Tool untuk menambah Leksem bahasa-bahasa di Indonesia

https://leksem-indonesia.toolforge.org/

Penggunaan

1. Memasukkan leksem secara manual

1. Klik Login di https://leksem-indonesia.toolforge.org/
2. Pilih salah satu kelas kata, misalnya:
   Leksem kelas nomina
   atau
   Leksem kelas nomina (peng-)
3. Isi masing-masing bentukan.
4. Apabila ada yang tidak *applicable*, lewati/kosongkan tidak apa-apa.
5. Hasil: https://www.wikidata.org/wiki/Lexeme:L523837

1. Memasukkan leksem secara massal

1. Klik Login
2. Pilih salah satu kelas kata, lalu klik di bagian “mode jumlah banyak”
3. Ikuti petunjuk di halaman tersebut. Masing-masing bentukan (*form*) dipisahkan oleh pipa. Apabila ada yang tidak *applicable*, lewati/kosongkan tidak apa-apa. (pipa + pipa)
4. Prasyarat: data harus disiapkan terlebih dahulu (lihat caranya di halaman berikutnya).
5. Pro Tip: mendaftarkan bot dan masuk dengan akun bot sebelum melakukan suntingan massal, atau berpotensi diblokir.

2. Mengapa memasukkan data leksem secara massal?

- Jumlah leksem yang puluhan hingga ratusan ribu per bahasa, tidak logis untuk mengandalkan memasukkan secara manual.
- Lebih cepat, lebih konsisten, lebih terstruktur (tidak ada yang terlewat)
- Lexeme ID-nya urut dan beraturan 🤗

- Data harus lengkap dan rapi – memakan waktu lama di persiapan data (*data cleanup*)
- Hanya lema saja, tidak perlu definisi. Bisa mengambil dari data kamus mana pun (mis. kamus bahasa daerah)
- Untuk masing-masing bentukan, harus memahami linguistik dan tata bahasa bahasa tersebut terlebih dahulu

3. Mempersiapkan data untuk input massal

1. Siapkan pdf kamus
2. Salin tempel semua isinya, ambil hanya lema (*headword*) saja. Abaikan definisi
3. Menggunakan *spreadsheet* atau program sejenisnya, buat setiap kolom bentukan yang dari lema tersebut
4. Pindahkan ke penyunting teks, ganti tab dengan pipa, simpan sebagai teks (txt)
5. Data siap dimasukkan ke leksem-indonesia.toolforge.org.

3. Memp massal

1. Siapkan pd
2. Salin tempe aja. Abaikan definisi
3. Menggunak setiap kolom bentukan y
4. Pindahkan an sebagai teks (txt)
5. Data siap d

Lister - [d:\Documents\nomina-peng.txt]

File Edit Options Encoding Help 4 %

```
pengabai|abai|pengabai|pengabaiku|pengabaimu|pengabainya|pengabailah|pengabaikulah|pengabai
pengabdi|abdi|pengabdi|pengabdiku|pengabdimu|pengabdinya|pengabdilah|pengabdikulah|pengabdi
pengabur|abur|pengabur|pengaburku|pengaburmu|pengaburnya|pengaburlah|pengaburkulah|pengabur
pengacah|acah|pengacah|pengacahku|pengacahmu|pengacahnya|pengacahlah|pengacahkulah|pengacah
pengacara|acara|pengacara|pengacaraku|pengacaramu|pengacaranya|pengacaralah|pengacarakulah|
pengacum|acum|pengacum|pengacumku|pengacummu|pengacumnya|pengacumlah|pengacumkulah|pengacum
pengada|ada|pengada|pengadaku|pengadamu|pengadanya|pengadalah|pengadakulah|pengadamulah|pen
pengadang|adang|pengadang|pengadangku|pengadangmu|pengadangnya|pengadanglah|pengadangkulah|
pengadil|adil|pengadil|pengadilku|pengadilmu|pengadilnya|pengadillah|pengadilkulah|pengadil
pengadopsi|adopsi|pengadopsi|pengadopsiku|pengadopsimu|pengadopsinya|pengadopsilah|pengadop
pengadu|adu|pengadu|pengaduku|pengadumu|pengadunya|pengadulah|pengadukulah|pengadumulah|pen
pengaduk|aduk|pengaduk|pengadukku|pengadukmu|pengaduknya|pengaduklah|pengadukkulah|pengaduk
pengadun|adun|pengadun|pengadunku|pengadunmu|pengadunnya|pengadunlah|pengadunkulah|pengadun
pengajar|ajar|pengajar|pengajarku|pengajarmu|pengajarnya|pengajarlah|pengajarkulah|pengajar
pengaktif|aktif|pengaktif|pengaktifku|pengaktifmu|pengaktifnya|pengaktiflah|pengaktifkulah|
pengaku|aku|pengaku|pengakuku|pengakumu|pengakunya|pengakulah|pengakukulah|pengakumulah|pen
pengakuisisi|akuisisi|pengakuisisi|pengakuisisiku|pengakuisisimu|pengakuisisinya|pengakuisi
pengalang|alang|pengalang|pengalangku|pengalangmu|pengalangnya|pengalanglah|pengalangkulah|
pengamal|amal|pengamal|pengamalku|pengamalmu|pengamalnya|pengamallah|pengamalkulah|pengamal
pengamalgam|amalgam|pengamalgam|pengamalgamku|pengamalgammu|pengamalgamnya|pengamalgamlah|p
pengaman|aman|pengaman|pengamanku|pengamanmu|pengamannya|pengamanlah|pengamankulah|pengaman
pengamanah|amanah|pengamanah|pengamanahku|pengamanahmu|pengamanahnya|pengamanahlah|pengaman
pengamat|amat|pengamat|pengamatku|pengamatmu|pengamatnya|pengamatlah|pengamatkulah|pengamat
pengambil|ambil|pengambil|pengambilku|pengambilmu|pengambilnya|pengambillah|pengambilkulah|
pengamen|amen|pengamen|pengamenku|pengamenmu|pengamennya|pengamenlah|pengamenkulah|pengamen
pengampelas|ampelas|pengampelas|pengampelasku|pengampelasmu|pengampelasnya|pengampelaslah|p
pengampu|ampu|pengampu|pengampuku|pengampumu|pengampunya|pengampulah|pengampukulah|pengampu
pengampuh|ampuh|pengampuh|pengampuhku|pengampuhmu|pengampuhnya|pengampuhlah|pengampuhkulah|
pengampun|ampun|pengampun|pengampunku|pengampunmu|pengampunnya|pengampunlah|pengampunkulah|
pengamuk|amuk|pengamuk|pengamukku|pengamukmu|pengamuknya|pengamuklah|pengamukkulah|pengamuk
penganalisis|analisis|penganalisis|penganalisisku|penganalisismu|penganalisisnya|penganalis
pengancam|ancam|pengancam|pengancamku|pengancammu|pengancamnya|pengancamlah|pengancamkulah|
pengandak|andak|pengandak|pengandakku|pengandakmu|pengandaknya|pengandaklah|pengandakkulah|
pengandam|andam|pengandam|pengandamku|pengandammu|pengandamnya|pengandamlah|pengandamkulah|
penganduh|anduh|penganduh|penganduhku|penganduhmu|penganduhnya|penganduhlah|penganduhkulah|
pengangguk|angguk|pengangguk|penganggukku|penganggukmu|pengangguknya|pengangguklah|penangg
penganggur|anggur|penganggur|penganggurku|penganggurmu|penganggurnya|penganggurlah|penangg
pengangkat|angkat|pengangkat|pengangkatku|pengangkatmu|pengangkatnya|pengangkatlah|pengangk
pengangkut|angkut|pengangkut|pengangkutku|pengangkutmu|pengangkutnya|pengangkutlah|pengangk
pengangon|angon|pengangon|pengangonku|pengangonmu|pengangonnya|pengangonlah|pengangonkulah|
penganiaya|aniaya|penganiaya|penganiayaku|penganiayamu|penganiayanya|penganiayalah|pengania
penganjur|anjur|penganjur|penganjurku|penganjurmu|penganjurnya|penganjurlah|penganjurkulah|
pengantar|antar|pengantar|pengantarku|pengantarmu|pengantarnya|pengantarlah|pengantarkulah|
pengantara|antara|pengantara|pengantaraku|pengantaramu|pengantaranya|pengantaralah|penganta
pengantre|antre|pengantre|pengantreku|pengantremu|pengantrenya|pengantrelah|pengantrekulah|
penganut|anut|penganut|penganutku|penganutmu|penganutnya|penganutlah|penganutkulah|penganut
pengapit|apit|pengapit|pengapitku|pengapitmu|pengapitnya|pengapitlah|pengapitkulah|pengapit
```

Sejarah

Sejarah singkat mengapa leksem-indonesia lahir dan apa yang membuatnya seperti itu

1. Tahap pertama

April 2021 - Mei 2021

- Dimulai dari usulan penggabungan leksem Indonesia dan Malaysia, yang kontroversial oleh Pengguna:Mahir256 pada bulan April 2021 https://www.wikidata.org/wiki/Wikidata_talk:WikiProject_Indonesia#Lexemes_in_%22Malay%22_and_%22Indonesian%22 (diskusi sangat panjang)
- Kemudian ada keinginan membuktikan perbedaan kedua bahasa, secara leksikal (bukan definisi maupun gramatikal), saya berkesimpulan cara terbaik adalah menunjukan bahwa ada pengaruh serapan imbuhan dari bahasa gaul, terutama dalam bentuk-bentuk verba (kata kerja), imbuhan-imbuhan nonformal yang tidak dikenal di Melayu.
- Saya menemukan https://lexeme-forms.toolforge.org/ dan menghubungi Lucas untuk membuatkan formulir untuk bahasa Indonesia, yang sudah saya persiapkan, https://www.wikidata.org/wiki/Wikidata:Wikidata_Lexeme_Forms/Indonesian
  namun dihalang-halangi oleh Mahir256 https://www.wikidata.org/wiki/Topic:W7pbjos433zw8uum
- Karena tidak ada itikad baik dari keduanya, saya memutuskan untuk melakukan *fork* kode Lucas

2. Tahap kedua

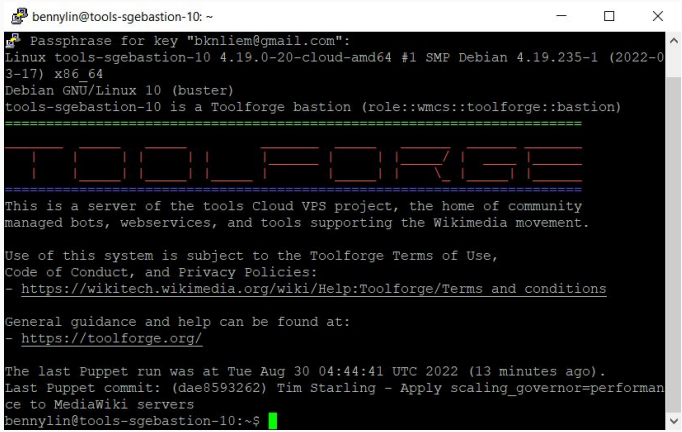
Juni 2021 - Agustus 2021

Supaya mudah, saya putuskan untuk melakukan fork di toolforge. Dokumentasi dapat dibaca di:

- https://wikitech.wikimedia.org/wiki/Portal:Toolforge
- https://wikitech.wikimedia.org/wiki/Help:Toolforge
- https://wikitech.wikimedia.org/wiki/Portal:Toolforge/Tool_Accounts
- https://wikitech.wikimedia.org/wiki/Help:Create_a_Wikimedia_developer_account#Toolforge_users

Langkah-langkah:

1. meminta akun
2. masuk log dengan PuTTY

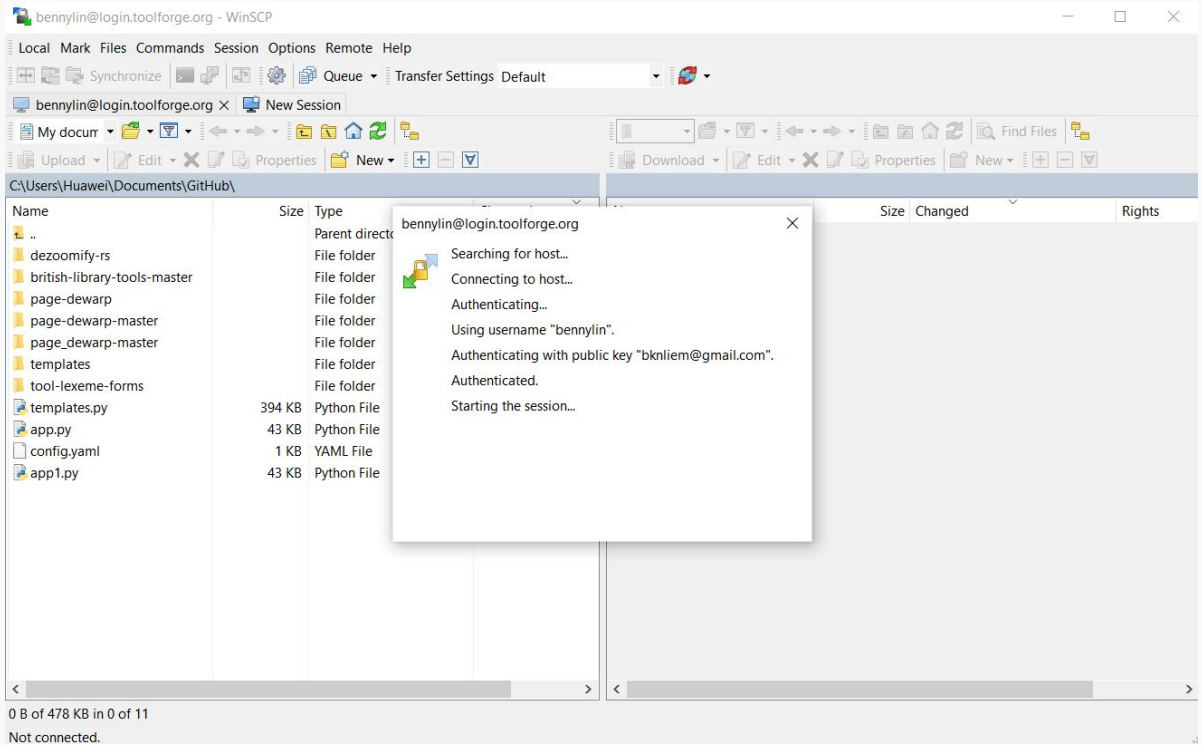


Tahap kedua (lanj.)

Juni 2021 - Agustus 2021

Langkah-langkah:

1. meminta akun
2. masuk log dengan PuTTY
3. masuk log dengan WinSCP



Tahap kedua (lanj.)

Juni 2021 - Juli 2021

Langkah-langkah:

1. meminta akun
2. masuk log dengan PuTTY
3. masuk log dengan WinSCP, unggah kode lexeme-forms dari Lucas ke /data/project/leksem-indonesia/www/python/src
4. unggah templates.py dan config.yaml kustom ke root folder /data/project/leksem-indonesia/
5. di putty, jalankan perintah: “become leksem-indonesia” lalu “cp templates.py $HOME/www/python/src/” Lakukan ini setiap kali ada perubahan templat (ada tambahan bahasa, dll.) kemudian jalankan perintah “webservice restart” supaya perubahannya berefek.

3. Tahap ketiga

Juni 2021 - Agustus 2021

(yang nomina tidak selesai, baru sampai di awalan peng-, tidak sempat dilanjutkan karena kehilangan interes / tujuan sudah tercapai, dan tidak sempat menyisihkan waktu lagi)

- Melengkapi bentukan (forms) masing-masing set data
- Melakukan riset
- Membersihkan data
- Memasukkan data lewat “mode jumlah banyak”

Leksem Indonesia (TODO) - nomina dan adjektiva
https://docs.google.com/spreadsheets/d/1DJNjpbqSwz-z-t0JOpeGvfTqwPptY3hl0d-WCzjDzao/edit#gid=674920041

Leksem Indonesia (done)-verba, tinggal verba dasar yg belum
https://docs.google.com/spreadsheets/d/1mFu_QGgjpYm4d-D0o1MDe5vYj22zu8q2pU6NDglOlDk/edit#gid=0

Leksem Jawa, Sunda, Gorontalo, dll.

Leksem Indonesia - non formal:
https://docs.google.com/spreadsheets/d/1t2aBiJKzXEQyniW1Mj9d2RyNZue-7UjsgKHg5EiEN7k/edit#gid=686325759

Pembagian dataset

Berdasarkan kelas kata, dan imbuhan

1. Berdasar kelas kata

Verba: ~15000 leksem

Nomina: ~43000

Adjektiva ~5000

Adverbia ~300

Numeralia ~200, pronomina (~50), partikel (~200): artikula 6, interjeksi 7, interogativa 45, konjungsi 24, preposisi 45,

1. Berdasar kelas kata

Verba: ~15000 leksem

Nomina: ~43000

Adjektiva ~5000

Adverbia ~300

Numeralia ~200, pronomina (~50), partikel (~200): artikula 6, interjeksi 7, interogativa 45, konjungsi 24, preposisi 45,

2. Nomina

Nomina: ~43000

- nomina dasar: 18000 leksem
- nomina majemuk dasar & turunan: 18000 leksem
- -an: ~1600 leksem nomina
- peng-: ~1500 leksem nomina
- peng-an: ~1500 leksem nomina
- per-an: ~500 leksem nomina
- ke-an: ~1400 leksem nomina
- se-: ~300 leksem nomina
- pe-, per-, keber-, keter, kepeng-, kese-, ketidak- (an)

2. Nomina

Nomina: ~43000

- nomina dasar: 18000 leksem
- nomina majemuk dasar & turunan: 18000 leksem
- -an: ~1600 leksem nomina
- **peng-: ~1500 leksem nomina**
- peng-an: ~1500 leksem nomina
- per-an: ~500 leksem nomina
- ke-an: ~1400 leksem nomina
- se-: ~300 leksem nomina
- pe-, per-, keber-, keter, kepeng-, kese-, ketidak- (an)

- Russian (101423)
- Estonian (83208)
- English (71657)
- Malayalam (63315)
- Swedish (36851)
- Latin (32183)
- Hebrew (29912)
- German (27488)
- Basque (22931)
- Spanish (20959)
- Bokmål (17431)
- Slovak (16475)
- Ukrainian (15967)
- Danish (14907)
- Czech (14193)
- French (13707)
- Indonesian (13119)

3. Nomina (peng-)

Nomina: peng-: ~1500 leksem nomina

- baru 100 dari 1282 yang sempat diunggah
- sisanya https://www.wikidata.org/wiki/User:Bennylin/peng-
- kita kerjakan sekarang

3. Nomina (peng-)

Nomina: peng-: ~1500 leksem nomina

- baru 100 dari 1282 yang sempat diunggah
- sisanya https://www.wikidata.org/wiki/User:Bennylin/peng-
- kita kerjakan sekarang

Tahap selanjutnya

- https://www.wikidata.org/wiki/User:Bennylin/Leksem_Indonesia

Selanjutnya

1. Membahas topik-topik yang masih didiskusikan

https://www.wikidata.org/wiki/Wikidata:Lexicographical_data/Documentation/Languages/id#Lemma

1. Spelling variants
2. Language codes
3. Regional languages & scripts!
4. Statements
5. Forms
6. Senses, gloss, definitions
7. Etc.: audio, pronunciation, examples, translations

2. Cobalah Query

https://www.wikidata.org/wiki/Wikidata:Lexicographical_data/Documentation/Languages/id#Queries

1. Indonesian nouns
2. Indonesian verbs
3. Indonesian adjectives
4. Get all existing Indonesian lexemes
5. Get the count of lexemes in Indonesian belonging to different lexical categories
6. Ordered by newest to oldest creation time

693001	wd:L693001	monumen	wd:Q1084	noun
692552	wd:L692552	betina	wd:Q1084	noun
692346	wd:L692346	bahwa	wd:Q36484	conjunction
692226	wd:L692226	steker	wd:Q1084	noun
691676	wd:L691676	atau	wd:Q36484	conjunction
691675	wd:L691675	juga	wd:Q380057	adverb
690861	wd:L690861	ke	wd:Q4833830	preposition
690715	wd:L690715	sebagai	wd:Q4833830	preposition
689613	wd:L689613	dari	wd:Q4833830	preposition
689607	wd:L689607	di	wd:Q4833830	preposition
689600	wd:L689600	dan	wd:Q36484	conjunction
619093	wd:L619093	adaptasi perubahan iklim	wd:Q1084	noun
618292	wd:L618292	mempergencar	wd:Q24905	verb
595062	wd:L595062	waktu	wd:Q1084	noun
583876	wd:L583876	borang	wd:Q1084	noun
582682	wd:L582682	fafifu wasweswos	wd:Q34698	adjective
582503	wd:L582503	hilir	wd:Q1084	noun
582501	wd:L582501	hulu	wd:Q1084	noun
580939	wd:L580939	cepu	wd:Q34698	adjective

3. Lihat pula

1. https://ordia.toolforge.org/language/Q9240 - daftar dan statistik leksem bahasa Indonesia di WIkidata
2. https://ordia.toolforge.org/language/ - statistik bahasa dengan leksem terbanyak di WIkidata
3. https://ordia.toolforge.org/search?q=air
4. https://machtsinn.toolforge.org/?lang=9240
5. Recent changes: https://www.wikidata.org/wiki/Special:RecentChanges?hidebots=1&hidecategorization=1&tagfilter=OAuth+CID%3A+2271&limit=50&days=30&urlversion=2

4. Alat-alat menarik

https://www.wikidata.org/wiki/Wikidata:Tools/Lexicographical_data

- https://orthohin.toolforge.org/add/id - menambahkan *sense* ke leksem tanpa *sense*
- https://hangor.toolforge.org/browse/id - semua leksem, bisa difilter dengan/tanpa *sense*

  Hauki is a proof-of-concept tool for searching and displaying Lexemes in a user-friendly way, using lemmas rather than L-numbers as identifiers. https://www.wikidata.org/wiki/User:Vesihiisi/Hauki

fafifu wasweswos

adjective [L582682]

Forms

fafifuwasweswos

Senses

- Kalimat yang mengandung banyak istilah teknis yang sulit dimengerti oleh orang awam.

Add a sense!

Examples (not attached to a Sense)

Sekali lagi, mending usahakan tetap jernih dalam diskusinya agar tidak jadi sekadar fafifuwasweswos.

Sudah, tidak usah diladeni. Apalagi yang fafifu wasweswos campur bahasa inggris.

Menurut saya demarkasi story ini terlalu longitudinal elementer. Greymorian dari Goetia saja tafsirnya tidak Mukhalafatu lil hawadisi.Apalagi nubuat hetero-cis-Apocrypha terhollowfikasi ini sangat ndakik-ndakik nggilani.

4. Alat-alat menarik

https://www.wikidata.org/wiki/Wikidata:Tools/Lexicographical_data

- https://orthohin.toolforge.org/add/id - menambahkan *sense* ke leksem tanpa *sense*
- https://hangor.toolforge.org/browse/id - semua leksem, bisa difilter dengan/tanpa *sense*
- https://dicare.toolforge.org/lexemes/party.php
- https://dicare.toolforge.org/lexemes/challenge.php
- https://ordia.toolforge.org/
- https://www.wikidata.org/wiki/Module:Lexeme-en

Snippet code

```
templates = {

        'nomina-indonesia': {
        '@attribution': {'users': ['Bennylin'], 'title':
'User:Bennylin/Leksem_Indonesia'},
        'label': 'Leksem kelas nomina',
        'language_item_id': 'Q9240',
        'language_code': 'id',
        'lexical_category_item_id': 'Q1084',
        'forms': [
        {
        'label': 'nama leksem',
        'example': 'Contoh: [anak]',
        'grammatical_features_item_ids': ['Q106644026'],
        },
        {
        'label': 'tunggal',
        'example': 'Contoh: [anak]',
        'grammatical_features_item_ids': ['Q110786'],
        },
        {
        'label': 'jamak',
        'example': 'Contoh: [anak-anak]',
        'grammatical_features_item_ids': ['Q146786'],
        },
        {
        'label': 'verba',
        'example': 'Contoh: [beranak]',
        'grammatical_features_item_ids': ['Q106614340'],
        },
dst.
```

templates.py

```
        'nomina-indonesia-peng': {
        '@attribution': {'users': ['Bennylin'], 'title':
'User:Bennylin/Leksem_Indonesia'},
        'label': 'Leksem kelas nomina (peng-)',
        'language_item_id': 'Q9240',
        'language_code': 'id',
        'lexical_category_item_id': 'Q1084',
        'forms': [
        {
        'label': 'kata benda (peng-)',
        'example': 'Contoh: [pengajar]',
        'grammatical_features_item_ids': ['Q106644026'],
        },
        {
        'label': 'kata dasar',
        'example': 'Contoh: [ajar]',
        'grammatical_features_item_ids': ['Q111029'],
        },

        {
        'section_break': True,
        'label': 'tunggal',
        'example': 'Contoh: [pengajar]',
        'grammatical_features_item_ids': ['Q110786'],
        },
        {
        'label': 'posesif orang pertama',
        'example': 'Contoh: [pengajarku]',
        'grammatical_features_item_ids': ['Q71470598'],
        },
dst
```

```
        'adjektiva-indonesia': {
        '@attribution': {'users': ['Bennylin'], 'title':
'User:Bennylin/Leksem_Indonesia'},
        'label': 'Leksem kelas adjektiva',
        'language_item_id': 'Q9240',
        'language_code': 'id',
        'lexical_category_item_id': 'Q34698',
        'forms': [
        {
        'label': 'nama leksem',
        'example': 'Contoh: [berapi-api]',
        'grammatical_features_item_ids': ['Q3482678'],
        },
        {
        'label': 'kata dasar',
        'example': 'Contoh: [api]',
        'grammatical_features_item_ids': ['Q111029'],
        },
        ],
        },
```

templates.py

```
        'adjektiva-indonesia-2': {
        '@attribution': {'users': ['Bennylin'], 'title':
'User:Bennylin/Leksem_Indonesia'},
        'label': 'Leksem kelas adjektiva (2)',
        'language_item_id': 'Q9240',
        'language_code': 'id',
        'lexical_category_item_id': 'Q34698',
        'forms': [
        {
        'label': 'nama leksem',
        'example': 'Contoh: [cantik]',
        'grammatical_features_item_ids': ['Q3482678'],
        },
        {
        'label': 'bentuk superlatif',
        'example': 'Contoh (ter-): [tercantik]',
        'grammatical_features_item_ids': ['Q1817208'],
        },
        {
        'label': 'nomina',
        'example': 'Contoh kata benda: [kecantikan]',
        'grammatical_features_item_ids': ['Q106614337'],
        },
        {
        'label': 'verba',
        'example': 'Contoh kata kerja: [mempercantik]',
        'grammatical_features_item_ids': ['Q106614338'],
        },
        {
        'label': 'tingkat eksesif',
        'example': 'Contoh (ke-an, terlalu ...): [kebanyakan]',
        'grammatical_features_item_ids': ['Q1385613'],
```

```
        'nomina-jawa': {
        '@attribution': {'users': ['Bennylin'], 'title':
'User:Bennylin/Leksem_Jawa'},
        'label': 'Leksem kelas nomina',
        'language_item_id': 'Q33549',
        'language_code': 'jav',
        'lexical_category_item_id': 'Q1084',
        'forms': [
        {
        'label': 'nama leksem',
        'example': 'Contoh: [anak]',
        'grammatical_features_item_ids': ['Q110786'],
        },
        ],
        },

        'adjektiva-jawa': {
        '@attribution': {'users': ['Bennylin'], 'title':
'User:Bennylin/Leksem_Jawa'},
        'label': 'Leksem kelas adjektiva',
        'language_item_id': 'Q33549',
        'language_code': 'jav',
        'lexical_category_item_id': 'Q34698',
        'forms': [
        {
        'label': 'nama leksem',
        'example': 'Contoh: [ayu]',
        'grammatical_features_item_ids': ['Q3482678'],
        },
        ],
        },
```

```
        'verba-jawa': {
        '@attribution': {'users': ['Bennylin'], 'title':
'User:Bennylin/Leksem_Jawa'},
        'label': 'Leksem kelas verba',
        'language_item_id': 'Q33549',
        'language_code': 'jav',
        'lexical_category_item_id': 'Q24905',
        'forms': [
        {
        'label': 'nama leksem',
        'example': 'Contoh: [tangi]',
        'grammatical_features_item_ids': ['Q179230'],
        },
        ],
        },
```

templates.py

Terima kasih!

Presentasi oleh Benny
untuk PPLL WMID
benny@wikimedia.or.id

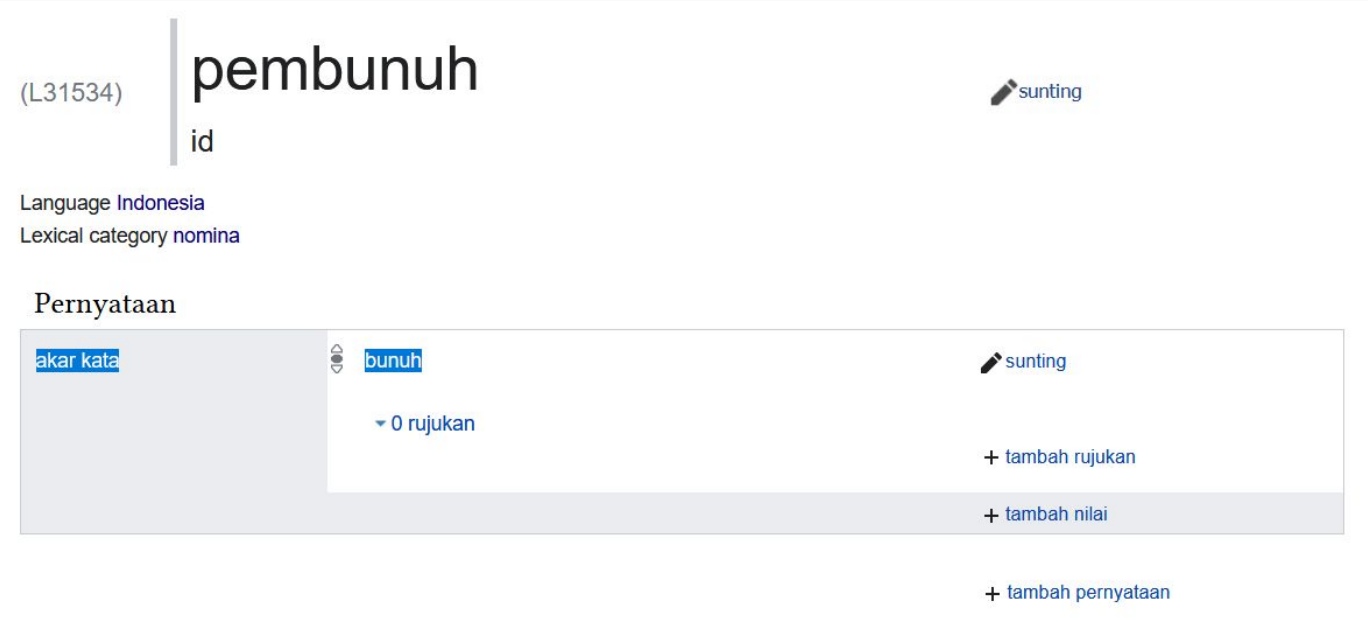
30-08-2022

Addendum

Yang lupa diutarakan pada pertemuan

Addendum

- Perlu tool lain (atau tool yang sudah ada perlu ditingkatkan) untuk bisa menambahkan “statement”, seperti akar kata, bandingkan https://www.wikidata.org/wiki/Lexeme:L31534



Addendum

- Perlu tool lain (atau tool yang sudah ada perlu ditingkatkan) untuk bisa menambahkan “statement”, seperti akar kata, bandingkan
  https://www.wikidata.org/wiki/Lexeme:L31534
  dengan
  https://www.wikidata.org/wiki/Lexeme:L693131

Addendum

- Perlu tool lain (atau tool yang sudah ada perlu ditingkatkan) untuk bisa menambahkan “statement”, seperti akar kata, bandingkan
- Big data, datanya akan dibuat di Wiktionary nantinya
- Datanya bersumber dari Wiktionary
  https://id.wiktionary.org/wiki/Kategori:id:Nomina
  https://id.wiktionary.org/wiki/Kategori:id:Verba
  dst.
- Pentingnya memasukkan pertama kali sudah benar semua, karena kalau sudah ada/sudah dibuat, sulit untuk menyunting.
- Bisa menggunakan Leksem-Indonesia untuk menyunting, mis.:
  https://leksem-indonesia.toolforge.org/template/nomina-indonesia-peng/edit/L31534
- Kesulitan utama memproses data dan memasukkan lema bahasa daerah adalah
  kamus-kamus bahasa daerah kebanyakan tidak dilengkapi dengan kelas kata
  sementara setiap leksem dikategorikan berdasarkan “kategori leksikal” (= kelas kata)
- [BUG]: Lexeme tidak dapat memiliki 2 leksem yang sama dengan kata dasar yang berbeda, misalnya
  https://www.wikidata.org/wiki/Lexeme:L696755, https://www.wikidata.org/wiki/Lexeme:L696801
  https://id.wiktionary.org/wiki/kelebatan, https://id.wiktionary.org/wiki/keluaran
  bisa dari akar kata “kelebat”/”keluaran” maupun “lebat”/”luar” (2 kata dengan “sense” yang berbeda)
- Belum ada proses penghapusan leksem di Wikidata(?)