{bahan-untuk}}$  roti, adonan, kue

juri  $\xrightarrow{\text{anggota-dari}}$  mahkamah, pengadilan, tribunal

**contoh** Mozart, Wolfgang Amadeus Mozart  $\xrightarrow{\text{contoh}}$  komponis, komposer, ...

juga ragam bahasa, bidang ilmu (domain) dsb.

(bdk. label-label di kamus)

Wordnet: Apa saja kegunaannya?

---

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

Mengelompokkan kata-kata terjemahan alternatif

Contoh:

`translate.google.co.id/#en/id/head`

# KEUNTUNGAN MENGGUNAKAN WORDNET

- Untuk pengguna awam:
  - Gratis (Daring)(+ Luring)
  - Multibahasa
- Untuk peneliti bahasa dan pekamus:
  - Jaringan semantik leksikal dengan informasi semantik, sintaks, definisi, contoh, domain, tautan derivasi kata
  - Pembuatan kamus (cetak) multilingual dengan  $\text{\LaTeX}$   
[www.overleaf.com/latex/examples/dictionary-template/pdztbwjxrpmz#.V7a16nV96kA](http://www.overleaf.com/latex/examples/dictionary-template/pdztbwjxrpmz#.V7a16nV96kA)
- Untuk peneliti linguistik komputasional:
  - Pemrosesan bahasa alami (NLP), analisis teks, disambiguasi makna, kecerdasan buatan, misalnya melalui NLTK (Natural Language Toolkit/Perangkat Bahasa Alami) untuk Python:  
[www.nltk.org](http://www.nltk.org)

## Berbagai macam Wordnet

---

- Banyak Wordnet untuk berbagai bahasa telah dikembangkan!
- 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)

- [compling.hss.ntu.edu.sg/omw/](http://compling.hss.ntu.edu.sg/omw/)
- Kumpulan berbagai Wordnet dengan lisensi sumber terbuka

Wordnet	Lang	Synsets	Words	Senses	Core	Licence	Data	Citation
<a href="#">Albanet</a>	<a href="#">als</a>	4,676	5,990	9,602	31%	<a href="#">CC BY 3.0</a>	<a href="#">als.zip (+xml)</a>	<a href="#">cite:als; (.bib)</a>
<a href="#">Arabic WordNet (AWN)</a>	<a href="#">arb</a>	10,165	14,595	21,751	48%	<a href="#">CC BY SA 3.0</a>	<a href="#">arb.zip (+xml)</a>	<a href="#">cite:arb; (.bib)</a>
<a href="#">BulTreeBank Wordnet (BTB-WN)</a>	<a href="#">bul</a>	4,999	6,783	9,056	100%	<a href="#">CC BY 3.0</a>	<a href="#">bul.zip (+xml)</a>	<a href="#">cite:bul; (.bib)</a>
<a href="#">Chinese Open Wordnet</a>	<a href="#">cmn</a>	42,312	61,533	79,809	100%	<a href="#">wordnet</a>	<a href="#">cmn.zip (+xml)</a>	<a href="#">cite:cmn; (.bib)</a>
<a href="#">Chinese Wordnet (Taiwan)</a>	<a href="#">qcn</a>	4,913	3,206	8,069	28%	<a href="#">wordnet</a>	<a href="#">qcn.zip (+xml)</a>	<a href="#">cite:qcn; (.bib)</a>
<a href="#">DanNet</a>	<a href="#">dan</a>	4,476	4,468	5,859	81%	<a href="#">wordnet</a>	<a href="#">dan.zip (+xml)</a>	<a href="#">cite:dan; (.bib)</a>
<a href="#">Greek Wordnet</a>	<a href="#">ell</a>	18,049	18,227	24,106	57%	<a href="#">Apache 2.0</a>	<a href="#">ell.zip (+xml)</a>	<a href="#">cite:ell; (.bib)</a>
<a href="#">Princeton WordNet</a>	<a href="#">eng</a>	117,659	148,730	206,978	100%	<a href="#">wordnet</a>	<a href="#">eng.zip (+xml)</a>	<a href="#">cite:eng; (.bib)</a>
<a href="#">Persian Wordnet</a>	<a href="#">fas</a>	17,759	17,560	30,461	41%	<a href="#">Free to use</a>	<a href="#">fas.zip (+xml)</a>	<a href="#">cite:fas; (.bib)</a>
<a href="#">FinnWordNet</a>	<a href="#">fin</a>	116,763	129,839	189,227	100%	<a href="#">CC BY 3.0</a>	<a href="#">fin.zip (+xml)</a>	<a href="#">cite:fin; (.bib)</a>
<a href="#">WOLF (Wordnet Libre du Français)</a>	<a href="#">fra</a>	59,091	55,373	102,671	92%	<a href="#">CeCILL-C</a>	<a href="#">fra.zip (+xml)</a>	<a href="#">cite:fra; (.bib)</a>
<a href="#">Hebrew Wordnet</a>	<a href="#">heb</a>	5,448	5,325	6,872	27%	<a href="#">wordnet</a>	<a href="#">heb.zip (+xml)</a>	<a href="#">cite:heb; (.bib)</a>
<a href="#">MultiWordNet</a>	<a href="#">ita</a>	34,728	40,343	61,558	83%	<a href="#">CC BY 3.0</a>	<a href="#">ita.zip (+xml)</a>	<a href="#">cite:ita; (.bib)</a>
<a href="#">Japanese Wordnet</a>	<a href="#">jpn</a>	57,184	91,964	158,069	95%	<a href="#">wordnet</a>	<a href="#">jpn.zip (+xml)</a>	<a href="#">cite:jpn; (.bib)</a>
<a href="#">Multilingual Central Repository</a>	<a href="#">cat</a>	45,826	46,531	70,622	81%	<a href="#">CC BY 3.0</a>	<a href="#">cat.zip (+xml)</a>	<a href="#">cite:cat; (.bib)</a>
<a href="#">Multilingual Central Repository</a>	<a href="#">eus</a>	29,413	26,240	48,934	71%	<a href="#">CC BY 3.0</a>	<a href="#">eus.zip (+xml)</a>	<a href="#">cite:eus; (.bib)</a>



# Wordnet Bahasa

---

- `wn-msa.sourceforge.net`
- Sumber terbuka
- Menggabungkan tiga wordnet:
  - Wordnet bahasa Melayu (Lim & Hussein, 2006)
  - Wordnet bahasa Indonesia (Riza, Budiono & Hakim, 2010)
  - Wordnet Bahasa Terbuka (Nurril Hirfana, Suerya & Bond, 2011)
- 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 ([gwc2016.racai.ro/](http://gwc2016.racai.ro/))

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

# BENKEL WORDNET BAHASA PERTAMA DAN KEDUA



Universitas Teknologi Nanyang, Singapura, 26–27 Oktober 2014  
[compling.hss.ntu.edu.sg/events/2014-ws-wn-bahasa](http://compling.hss.ntu.edu.sg/events/2014-ws-wn-bahasa)



Universitas Teknologi Nanyang, Singapura, 15–16 Januari 2016  
[compling.hss.ntu.edu.sg/events/2016-ws-wn-bahasa/](http://compling.hss.ntu.edu.sg/events/2016-ws-wn-bahasa/)

# BENKEL WORDNET BAHASA KETIGA



Badan Bahasa



Universitas Indonesia

Tahun 2017?



Malaysia



Brunei Darussalam

Tahun 2018?

## 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 dan memperbaiki data menggunakan:
  - Kamus Besar Bahasa Indonesia (Indonesia)
  - Penerjemahan judul artikel Wikipedia dan Wikidata (Malaysia)  
*Flora dan fauna*
  - Data bahasa Melayu Brunei (Brunei Darussalam?)

- Beberapa konsep yang biasa/sering dipakai masih belum ada
- Sedikit kalimat definisi
- Kekeliruan (kata-kata dalam bahasa Malaysia dan bahasa Indonesia)
- Kesalahan ketik dan konsep yang tidak ada padanannya

“Dengan Wordnet, mari kita bersama memajukan leksikografi dan penelitian bahasa di Brunei Darussalam, Indonesia, dan Malaysia”

Pertanyaan?

# Ucapan terima kasih

- Terima kasih kepada Badan Bahasa (Ibu Dora Amalia) atas kesempatan yang diberikan di MABBIM
- Terima kasih kepada Lim Lian Tze untuk slides presentasi tentang Wordnet
- Terima kasih kepada Francis Bond atas dukungan dan bantuannya untuk Wordnet

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.

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.

Nurril 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).