Building an online Indonesian dictionary from Word and Excel files

David **Moeljadi**

Division of Linguistics and Multilingual Studies,
Nanyang Technological University, Singapore

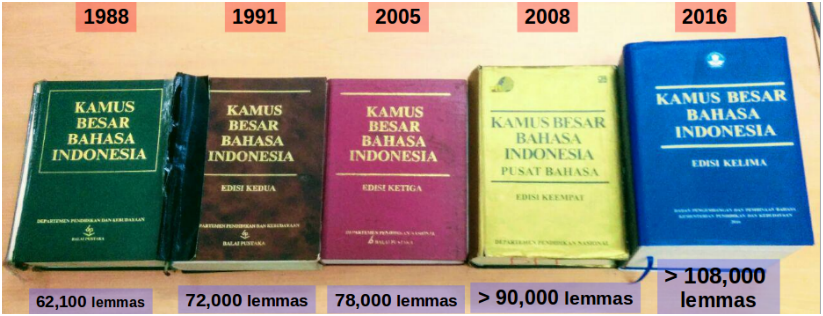
NIE-ELL Postgraduate Conference (PGC),
National Institute of Education (NIE), Singapore

20 April 2017

Outline

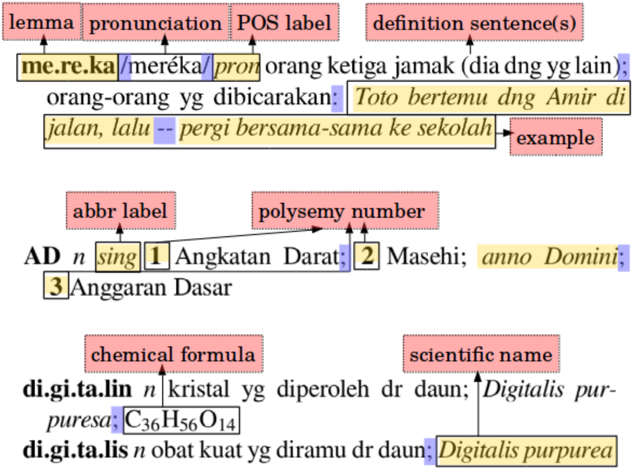
1. Kamus Besar Bahasa Indonesia (KBBI)
2. From Word and Excel to Database
3. Features in the Online KBBI V

Kamus Besar Bahasa Indonesia (KBBI)


1988
1991
2005
2008
2016
KAMUS BESAR BAHASA INDONESIA
KAMUS BESAR BAHASA INDONESIA EDISI KEDUA
KAMUS BESAR BAHASA INDONESIA EDISI KETIGA
KAMUS BESAR BAHASA INDONESIA PUSAT BAHASA EDISI KEEMPAT
KAMUS BESAR BAHASA INDONESIA EDISI KELIMA
62,100 lemmas
72,000 lemmas
78,000 lemmas
> 90,000 lemmas
> 108,000 lemmas

- the official dictionary of the Indonesian language
- published by *Badan Pengembangan dan Pembinaan Bahasa* (The Language Development and Cultivation Agency) or *Badan Bahasa* under Ministry of Education and Culture, Republic of Indonesia
- KBBI Fourth Edition (KBBI IV) [5] had its data in Microsoft Excel and Word files

Dictionary entries in KBBI


lemma
pronunciation
POS label
definition sentence(s)
example
abbr label
polysemy number
chemical formula
scientific name

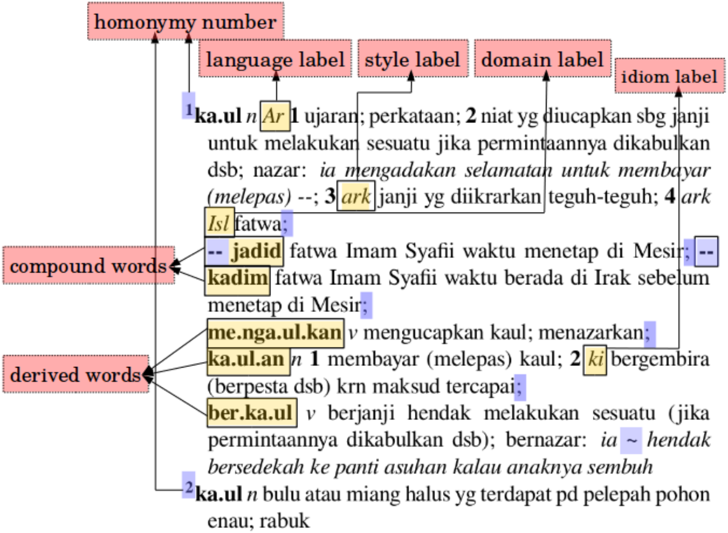
me.re.ka /meréka/ *pron* orang ketiga jamak (dia dng yg lain); orang-orang yg dibicarakan: *Toto bertemu dng Amir di jalan, lalu -- pergi bersama-sama ke sekolah*

AD *n sing* **1** Angkatan Darat; **2** Masehi; *anno Domini*; **3** Anggaran Dasar

di.gi.ta.lin *n* kristal yg diperoleh dr daun; *Digitalis purpuresa*; $C_{36}H_{56}O_{14}$

di.gi.ta.lis *n* obat kuat yg diramu dr daun; *Digitalis purpurea*

Dictionary entries in KBBI

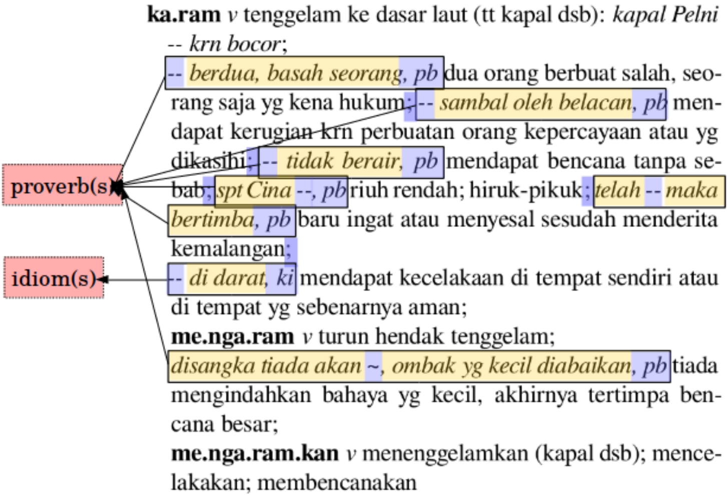


homonymy number
language label
style label
domain label
idiom label
compound words
derived words

^{1}**ka.ul** *n* *Ar* **1** ujaran; perkataan; **2** niat yg diucapkan sbg janji untuk melakukan sesuatu jika permintaannya dikabulkan dsb; nazar: *ia mengadakan selamatan untuk membayar (melepas)* --; **3** *ark* janji yg diikrarkan teguh-teguh; **4** *ark* *Isl* fatwa;
-- jadid fatwa Imam Syafii waktu menetap di Mesir; **-- kadim** fatwa Imam Syafii waktu berada di Irak sebelum menetap di Mesir;
me.nga.ul.kan *v* mengucapkan kaul; menazarkan;
ka.ul.an *n* **1** membayar (melepas) kaul; **2** *ki* bergembira (berpesta dsb) krn maksud tercapai;
ber.ka.ul *v* berjanji hendak melakukan sesuatu (jika permintaannya dikabulkan dsb); bernazar: *ia* ~ *hendak bersedekah ke panti asuhan kalau anaknya sembuh*

^{2}**ka.ul** *n* bulu atau miang halus yg terdapat pd pelepah pohon enau; rabuk

Dictionary entries in KBBI


proverb(s)
idiom(s)

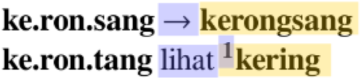
ka.ram *v* tenggelam ke dasar laut (tt kapal dsb): *kapal Pelni -- krn bocor*; *-- berdua, basah seorang, pb* dua orang berbuat salah, seorang saja yg kena hukum; *-- sambal oleh belacan, pb* mendapat kerugian krn perbuatan orang kepercayaan atau yg dikasihi; *-- tidak berair, pb* mendapat bencana tanpa sebab; *spt Cina --, pb* riuh rendah; hiruk-pikuk; *telah -- maka bertimba, pb* baru ingat atau menyesal sesudah menderita kemalangan; *-- di darat, ki* mendapat kecelakaan di tempat sendiri atau di tempat yg sebenarnya aman;

me.nga.ram *v* turun hendak tenggelam; *disangka tiada akan ~, ombak yg kecil diabaikan, pb* tiada mengindahkan bahaya yg kecil, akhirnya tertimpa bencana besar;

me.nga.ram.kan *v* menenggelamkan (kapal dsb); mencelakakan; membencanakan

Dictionary entries in KBBI

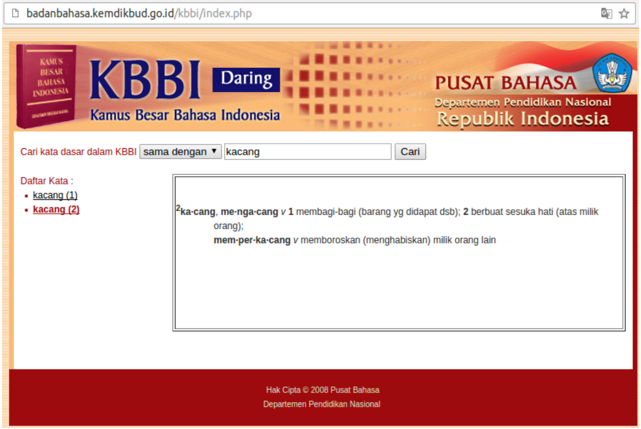
Cross-references



ke.ron.sang → **kerongsang**

ke.ron.tang lihat ^{1}**kering**

The Online KBBI before October 2016


badanbahasa.kemdikbud.go.id/kbbi/index.php
KAMUS BESAR BAHASA INDONESIA
KBBI
Daring
Kamus Besar Bahasa Indonesia
PUSAT BAHASA
Departemen Pendidikan Nasional
Republik Indonesia
Cari

Cari kata dasar dalam KBBI: sama dengan ▾ kacang

Daftar Kata :

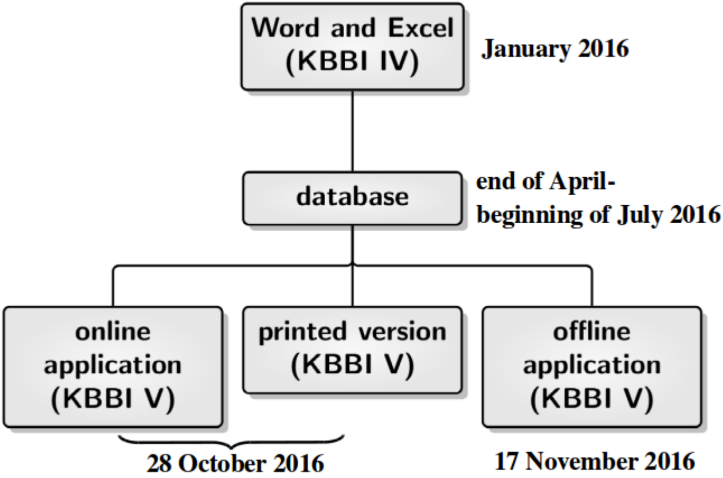
- kacang (1)
- kacang (2)

[2]**ka·cang**, **me·nga·cang** *v* **1** membagi-bagi (barang yg didapat dsb); **2** berbuat sesuka hati (atas milik orang);
mem·per·ka·cang *v* memboroskan (menghabiskan) milik orang lain

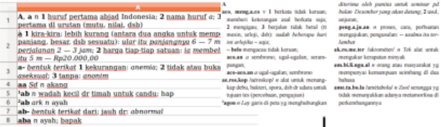
Hak Cipta © 2008 Pusat Bahasa
Departemen Pendidikan Nasional

- data from KBBI III, for simple word search by root (*kata dasar*)
- the result is exactly in the same format as the one in the printed dictionary
- the data was not structured, no database

From KBBI IV to KBBI V



From KBBI IV to KBBI V



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

A

aco, meng.a.co *v* **1** berkata tidak keruan; memberi keterangan asal berkata saja; **2** mengigau; **3** berjalan tidak betul (tt mesin, arloji, dsb): *sudah beberapa hari ini arlojiku – saja;*

– belo mengacau tidak keruan;

aco.an *a* sembrono; ugal-ugalan; serampangan;

aco-aco.an *a* ugal-ugalan; sembrono

ae.ros.kop /aéroskop/ *n* alat untuk menangkap debu, bakteri, spora, dsb dr udara untuk tujuan tes (percobaan, pengujian)

²agon *n Lay* garis di peta yg menghubungkan

diterima oleh panitia untuk seminar pd bulan Desember yang akan datang; **2** usul; anjuran;

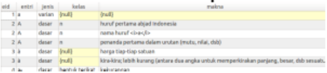
peng.a.ju.an *n* proses, cara, perbuatan mengajukan; pengusulan: *– usulnya itu terlambat*

ak.ro.me.ter /akrométer/ *n Tek* alat untuk mengukur kerapatan minyak

am.bi.li.ngu.al *n* orang atau masyarakat yg mempunyai kemampuan seimbang dl dua bahasa

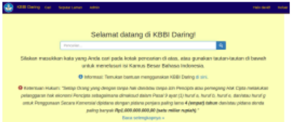
ame.ta.bo.la /amétabola/ *n Zool* serangga yg tidak menunjukkan adanya metamorfosa di perkembangannya

January 2016



eid	entri	jenis	kelas	makna
1	a	varian	(null)	(null)
2	A	dasar	n	huruf pertama abjad Indonesia
2	A	dasar	n	nama huruf <i>a</i>
2	A	dasar	n	penanda pertama dalam urutan (mutu, nilai, dsb)
3	à	dasar	(null)	harga tiap-tiap satuan
3	à	dasar	(null)	kira-kira; lebih kurang (antara dua angka untuk memperkirakan panjang, besar, dsb sesuatu...
4	a-	dasar	bentuk terikat	kekurangan

end of April-beginning of July 2016



Selamat datang di KBBI Daring!

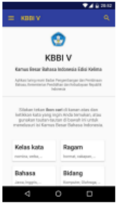
Silakan masukkan kata yang Anda cari pada kotak pencarian di atas, atau gunakan tautan-tautan di bawah untuk menelusuri isi Kamus Besar Bahasa Indonesia.

Informasi: Temukan bantuan menggunakan KBBI Daring di sini.

Ketentuan Hukum: "*Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf a, huruf b, huruf e, dan/atau huruf g untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama* **4 (empat) tahun** *dan/atau pidana denda paling banyak* **Rp1.000.000.000,00 (satu miliar rupiah)**."

Baca selengkapnya »

28 October 2016



KBBI V

Kamus Besar Bahasa Indonesia Edisi Kelima

Aplikasi luring resmi Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan dan Kebudayaan Republik Indonesia

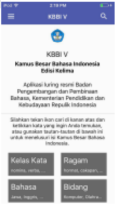
Silakan tekan **ikon cari** di kanan atas dan ketikkan kata yang ingin Anda temukan, atau gunakan tautan-tautan di bawah ini untuk menelusuri isi Kamus Besar Bahasa Indonesia.

Kelas kata
nomina, verba, ...

Ragam
hormat, cakapan, ...

Bahasa
Jawa, Inggris, ...

Bidang
Komputer, Olahraga, ...



KBBI V

Kamus Besar Bahasa Indonesia Edisi Kelima

Aplikasi luring resmi Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan dan Kebudayaan Republik Indonesia

Silahkan tekan ikon cari di kanan atas dan ketikkan kata yang ingin Anda temukan, atau gunakan tautan-tautan di bawah ini untuk menelusuri isi Kamus Besar Bahasa Indonesia.

Kelas Kata
nomina, verba, ...

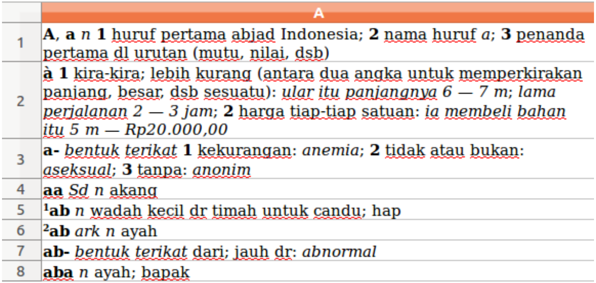
Ragam
hormat, cakapan, ...

Bahasa
Jawa, Inggris, ...

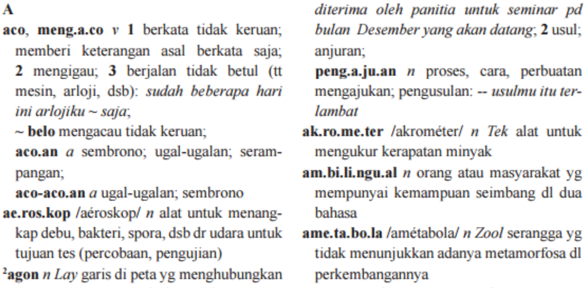
Bidang
Komputer, Olahra...

17 November 2016

Word and Excel files



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



A

aco, **meng.a.co** *v* **1** berkata tidak keruan; memberi keterangan asal berkata saja; **2** mengigau; **3** berjalan tidak betul (tt mesin, arloji, dsb): *sudah beberapa hari ini arlojiku ~ saja*;

~ **belo** mengacau tidak keruan;

aco.an *a* sembrono; ugal-ugalan; serampangan;

aco-aco.an *a* ugal-ugalan; sembrono

ae.ros.kop /aéroskop/ *n* alat untuk menangkap debu, bakteri, spora, dsb dr udara untuk tujuan tes (percobaan, pengujian)

^{2}**agon** *n Lay* garis di peta yg menghubungkan

diterima oleh panitia untuk seminar pd bulan Desember yang akan datang; **2** usul; anjuran;

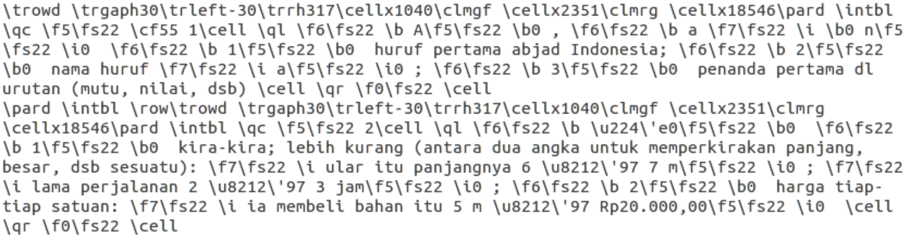
peng.a.ju.an *n* proses, cara, perbuatan mengajukan; pengusulan: -- *usulmu itu terlambat*

ak.ro.me.ter /akrométer/ *n Tek* alat untuk mengukur kerapatan minyak

am.bi.li.ngu.al *n* orang atau masyarakat yg mempunyai kemampuan seimbang dl dua bahasa

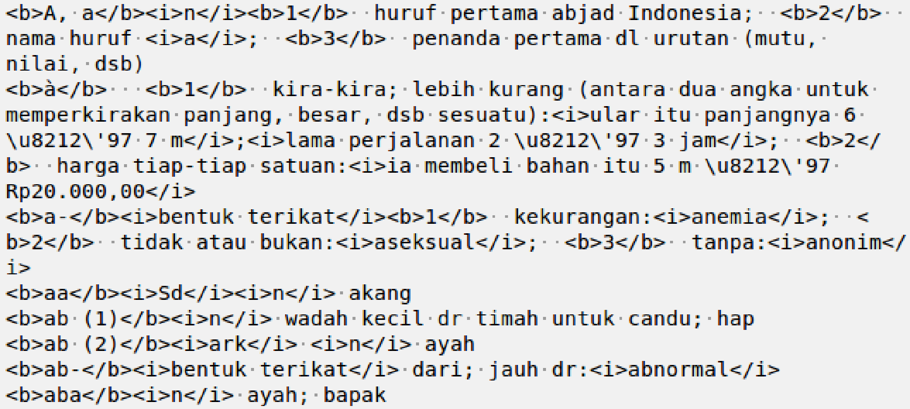
ame.ta.bo.la /amétabola/ *n Zool* serangga yg tidak menunjukkan adanya metamorfosa dl perkembangannya

From Word and Excel to Rich Text Format (rtf)



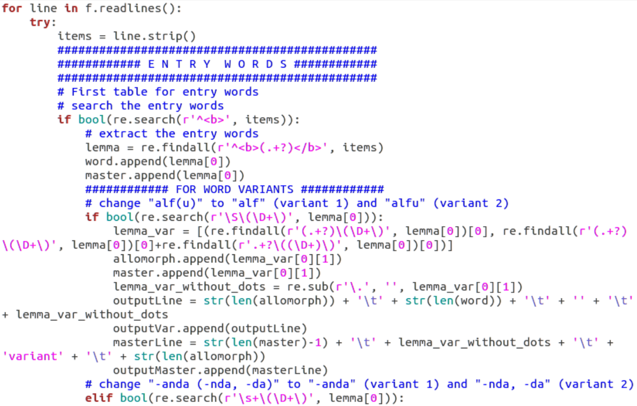
```
\trowd \trgaph30\trleft-30\trrh317\cellx1040\clmgf \cellx2351\clmrg \cellx18546\pard \intbl
\qc \f5\fs22 \cf55 1\cell \ql \f6\fs22 \b A\f5\fs22 \b0 , \f6\fs22 \b a \f7\fs22 \i \b0 n\f5
\fs22 \i0  \f6\fs22 \b 1\f5\fs22 \b0  huruf pertama abjad Indonesia; \f6\fs22 \b 2\f5\fs22
\b0  nama huruf \f7\fs22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

From rtf to HyperText Markup Language (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>  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>
<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
```

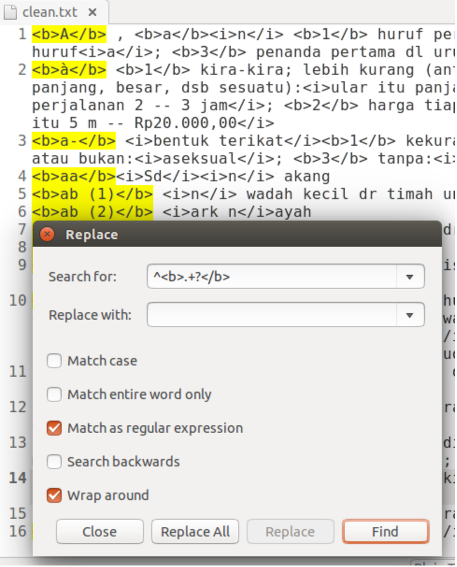
Using Python...



```
for line in f.readlines():
    try:
        items = line.strip()
        ##############################################
        ############ E N T R Y  W O R D S ############
        ##############################################
        # 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'(.+?)
\(\D+\)', lemma[0])[0]+re.findall(r'.+?\((\D+)\)', 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])):
```

The data was broken down by lemmas, sublemmas (derived words, compounds, proverbs, and idioms), labels, pronunciations, definitions, examples, scientific names, and chemical formulas using **regular expression**, a language for specifying text search strings which requires a pattern that we want to search for and a corpus of texts to search through [4].

Regular expression



```
<b>A</b> , <b>a</b><i>n</i> <b>1</b> huruf pe
huruf<i>a</i>; <b>3</b> penanda pertama dl ur
<b>à</b> <b>1</b> kira-kira; lebih kurang (an
panjang, besar, dsb sesuatu):<i>ular itu panj
perjalanan 2 -- 3 jam</i>; <b>2</b> harga tia
itu 5 m -- Rp20.000,00</i>
<b>a-</b> <i>bentuk terikat</i><b>1</b> kekur
atau bukan:<i>aseksual</i>; <b>3</b> tanpa:<i
<b>aa</b><i>Sd</i><i>n</i> akang
<b>ab (1)</b> <i>n</i> wadah kecil dr timah u
<b>ab (2)</b> <i>ark n</i>ayah
```

Replace

Search for: ^<b>.+?</b>

Replace with:

□ Match case

□ Match entire word only

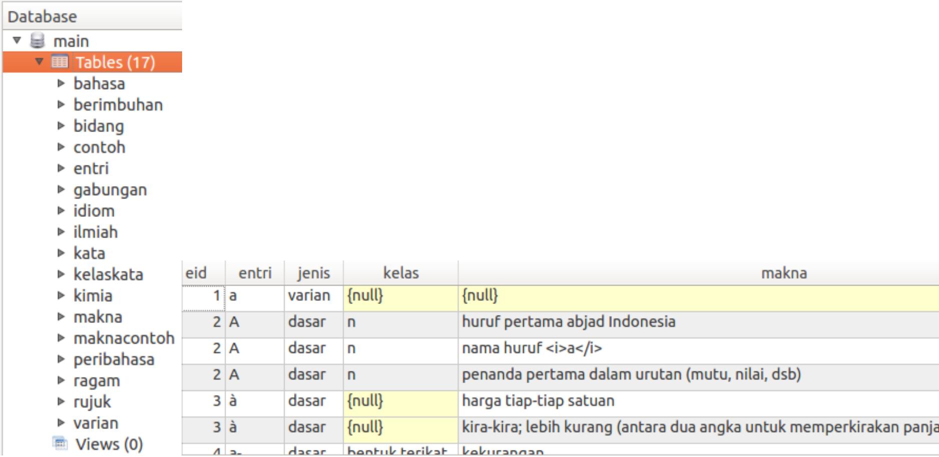
☑ Match as regular expression

□ Search backwards

☑ Wrap around

KBBI Database

SQLite (`www.sqlite.org`)



Database

- main
  - Tables (17)
    - bahasa
    - berimbuhan
    - bidang
    - contoh
    - entri
    - gabungan
    - idiom
    - ilmiah
    - kata
    - kelaskata
    - kimia
    - makna
    - maknacontoh
    - peribahasa
    - ragam
    - rujuk
    - varian
  - Views (0)

eid	entri	jenis	kelas	makna
1	a	varian	{null}	{null}
2	A	dasar	n	huruf pertama abjad Indonesia
2	A	dasar	n	nama huruf <i>a</i>
2	A	dasar	n	penanda pertama dalam urutan (mutu, nilai, dsb)
3	à	dasar	{null}	harga tiap-tiap satuan
3	à	dasar	{null}	kira-kira; lebih kurang (antara dua angka untuk memperkirakan panja
4	a-	dasar	bentuk terikat	kekurangan

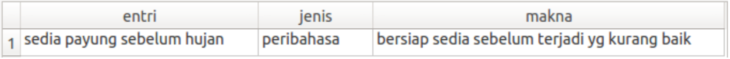
The current state of the KBBI Database

- Lemmas: 48,140
- Derived words: 26,197
- Compound words: 30,375
- Proverbs: 2,039
- Idioms: 267
- Entries (total): 108,238
- Definition sentences: 126,635
- Examples: 29,251

What can we get from KBBI Database? I

1. More specific and targeted word lookups, e.g.
   - looking up phrases and MWEs such as compound words, idioms, and proverbs as well as derived words

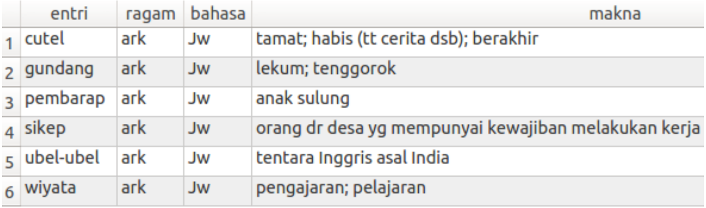
```
SELECT entri, jenis, makna FROM baseview WHERE entri="sedia payung sebelum hujan";
```



	entri	jenis	makna
1	sedia payung sebelum hujan	peribahasa	bersiap sedia sebelum terjadi yg kurang baik

   - looking up entries by their labels (part-of-speech, language, and domain labels)

```
SELECT entri, ragam, bahasa, makna FROM baseview WHERE ragam="ark" and bahasa="Jw";
```



	entri	ragam	bahasa	makna
1	cutel	ark	Jw	tamat; habis (tt cerita dsb); berakhir
2	gundang	ark	Jw	lekum; tenggorok
3	pembarap	ark	Jw	anak sulung
4	sikep	ark	Jw	orang dr desa yg mempunyai kewajiban melakukan kerja
5	ubel-ubel	ark	Jw	tentara Inggris asal India
6	wiyata	ark	Jw	pengajaran; pelajaran

What can we get from KBBI Database? II

2. Lexicography analysis
   - extracting the most frequent words in the definition sentences → can be used as a lexical set for the Indonesian learner's dictionary

Word	Freq.	Word	Freq.	Word	Freq.
yang	43,613	untuk	10,312	pada	6,793
dan	26,221	dalam	8,638	orang	6,110
atau	14,414	di	8,537	tentang	4,746
sebagainya	12,410	tidak	7,756	seperti	3,422
dengan	12,016	dari	7,280	...	...

   - extracting the most frequent genus terms in the definition sentences

Word	Freq.	Word	Freq.	Word	Freq.
orang	2,703	perihal	823	sesuatu	573
proses	1,858	tempat	806	kata	557
alat	1,595	menjadikan	745	pohon	547
tidak	1,526	yang	664	mempunyai	526
bagian	835	hasil	656	...	...

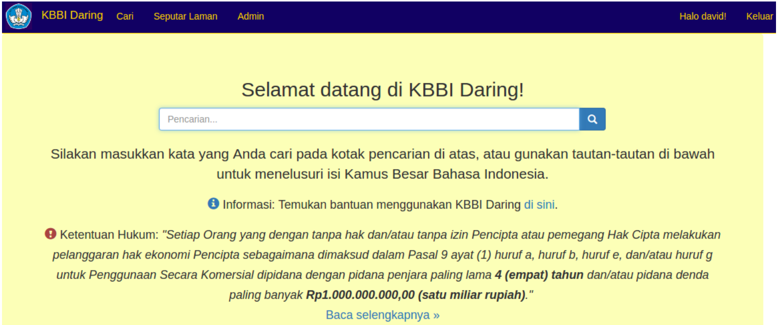
What can we get from KBBI Database? III

3. Linguistic analysis
   - grouping the derived words based on affixes and patterns of reduplication in Indonesian

Affix/Redup.	**Example**	**Number**	**Percentage**
meN-	**meng**abadi	5,185	21.1%
meN-...-kan	**meng**abadi**kan**	2,884	11.7%
ber-	**ber**abang	2,704	11.0%
-an	abai**an**	1,873	7.6%
peN-...-an	**peng**abadi**an**	1,780	7.2%
...	...	...	...
	Total	24,587	100.0%

4. Online and offline applications etc.

The Online KBBI V



Selamat datang di KBBI Daring!

Silakan masukkan kata yang Anda cari pada kotak pencarian di atas, atau gunakan tautan-tautan di bawah untuk menelusuri isi Kamus Besar Bahasa Indonesia.

ℹ Informasi: Temukan bantuan menggunakan KBBI Daring di sini.

❗ Ketentuan Hukum: *"Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf a, huruf b, huruf e, dan/atau huruf g untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama* ***4 (empat) tahun*** *dan/atau pidana denda paling banyak* ***Rp1.000.000.000,00 (satu miliar rupiah)****."*

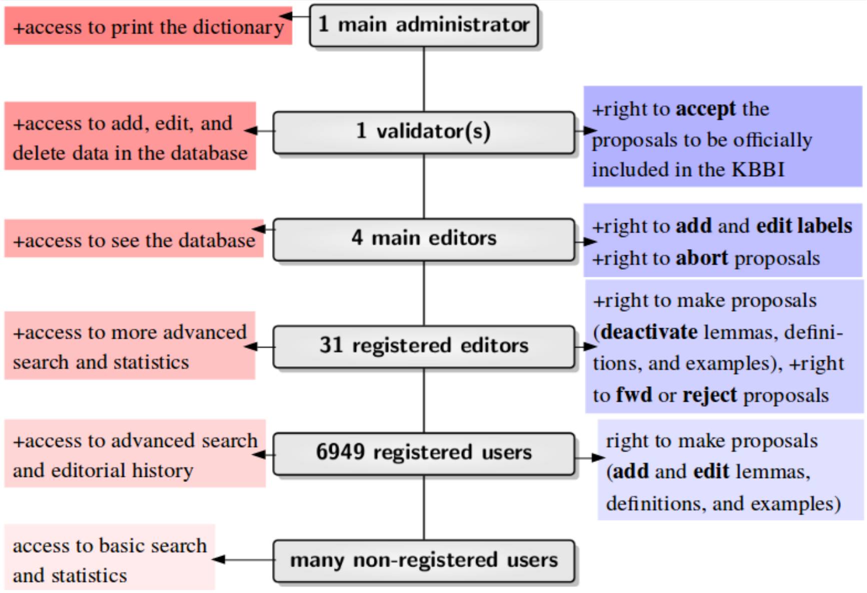
Baca selengkapnya »

- officially launched on 28 October 2016 [1], its user interface and the system were made using ASP.NET (`www.asp.net`).
- `https://kbbi.kemdikbud.go.id/`
- **Dictionary Writing System (DWS)** [2] which enables lexicographers to compile and edit dictionary text, as well as to facilitate project management, typesetting, and output to printed or electronic media

Some features in the Online KBBI

	Before 28 Oct 2016	**After 28 Oct 2016**
Word search	basic (by roots)	advanced (+by labels etc.)
Lexicographical workflow	done within the editorial board in Badan Bahasa	+online public participation to add, edit, and deactivate lemmas, definitions, and examples (crowdsourcing)
Security system	data can be easily crawled	customized security system to protect the data from web crawlers
Print function	no print function	print function can convert the data in the database to print format

Lexicographical workflow in the Online KBBI



1 main administrator

+access to print the dictionary

1 validator(s)

+access to add, edit, and delete data in the database

+right to **accept** the proposals to be officially included in the KBBI

4 main editors

+access to see the database

+right to **add** and **edit labels**
+right to **abort** proposals

31 registered editors

+access to more advanced search and statistics

+right to make proposals (**deactivate** lemmas, definitions, and examples), +right to **fwd** or **reject** proposals

6949 registered users

+access to advanced search and editorial history

right to make proposals (**add** and **edit** lemmas, definitions, and examples)

many non-registered users

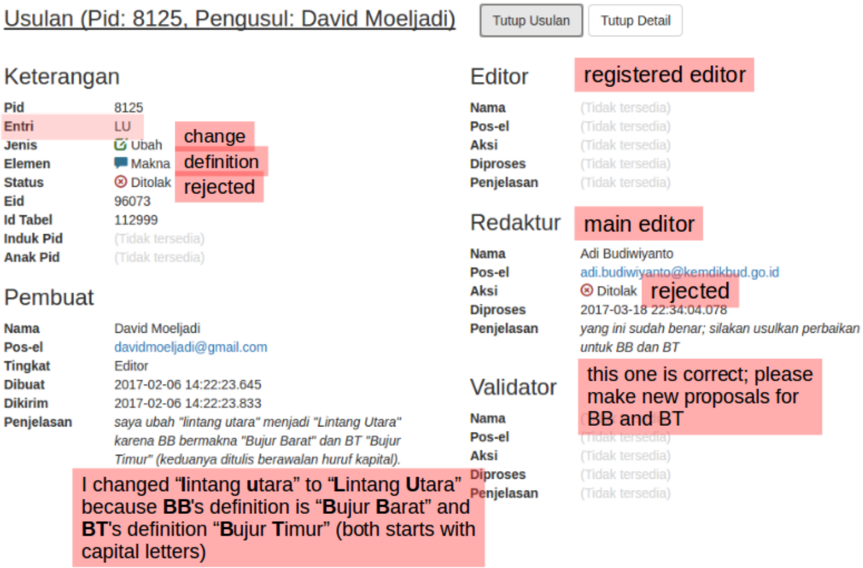
access to basic search and statistics

How a new lemma can be included in KBBI?

1. Having a unique concept
   NOT OK **si.ha.lu.an** *v* saling bertemu (cf. **ber.se.mu.ka**)
2. According to the Indonesian spelling rules
   NOT OK **ojeg** *n* sepeda atau sepeda motor yang ditambangkan dengan cara memboncengkan penumpang atau penyewanya (cf. **ojek**)
3. Euphonic (being pleasing to the ear)
   NOT OK **la.bu.la.bu.wai** *n* nasi yang diberi air putih ditambah garam atau ikan asin
4. Having positive connotations
5. Having a high frequency of use

Dora Amalia, p.c.

Rejected proposal



Usulan (Pid: 8125, Pengusul: David Moeljadi)

Keterangan

Pid	8125
Entri	LU
Jenis	Ubah
Elemen	Makna
Status	Ditolak
Eid	96073
Id Tabel	112999
Induk Pid	(Tidak tersedia)
Anak Pid	(Tidak tersedia)

change

definition

rejected

Pembuat

Nama	David Moeljadi
Pos-el	davidmoeljadi@gmail.com
Tingkat	Editor
Dibuat	2017-02-06 14:22:23.645
Dikirim	2017-02-06 14:22:23.833
Penjelasan	*saya ubah "lintang utara" menjadi "Lintang Utara" karena BB bermakna "Bujur Barat" dan BT "Bujur Timur" (keduanya ditulis berawalan huruf kapital).*

I changed “**l**intang **u**tara” to “**L**intang **U**tara” because **BB**'s definition is “**B**ujur **B**arat” and **BT**'s definition “**B**ujur **T**imur” (both starts with capital letters)

Editor

registered editor

Nama	(Tidak tersedia)
Pos-el	(Tidak tersedia)
Aksi	(Tidak tersedia)
Diproses	(Tidak tersedia)
Penjelasan	(Tidak tersedia)

Redaktur

main editor

Nama	Adi Budiwiyanto
Pos-el	adi.budiwiyanto@kemdikbud.go.id
Aksi	Ditolak
Diproses	2017-03-18 22:34:04.078
Penjelasan	*yang ini sudah benar; silakan usulkan perbaikan untuk BB dan BT*

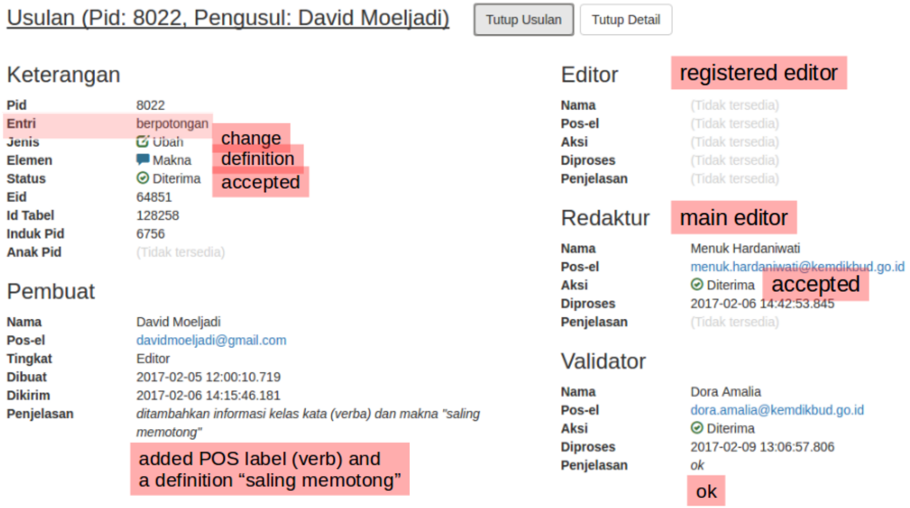
rejected

this one is correct; please make new proposals for BB and BT

Validator

Nama	(Tidak tersedia)
Pos-el	(Tidak tersedia)
Aksi	(Tidak tersedia)
Diproses	(Tidak tersedia)
Penjelasan	(Tidak tersedia)

Accepted proposal



Usulan (Pid: 8022, Pengusul: David Moeljadi)

Keterangan

Pid	8022
Entri	berpotongan
Jenis	Ubah
Elemen	Makna
Status	Diterima
Eid	64851
Id Tabel	128258
Induk Pid	6756
Anak Pid	(Tidak tersedia)

change

definition

accepted

Pembuat

Nama	David Moeljadi
Pos-el	davidmoeljadi@gmail.com
Tingkat	Editor
Dibuat	2017-02-05 12:00:10.719
Dikirim	2017-02-06 14:15:46.181
Penjelasan	*ditambahkan informasi kelas kata (verba) dan makna "saling memotong"*

added POS label (verb) and a definition “saling memotong”

Editor

registered editor

Nama	(Tidak tersedia)
Pos-el	(Tidak tersedia)
Aksi	(Tidak tersedia)
Diproses	(Tidak tersedia)
Penjelasan	(Tidak tersedia)

Redaktur

main editor

Nama	Menuk Hardaniwati
Pos-el	menuk.hardaniwati@kemdikbud.go.id
Aksi	Diterima
Diproses	2017-02-06 14:42:53.845
Penjelasan	(Tidak tersedia)

accepted

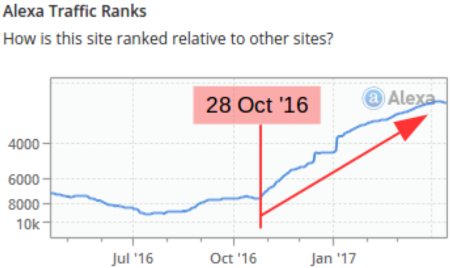
Validator

Nama	Dora Amalia
Pos-el	dora.amalia@kemdikbud.go.id
Aksi	Diterima
Diproses	2017-02-09 13:06:57.806
Penjelasan	*ok*

ok

Current situation (as of 20 April 2017)

- Word lookups
  - Total: 2,733,592 (10.93/minute, 653.90/hour, 15,741.62/day)
- Proposals
  - Total: 8,375 (48.23/day)
  - Accepted: 2,681
  - Rejected: 494
  - Being processed: 4,732
- Popularity (according to Alexa Traffic Ranks `www.alexa.com`)
  - Global rank: 2,548
  - Rank in Indonesia: 64



Alexa Traffic Ranks

How is this site ranked relative to other sites?

28 Oct '16

Future work

- add etymological information
- connect to corpora
- link to other lexical resources such as Wordnet Bahasa [3]

Acknowledgments

- Thanks to Dora Amalia for the KBBI IV data and her support
- Thanks to Francis Bond and Luis Morgado da Costa for the precious advice on the database structure
- Thanks to Ivan Lanin for improving the database
- Thanks to Ian Kamajaya for building the Online KBBI
- Thanks to Randy Sugianto for creating the Android application
- Thanks to Jaya Satrio Hendrick for designing the Android and iOS applications
- Thanks to Lie Gunawan for creating the iOS application
- Thanks to NTU HSS library support staff: Rashidah Ismail, Raihana Abdul Wahid, and Tan Chuan Ko for allowing me to borrow KBBI IV paper dictionary for months; and to Wong Oi May who helped us order the dictionary

References

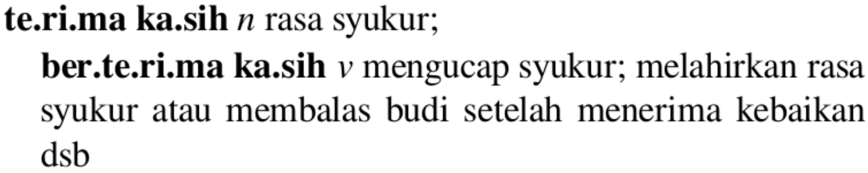
Dora Amalia, ed. *Kamus Besar Bahasa Indonesia*. 5th ed. Jakarta: Badan Pengembangan dan Pembinaan Bahasa, 2016.

B. T. Sue Atkins and Michael Rundell. *The Oxford Guide to Practical Lexicography*. Oxford University Press, 2008.

Francis Bond et al. “The combined Wordnet Bahasa”. In: *NUSA: Linguistic studies of languages in and around Indonesia* 57 (2014), pp. 83–100.

Daniel Jurafsky and James H. Martin. *Speech and Language Processing*. 2nd ed. New Jersey: Pearson Education, Inc., 2009.

Dendy Sugono, ed. *Kamus Besar Bahasa Indonesia Pusat Bahasa*. 4th ed. Jakarta: PT Gramedia Pustaka Utama, 2008.



te.ri.ma ka.sih *n* rasa syukur;

ber.te.ri.ma ka.sih *v* mengucap syukur; melahirkan rasa syukur atau membalas budi setelah menerima kebaikan dsb