

KBBI Daring: A Revolution in The Indonesian Lexicography

Ian Kamajaya¹, David Moeljadi², Dora Amalia³

¹ASTrio Pte Ltd, Singapore

² Nanyang Technological University, Singapore

³ Badan Pengembangan dan Pembinaan Bahasa, Indonesia

E-mail: ian@astriotech.com, D001@e.ntu.edu.sg, dora.amalia@kemdikbud.go.id

Abstract

Kamus Besar Bahasa Indonesia (KBBI) is the official dictionary of the Indonesian language, published by Badan Pengembangan dan Pembinaan Bahasa (The Language Development and Cultivation Agency) or Badan Bahasa, under the Ministry of Education and Culture, Republic of Indonesia. The current, fifth edition of KBBI (Amalia, 2016) was launched on 28 October 2016 and contains more than 100,000 entries and 120,000 senses. It is available in three formats: printed, online, and offline mobile applications. The online version, called KBBI Dalam Jaringan or KBBI Daring (kbbi.kemdikbud.go.id), is categorized as Dictionary Writing System (DWS) (Atkins & Rundell, 2008). Through it, we invite online public participation to make proposals to add and to edit entries, senses, and examples. We are changing our workflow from manual to computerized work which has greatly reduced the time needed to make a dictionary. KBBI Daring greatly expands the database which was previously made for the fourth edition of KBBI (Sugono, 2008) using the data in Microsoft Excel and Word files (Moeljadi et al., 2017), fitting to its online usage. This paper describes our efforts in building the KBBI Daring which has revolutionized both the way people use a dictionary and the lexicographical workflow of the editorial staff in Indonesia.

Keywords: online dictionary; Dictionary Writing System; Indonesian language; electronic lexicography; lexicographical workflow

1. Introduction

Kamus Besar Bahasa Indonesia (KBBI) is the official dictionary of the Indonesian language,¹ published by Badan Pengembangan dan Pembinaan Bahasa (The Language Development and Cultivation Agency) or Badan Bahasa, under the Ministry of Education and Culture, Republic of Indonesia. Up until present, KBBI is the most comprehensive and the most authoritative reference for the Indonesian language. Its first edition, published in 1988, has 62,000 entries. The number of entries increased to 72,000 or about 10,000 entries over three years in the second edition (1991). Its third edition, published in 2001, contains 78,000 entries and seven years later, the number of entries in the fourth edition increased to more than 92,000. Its latest, fifth edition was released for the first time in 2016 in three formats: printed, online, and offline mobile versions.²

The online KBBI before 28 October 2016, launched in 2006, used the data from the third edition of KBBI and allowed searches only by headwords. For example, to look up *mengacang*, a user must first look up the root word *kacang*, as shown in Figure 1. The sound assimilation process (a morphophonemic process) in *meN-* prefix makes the first

¹ Indonesian (ISO 639-3: ind), called *bahasa Indonesia* (lit. “the language of Indonesia”) by its speakers, is a Western Malayo-Polynesian language of the Austronesian language family. Within this subgroup, it belongs to the Malayic branch with Standard Malay in Malaysia and other Malay varieties (Lewis, 2009). It is spoken mainly in the Republic of Indonesia as the sole official and national language and as the common language for hundreds of ethnic groups living there (Alwi et al., 2014: 1–2). In Indonesia it is spoken by around 43 million people as their first language and by more than 156 million people as their second language (2010 census data). The lexical similarity is over 80% with Standard Malay (Lewis, 2009). It is written in Latin script. Morphologically, Indonesian is mildly agglutinative, compared to Finnish or Turkish. It has a variety of prefixes, suffixes, circumfixes, and reduplication.

² The printed version and the online application of KBBI were launched on 28 October 2016. The offline Android and iOS mobile applications were launched on 17 November 2016.

sound /k/ in the root word *kacang* become a nasal sound /ŋ/ (orthographically written as <ng>). This may present some difficulties if the user is not familiar with Indonesian morphological rules. In addition, it was not designed to support targeted lookups, such as search by word classes and entry types. Furthermore, it was not built to support addition or modification of any of its data – neither by editorial staff nor by other users. In short, it was solely built for the purpose of searching dictionary entries by headwords.

The new online KBBI called “KBBI Dalam Jaringan” or “KBBI Daring”³ was launched on 28 October 2016. Compared to its predecessor, it is designed with richer features which allow targeted and flexible, rather than exact, lookups. Besides, and more importantly, it serves as a unified and reliable platform for Indonesian lexicographers—which include not only professional lexicographers, but also common, non-professional Indonesian language users—across the world to enrich and to edit KBBI, increasing the efficiency of the editorial workflow. Although this concept is by no means new, it is a revolutionary work in Indonesia (or rather, globally, for Indonesian language users). Prior to KBBI Daring, editorial workflow in Indonesia was greatly scattered, unreliable, marked by many loss of editorial requests (such as enrichment or correction request from Indonesian language users – oftentimes sent by letter or by raising the issues during Indonesian language seminars), and slow response from Badan Bahasa to edit and to enrich KBBI against the emerging globalization that introduces a lot of new cultures, concepts, and technologies which demand a vast amount of new words and concepts to be included in KBBI. Figure 2 shows the main page of KBBI Daring. This paper describes our efforts in building KBBI Daring and is organized as follows: Section 2 describes how the editorial workflow was before KBBI Daring was built. Section 3 explains the features of KBBI Daring which are intended to deal with the situations described in Section 2. Section 4 reports the receptions from users and editorial staff. Section 5 concludes and mentions some future works.

2. Dictionary Use and Lexicographical Work before KBBI Daring

Before KBBI Daring was launched on 28 October 2016, we had manual lexicographical work, less public participations, and inconsistencies in the dictionary format.

2.1 Lexicographical Workflow

Editorial work for KBBI includes adding new entries, checking the accuracy of spellings, definitions, and examples, as well as formatting and layout. Before KBBI Daring was launched, the editorial staff worked manually. The data collectors collected some data sources, such as magazines, newspapers, and books in order to find new words not listed in KBBI by looking up those words in the paper dictionary. They recorded the new words on small pieces of card. In the first and second editions of KBBI, the cards were manually sorted in alphabetical order and placed in a special catalog cabinet. Other staff inputted the data of the new words into a file in a computer and did the formatting and layout. After the dictionary was printed, the editors wrote some notes on the pages if they found some errors and other staff fixed them in the file.

³ <https://kbbi.kemdikbud.go.id/>

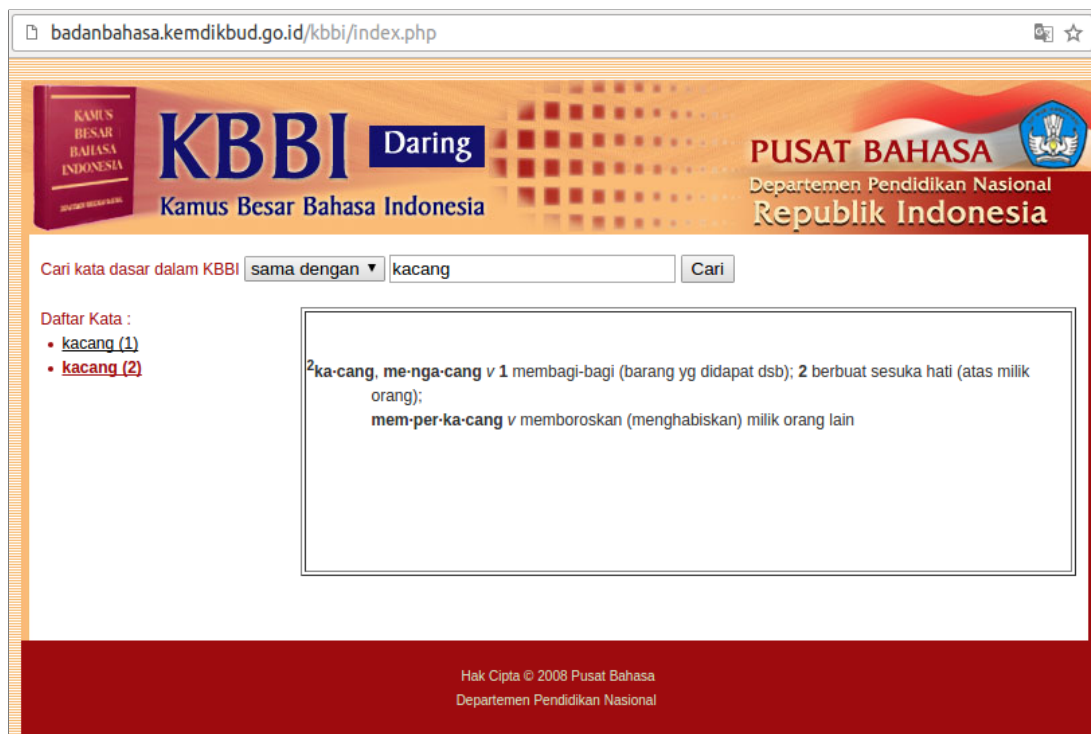


Figure 1: Screenshot of the online KBBI before 28 October 2016



Figure 2: Screenshot of the main page of KBBI Daring

2.2 Dictionary Use and Public Participations

Indonesian has a vast number of speakers (see Footnote 1). They rely on printed KBBI throughout Indonesia as the most authoritative source, while the printed dictionaries can be accessed only by limited users, i.e. those who can afford to buy them. In addition, Indonesia’s vast geographical condition also makes the distribution more difficult. Public contributions were also very limited. The public submitted proposals via post, email, and direct personal communication. They did not know whether their proposals had been processed or not. They did not even know whether their proposals had arrived at the editorial staff or not. There was no channel to inform them of the status of their proposals. Due to this limited access, the proposals received were also very small in number. Furthermore, the proposals submitted were checked by the editorial staff in Badan Bahasa without any help from experts outside the editorial team which might speed up the editorial work, mainly due to the limited access of the experts to the editorial requests.

2.3 Dictionary Format

Since KBBI was compiled and formatted manually by hand, errors (such as typos and inconsistencies in the formatting) are inevitable. In addition to some sporadic errors found and mentioned in Moeljadi et al. (2017), other errors in KBBI that were detected by our diagnostic tools in KBBI Daring (see Section 3.4) are, for example, compounds having word class labels (or contrarily, root words or derived words not having word class labels), duplicate entries, examples not containing the headwords used, and root words or derived words not having syllabifications.

3. KBBI Daring Features

KBBI Daring was built to deal with the issues described in Sections 1 and 2. This section explains some major features of KBBI Daring.

3.1 Dictionary Data Structure

KBBI Daring uses KBBI Database (Moeljadi et al., 2017). The database file is an SQLite file. The data structure of KBBI consists of four types of data: *entry*, *sense*, *example*, and *category*. The relationship between *entry* and *sense*, as well as the one between *sense* and *example* are one-to-many. The *category* is a list of descriptions or a metadata for *entry*, *sense*, and *example*. Figure 3 illustrates the KBBI data structure. An entry can be a fixed expression (*ungkapan*)⁴ or a root word (*kata dasar*). A fixed expression should have at least one sense and one example. In this case, one fixed expression may have one to multiple senses and one sense may have one to multiple examples. A root word should have at least one cross-reference, one sense, one compound, or one derived word. In this case, one root word may have zero to multiple senses and one sense may have zero to multiple examples. A root word may also have variant(s), proverb(s), and idiom(s). A proverb or an idiom should have at least one sense. A compound should have at least one cross-reference or one sense. One sense may have zero to multiple examples. Similar to the root word, a derived word should have at least one cross-reference, one sense, or one compound. It may also have variant(s), proverb(s), and idiom(s). The root word can be in the form of compound if it can be affixed and have derived word(s).

⁴ Fixed expressions are commonly used foreign phrases in Indonesian written works, such as *ad hoc*.

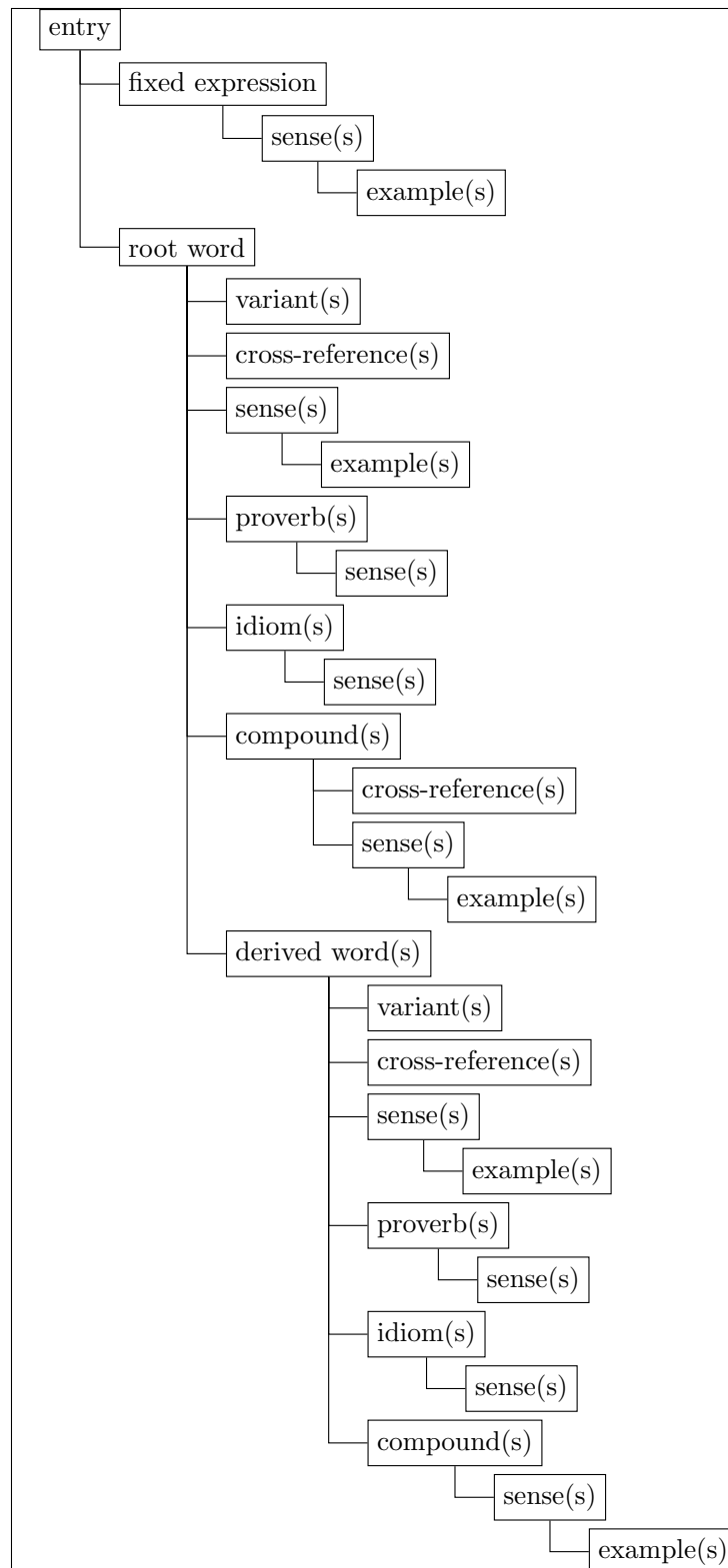


Figure 3: The KBBI data structure

3.2 User Groups and Privileges

User groups or user roles in KBBI Daring are primarily designed to represent the actual lexicographical workflow.⁵ Each user group is the realization of a certain group of Indonesian language users in the real-world counterpart. Therefore, KBBI Daring implements six groups of users: non-registered users, registered users, registered editors, main editors, validators, and one main administrator. Table 1 shows the privileges of each user group. A user in a higher group has all the privileges (features) of the lower groups. Figure 4 shows KBBI Daring system and access rights.

User group	(Additional) Features
Non-registered users	• Basic search tool
Registered users	+ Criteria-based search tool + Basic proposal tools
Registered editors	+ Advanced search tools + Advanced proposal tool + Basic editorial tools + Basic diagnostic tool
Main editors	+ Advanced editorial tools + Lexicographical tools
Validators	+ Validation privilege + Advanced diagnostic tool
Main administrator	+ Mass diagnostic and recovery tools + Printing tool

Table 1: (Additional) features for each user group in KBBI Daring

3.2.1 Non-registered Users

Non-registered users are those who use the KBBI Daring without registering their email addresses. They are only allowed to search using a basic search feature and are not allowed to make any proposal.

3.2.2 Registered Users

Registered users are those who use the KBBI Daring after registering their email addresses. They are given two additional privileges: to search based on predefined criteria and to make proposals to add or to edit the dictionary data. The criteria-based search tool is available on the registered user's main page. The users can search entries by their initial letters, word classes, styles, entry types, languages, and domains. Figure 5 shows some

⁵ Albeit in a website setting, where an additional administrator group is needed and non-registered users are given limited access to prevent anonymous stealing of the data.

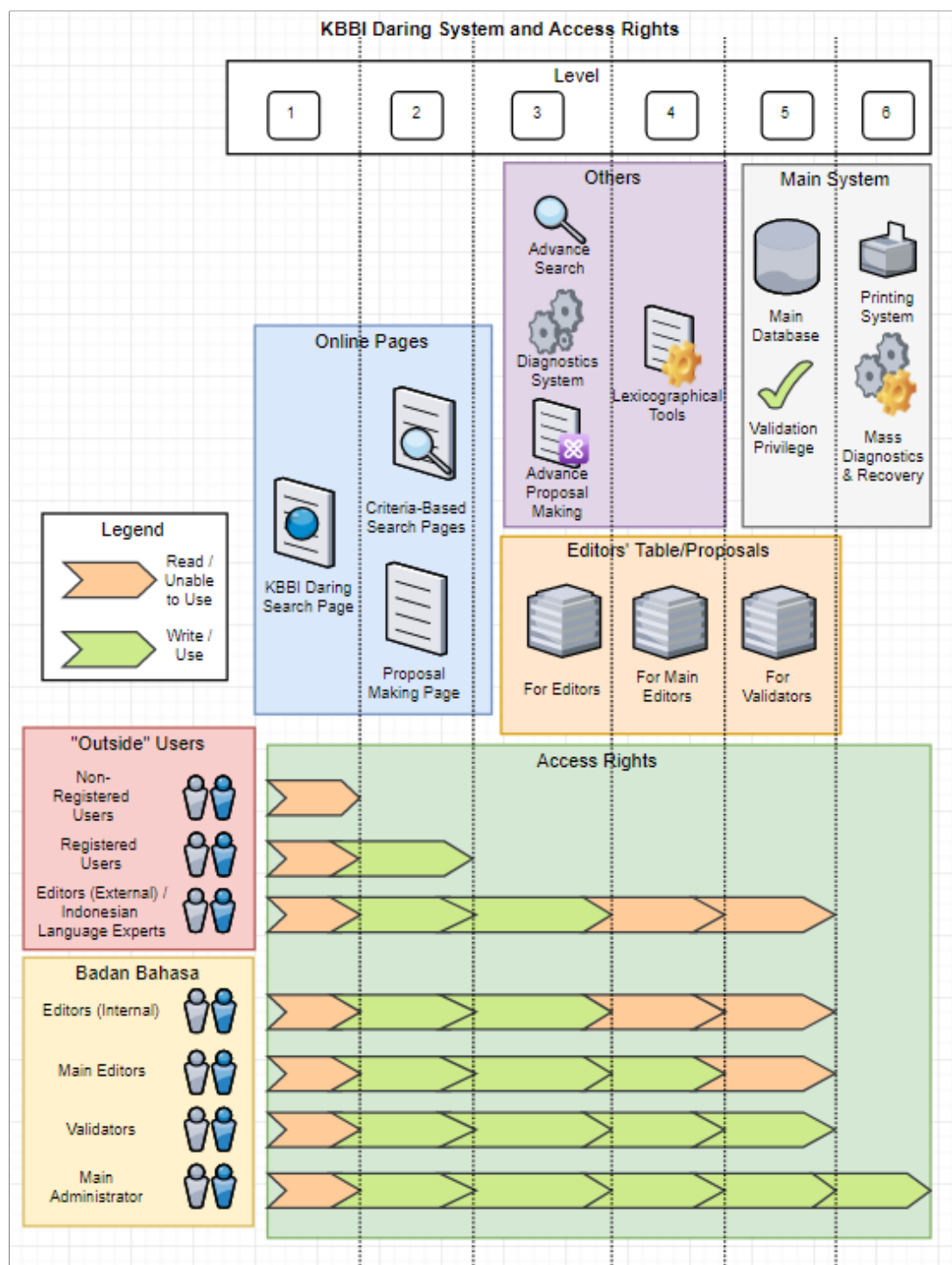


Figure 4: KBBI Daring system and access rights

proposal tools in the registered users' search result. The proposal system will be explained in Section 3.3. Figure 6 shows an example of a proposal page. Figure 7 shows an account management page where the users can check the status of their proposals.

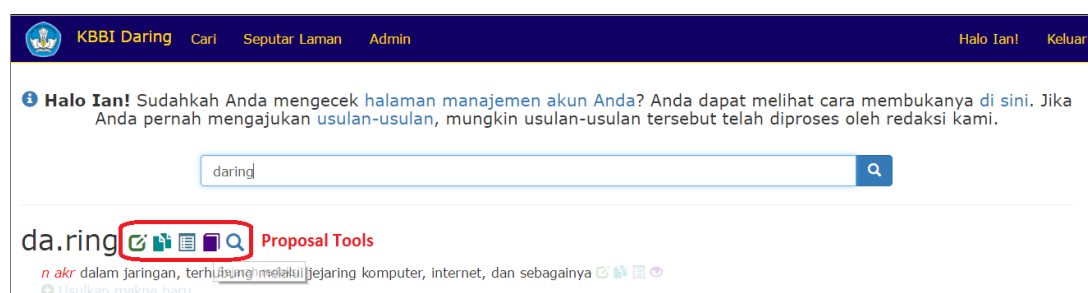


Figure 5: Screenshot of a search result for registered users

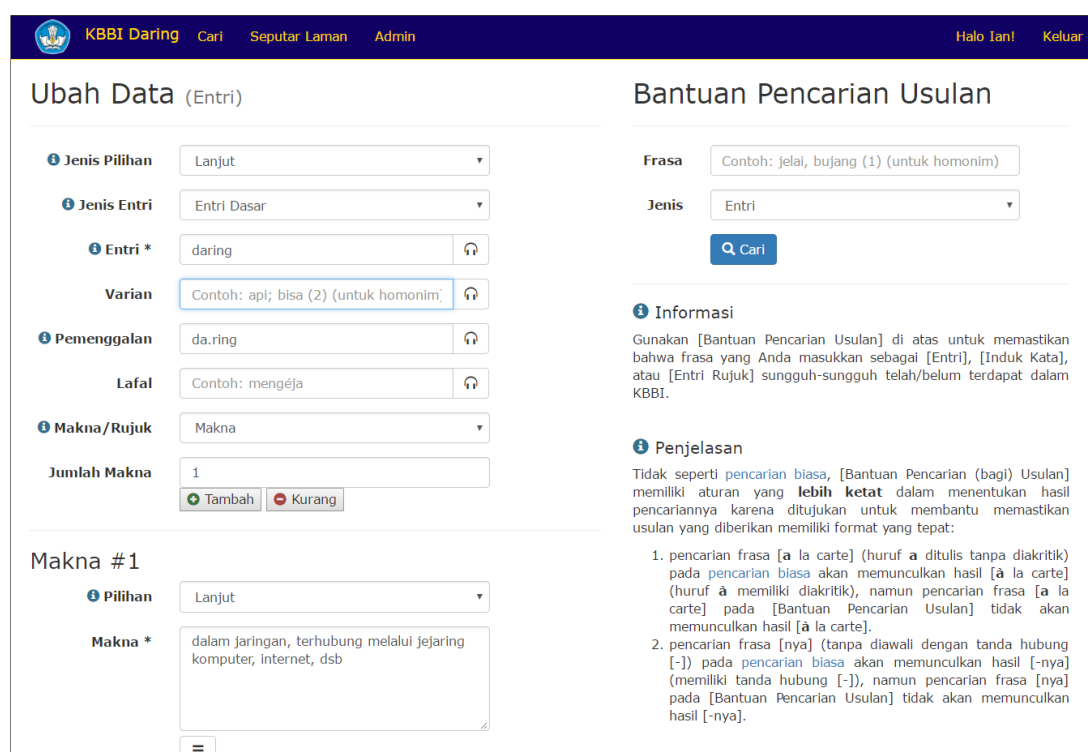


Figure 6: Screenshot of a proposal page

3.2.3 Registered Editors

Registered editors are the registered users who are granted privileges as KBBI editors because they understand linguistic issues and lexicographical theory and have attended lexicography workshops. They are given additional privileges to access advanced search tools, advanced proposal tools, basic diagnostic tools, and basic editorial tools. The advanced search tools allow them to search in the dictionary data structure (entry, sense, example, or category).



Figure 7: Screenshot of an account management page

The advanced proposal tool allows registered editors to make a deactivation proposal. The basic diagnostic tool allows registered editors to diagnose an entry to check if it has a formatting error. Figure 8 shows a search result page with the advanced proposal tool and the basic diagnostic tool.

The basic editorial tools allow registered editors to review the submitted proposals proposed by the registered users. A registered editor may accept and pass it up to the main editors, or take it over/change it, or return it to the registered user with comments, or reject it. Proposals which are taken over/changed become the registered editor's proposals. Proposals which are returned can be resubmitted but proposals which are rejected cannot. Figure 9 shows a list of proposals submitted by registered users to be further processed/reviewed by registered editors.

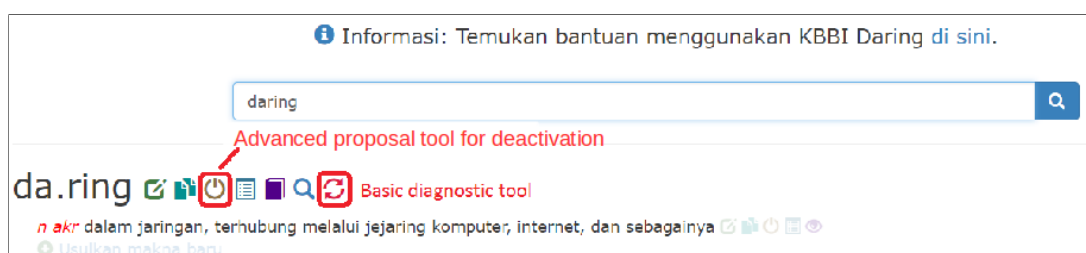


Figure 8: Screenshot of a search result with the advanced proposal tool and the basic diagnostic tool

3.2.4 Main Editors

The main editors are the ones inside Badan Bahasa whose main responsibility is to create KBBI. Compared to the registered editors, they have two additional privileges: access to

Daftar Usulan Bagi Editor								
Proposals to be reviewed by registered editors								
awal		Halaman 1 / 53		Penyaringan		(Usulan [1] persean - [20] apokrifa dari 1043 usulan)		
No	Entri	Jenis	Pid & Id Tabel	Jumlah	Pengusul & Editor	Redaktur & Validator	Status	Aksi
1	persean Eid: 111384		0000014026	1 1 0	Diki hidayatullah (2017-05-17 11:58:10.018) (Tidak tersedia)	(Tidak tersedia) (Tidak tersedia)	Diproses (Editor)	Detail Pass Teruskan Return Kembali Alihkan/Ubah Tolak
2	adil benar Eid: 111350		0000013952	1 1 1	Divi Karsena (2017-05-11 20:15:30.490) (Tidak tersedia)	(Tidak tersedia) (Tidak tersedia)	Diproses (Editor)	Detail Teruskan Return Kembali Alihkan/Ubah Tolak
3	adven Eid: 699		0000011065	0 1 0	Daniel Budiman (2017-03-29 23:12:36.637) (Tidak tersedia)	(Tidak tersedia) (Tidak tersedia)	Diproses (Editor)	Detail Teruskan Take over/Edit Alihkan/Ubah Tolak
4	agrikultur Eid: 110614		0000010192	1 0 0	ardianto bahтияr (2017-03-08 12:27:46.645) (Tidak tersedia)	(Tidak tersedia) (Tidak tersedia)	Diproses (Editor)	Detail Teruskan Reject Kembali Alihkan/Ubah Tolak

Figure 9: List of proposals to be reviewed by registered editors

advanced editorial tools and access to lexicographical tools. The advanced editorial tools give the main editors additional options to pass down, to archive, or to abort a proposal. A proposal which is passed down is returned to the registered editors for reviewing. A proposal which is archived is not aborted, but cannot be further processed. A proposal which is aborted is essentially deleted – not only rejected or returned. The lexicographical tools allow the main editors to reorder the polysemies (senses) of an entry and to “redirect” or “reattach” a sense to a different entry. The polysemy reordering is particularly useful to determine which senses should appear first or later on a search result page and on paper. The advanced editorial tools and lexicographical tools are only given to the main editors and above. Figure 10 shows the (additional) advanced editorial tools for processing proposals in the main editors’ table. Figure 11 shows two lexicographical tools for the main editors to reorder the polysemies and to “redirect” a sense to a different entry.

3.2.5 Validators

Validators are the editors who have a right to decide whether a proposal should be accepted and are the last examiners of the proposals. Validators have two additional privileges compared to the main editors: validation privilege and advanced diagnostic tool. Due to the validation privilege, once a validator accepts a proposal, the change will be reflected in the website. The advanced diagnostic tool allows the validators to diagnose multiple elements at the same time. Figure 12 shows a validator’s option to accept a proposal. Figure 13 shows the advanced diagnostic tool to diagnose multiple elements.

3.2.6 Main Administrator

The main administrator is an additional role to the actual editorial roles. It is designed, however, to allow a single most privileged user to use the printing application in the website as well as to do mass diagnostics and to recover the data. Since these features

Daftar Usulan Bagi Redaktur

▼ Penyaringan

Proposals to be reviewed by main editors

awal <<< << < Halaman 1 / 166 > >> >>> akhir (Usulan [1] cepirit [20] nya dari 3313 usulan)

No	Entri	Jenis	Pid & Id Tabel	Jumlah	Pengusul & Editor	Redaktur & Validator	Status	Aksi
1	cepirit Fid: 110507		0000010060	1 1 0	Dewi Khairiah (2017-04-28 15:26:18.864) (Tidak tersedia)	(Tidak tersedia) Dora Amalia (2017-02-28 10:30:56.140)	Diproses (Redaktur)	Detail Teruskan Kembalikan Turunkan Alihkan/Ubah Tolak Gugurkan Arsipkan
2	peresean Eid: 111384		0000014026	1 1 0	Diki hidayatullah (2017-05-17 11:58:10.018) (Tidak tersedia)	(Tidak tersedia) (Tidak tersedia)	Diproses (Editor)	Detail Teruskan Kembalikan Alihkan/Ubah Tolak Gugurkan Arsipkan
3	2bongkrok Eid: 109666		0000005922	1 1 0	Kaher Dwi P. (2017-02-07 07:36:27.653) (Tidak tersedia)	Azhari Dasman Damis (2017-01-20 13:19:22.518) (Tidak tersedia)	Diproses (Redaktur)	Detail Teruskan Kembalikan Turunkan Alihkan/Ubah Tolak Gugurkan Abort Arsipkan

Figure 10: Advanced editorial tools

makan

Additional option to reorder the polysemies

ma.kan¹

Redirect sense to different entry

- memasukkan makanan pokok ke dalam mulut serta mengunyah dan menelannya: mereka -- tiga kali sehari
- memasukkan sesuatu ke dalam mulut, kemudian mengunyah dan menelannya: ia sedang -- pisang
- memasukkan sesuatu ke dalam mulut dan mengunyah-ngunyahnya: Nenek sedang -- sirih
- memasukkan sesuatu ke dalam mulut dan menelannya: pasien harus -- pil
- mengisap: -- candu
- memakai; memerlukan; menghabiskan (waktu, biaya, dan sebagainya): pembangunan jembatan ini -- waktu lama; upacara adat itu -- ongkos besar
- menyerang, mematikan, mengambil (dalam permainan catur): gajah -- bidak putih
- bekerja sebagaimana mestinya (tentang rem, gigi roda, dan sebagainya)
- melukai: air keras itu -- kulit
- mengenai; menembus: ditembaknya tiga kali, tetapi tidak --
- memperoleh sesuatu; mencapai sesuatu: layarnya tidak -- tidak memperoleh angin; sauhnya dapat -- mencapai dasar laut
- (dapat) masuk (tentang barang yang dimasukkan ke lubang, ke air): kapal ini lima meter --nya ke dalam air
- ki mengambil; mempergunakan dan sebagainya secara tidak sah (milik orang lain atau negara): ia telah -- pupuk milik koperasi
- ki meniduri perempuan (biasanya dalam arti hubungan gelap): pemuda itu -- anak gadis tetangganya sampai hamil
- ki rezeki: memberi --; diberi --

Usulkan makna baru

Figure 11: Lexicographical tools

Daftar Usulan Bagi Validator

Penyaringan

Proposals to be reviewed by validators

awal

Halaman 1 / 266

akhir

(Usulan [1] cepirit - [20] -nya dari 5303 usulan)

No	Entri	Jenis	Pid & Id Tabel	Jumlah	Pengusul & Editor	Redaktur & Validator	Status	Aksi	Ekstra
1	cepirit Eid: 110507		0000010060	1 1 0	Dewi Khairiah (2017-04-28 15:26:18.864) (Tidak tersedia)	(Tidak tersedia) Dora Amalia (2017-02-28 10:30:56.140)	Diproses (Redaktur)	Detail Terima Kembalikan Turunkan Alihkan/Ubah Tolak Gugurkan Arsipkan	-

Figure 12: Validator's option to accept a proposal

Daftar Entri ▼ Penyaringan

awal ◀◀◀ ◀◀ ◀ Halaman 22 / 2730 ▶ ▶▶ ▶▶▶ akhir (Entri [841] - [880] dari 109186 entri)

Advance diagnostic tools
diagnosing multiple
elements at once

Eid	Entri	Jenis	Silabel	Lafal	Induk	Rujuk	Aktif	Lampiran	Diagnosis	Aksi
772	afektif	dasar	afek.tif	afèktif			✓		🟢	Ubah Detail Non-Aktifkan Diagnosis
98334	afektivitas	dasar	afek.ti.vi.tas	afèktivitas			✓		🟢	Ubah Detail Non-Aktifkan Diagnosis
102682	afektivitas negatif	gabungan			98334		✓		🟢	Ubah Detail Non-Aktifkan Diagnosis
102683	afektivitas positif	gabungan			98334		✓		🟢	Ubah Detail Non-Aktifkan Diagnosis
773	aferen	dasar	afe.ren	afèrén			✓		⚠ 1	Ubah Detail Non-Aktifkan Diagnosis
774	aferesis	dasar	afe.re.sis	afèrésis			✓		🟢	Ubah Detail Non-Aktifkan Diagnosis

Figure 13: Advanced diagnostic tool

will consume high resources in the server hardware, they are given only to one main administrator.

3.3 Proposal System

The proposal system in KBBI Daring is a guided, non-anonymous, transparent, crowdsourcing system. Its design is closely tied with the user groups and their privileges in order to transform the traditional lexicographical workflow to its current, public-friendly, transparency-imbued, web-based form – making it a lot more accessible for crowdsourcing while having an official body to guide the overall process.

Though most KBBI Daring users are non-registered users, the privilege to use the proposal system is only given to registered users and above. The reason is rather obvious: anonymous contribution for the official and the most authoritative reference for the Indonesian language is hardly a promising idea. Furthermore, such a feature may do more harm than good as it can be exploited anonymously to send “junk” proposals. Forcing registered accounts to access the proposal tools would limit the number of proposals generated from anonymous sources. People are required to register and to verify their email addresses for their accounts to be registered on the KBBI Daring website. Furthermore, fake and temporary email domains are filtered by the account registration system in KBBI Daring, leaving mostly only valid email addresses to be registered.

The proposal can target any one of the following three data types: entry, sense, or example. Additionally, the proposal made must be one of the following types: add, edit, or deactivate (only for registered editors and above). In the end, the validator must decide whether the proposal is acceptable or not. When a proposal is accepted, the targeted item in the database is replaced with the proposed item, the proposal data are logged (for historical references), and the changes are immediately reflected in the website. The contributor’s names, i.e. the proposal maker, the editors, and the validator, are shown in the editorial history. Figure 14 shows the editorial history of entry *Yesus Kristus* “Jesus Christ”. The editorial history, which is accessible to registered users and above, contains all essential information: details of the proposal, the proposer, the registered editor, the main editor,

and the validator of the proposal – each with their respective explanation or comment, the acceptance date and time, as well as the revision number. Thus, it enforces transparency of all the items added or changed in KBBI Daring through proposals.

The registered user group is designed to be the major group (in terms of number) in KBBI Daring which is given the privileges to participate in the enrichments and corrections of KBBI by creating reasonable proposals. It is the main crowdsourcing group which is designed to represent the “common” Indonesian language users (compared with different user groups in the following paragraphs) who are willing to contribute. Anyone who registers can immediately become a contributor. However, between the registered users and the validators, there are two groups: registered editors and main editors.

Registered editors review the proposals from registered users. They “recommend” good proposals to be passed up to the main editors, edit potential proposals with few flaws to be more acceptable, return and guide registered users to make better-formed proposals, or immediately reject the proposals when they are considered unacceptable in the first place (for example, if a user proposes a duplicate entry or a duplicate sense). The registered editor group is designed to represent the experts who want to participate in the lexicographical workflow of KBBI. They consist mostly of adept people in lexicography and linguistics. Consequently, the registered editors are expected to create good proposals and to be another major contributing group.

Main editors and validators consist exclusively of people inside Badan Bahasa who are responsible for KBBI, e.g. the chief editor of KBBI and the head of the lexicography subdepartment. Technically, they can make proposals, but they are not expected to do it as their prime task. Instead, using their official positions, their main task is to verify the acceptableness of proposals created by the registered users and the registered editors, and to focus more on the lexicographical work.

3.4 Mass Diagnostic and Recovery Tools

KBBI Daring also helps the editors find potential errors inherited from the previous editions, as mentioned in Section 2.3, or created by the users, and helps correct them. It is designed with mass diagnostic and recovery tools for that purpose. Figure 15 shows an example of mass diagnostic results.

Some errors, e.g. pronunciation containing letters other than *é*, *ê*, and *è*⁶ or definitions having certain words in the lengthened forms instead of the shortened forms (i.e. the word *seperti* “like, as” should be written in its shortened form *spt* in KBBI), are errors with a definite (single) solution. They are automatically correctable. On the other hand, other errors, e.g. duplicate entries with different senses or absence of syllables in root words or derivative words, are errors with non-definite solutions. They can only be corrected by humans. Table 2 shows the list of errors diagnosable using the mass diagnostic tool.

The recovery tool is designed to correct errors which have a definite (single) solution. It is not designed to correct errors with multiple viable solutions. Instead, the diagnostic tool

⁶ The pronunciation field in KBBI only deals with entries having the letter <e>. The Indonesian language has the sounds [e], [ə], and [ɛ]. However, they are not orthographically distinguished in the current spelling system, all of them are written as <e> (Alwi et al., 2014). The pronunciation field in KBBI indicates them as *é*, *ê*, and *è* respectively.

Revisi #1 (Buat Entri) Tutup

First Revision

The proposal

Eid 109197

Entri **Yesus Kristus**

Jenis Entri **Gabungan**

Induk Kata Yesus (Eid: 107835)

Id Entri (Tidak tersedia)

Id Homonim (Tidak tersedia)

Aktif ☒

Varian (Tidak tersedia)

Pemenggalan (Tidak tersedia)

Lafal yésus kristus

Makna/Rujuk makna

Jenis Rujuk (Tidak tersedia)

Entri Rujuk (Tidak tersedia)

Jumlah 2

Makna

Makna #1

Mid 127469

Polisem 1

Makna Sang Mesias (Juru Selamat dunia) dalam ajaran agama Kristen; Pribadi kedua Allah Tritunggal; Firman yg menjadi daging; Anak (Tunggal) Allah, yg juga memiliki gelar-gelar lain seperti (1) Anak Daud, (2) Anak Manusia, (3) Imanuel, (4) Yang Kudus dr Allah, (5) Anak Domba Allah (yg disembelih dan yg menghapus dosa dunia), sesuai dng berbagai macam fungsi yg dijalankan-Nya. Dipercaya sbg yg akan menghakimi seluruh umat manusia dan malaikat pd akhir zaman; Yesus yg diurapi

Ragam (Tidak tersedia)

Ragam (Tidak tersedia)

Varian (Tidak tersedia)

Kelas Kata (Tidak tersedia)

Bahasa Kris (Agama Kristen)

Bidang Bukan Klasan

Klasan? (Tidak tersedia)

Tipe (Tidak tersedia)

Penyingkat (Tidak tersedia)

Redaksi

No Pid 3333

Diterima 2017-02-16 22:13:11.035

Jenis +

Induk Pid (Tidak tersedia)

Anak Pid 9497

Acceptance Date Time:
16 February 2017 22:13, local time

Pengusul Tutup

Proposer and his comment

Nama Ian Kamajaya

Tingkat Pengguna Terdaftar

Waktu 2017-01-05 16:17:54.461

Penjelasan Edit - 5 Jan 2017: Sesuai dengan permintaan validator, jumlah polisem telah dikurangi dari 11 menjadi 5 saja, yaitu yang paling umum, hanya yang muncul pada 4 kitab Injil. Awal - 4 Des 2016: Telah terdapat cukup banyak lema dalam KBBI yang merujuk pada "Yesus Kristus" baik dalam bidang agama Kristen, agama Islam, maupun agama Katolik, tetapi entri "Yesus Kristus" sendiri belum terdapat dalam KBBI. Khusus untuk pengertian secara Kristen, saya menambahkan selengkap mungkin gelar-gelar yang diberikan pada Yesus Kristus seperti yang tertulis dalam Alkitab untuk memudahkan rujukan entri-entri baru pada Yesus Kristus (jika ada) di kemudian hari. Saat ini sudah ada entri berupa gelar-gelar Yesus Kristus seperti Anak Allah dan Mesias yang sudah terdapat dalam KBBI.

Editor Tutup

Registered Editor and his comment

Nama David Moeljadi

Waktu 2016-12-05 08:48:41.809

Penjelasan (Tidak tersedia)

Redaktur Tutup

Main Editor and her comment

Nama Menuk Hardaniwati

Waktu 2017-01-13 10:24:24.214

Penjelasan usulan dapat dimengerti

Validator Tutup

Validator and her comment

Nama Dora Amalia

Waktu 2017-02-16 22:13:11.035

Penjelasan Penjelasan Otomatis: Usulan ini dikembalikan pada

Figure 14: The editorial history of an entry

is designed to detect such cases, to help human editors identify and correct the mistakes. Table 3 shows the list of errors which are correctable by the recovery tool.

3.5 Printing Tool

The printing tool is integrated as a part of KBBI Daring and is only accessible to the main administrator. The printing tool is built to make consistent formatting effects for the printed version. The major contribution is to eliminate human errors in the formatting. The major challenge is the complex and potentially-growing formatting effects. However, manual editing has advantages over machines because it can handle exceptions in the formatting rules more flexibly, especially if the exceptions are only small parts of the dictionary, yet varying enough to be handled non-uniformly with ease using the printing tool. Therefore, the printing tool is primarily designed to create the base version of the printed version – solid enough as not containing formatting errors, but flexible enough for human agents to do a little finishing touch before being sent to the printing company in a ready-to-print format.

The printing tool is designed with Microsoft-created .Net dynamic link library (.dll) Microsoft.Office.Interop.Word to produce the base version in .doc format. The setting page is provided in KBBI Daring to allow the main administrator to choose the format and the range of the printing. Afterwards, the printing command is given to the server and the server starts to write the .doc document for printing. Once finished, the resulting document is made available in KBBI Daring, accessible only by the main administrator. It can be downloaded and further processed by human agents to make the ready-to-print format of KBBI.

Data type	List of diagnoses
Entry	<ul style="list-style-type: none"> • Does it have duplicates with different senses? • Does it have identical variants? • Does it have syllabification (where it should not have and vice versa)? • Does it have pronunciation (where it should not have and vice versa)? • Does the pronunciation contain letters other than é, ê, and è? • For an entry having a root word, does it refer to an active root word? • For a cross-reference entry, does it have a correct reference item?
Sense	<ul style="list-style-type: none"> • Does it refer to a particular entry? • Does it have a word class (where it should not have and vice versa)? • Does it contain repeated scientific names (binomial names)?
Example	<ul style="list-style-type: none"> • Does it refer to a particular entry? • Does it refer to a particular sense? • Does it contain the corresponding headword? • Does the headword in the example have the same spelling? • Does it have misplaced spaces or punctuations?
Entry, Sense, and Example	<ul style="list-style-type: none"> • Does it have “odd” or “missing” IDs? • Does it contain shortened form of certain words while they should be written in their lengthened form (and vice versa)?

Table 2: Errors diagnosable by the mass diagnostic tool

Data type	Error	Correction
Entry	should be without syllabification but has syllabification	remove the syllabification
	does not contain ‘e’ but has pronunciation	remove the pronunciation
	contains ‘e’ but does not have pronunciation	give the default pronunciation ê
Sense	a compound which has a word class	remove the word class
	contains repeated scientific names (binomial names)	remove one of the scientific names
Entry, Sense, and Example	contains shortened form of certain words while they should be written in their lengthened form	lengthen the word
	contains lengthened form of certain words while they should be written in their shortened form	shorten the word

Table 3: Some errors and the corrections using the recovery tool


<div>  KBBI Daring </div> <div> Cari Admin Data </div> <div> Halo Ian! Keluar </div>			
<div> <div>Daftar Diagnosis Massal</div> <div> <div>Penyaringan</div> <div>1</div> </div> </div> <div> <div>awal</div> <div>◀◀</div> <div>◀</div> <div>▶▶</div> <div>▶</div> <div>akhir</div> </div> <div> <div>Halaman 1 / 27</div> <div>(Diagnosis Massal [1] - [40] dari 1058 diagnosis massal)</div> </div>			
<div> <div>Keterangan Diagnosis</div> <div>Waktu Mulai: 2017-03-30 17:19:05.156, Waktu Berakhir: 2017-03-30 17:39:35.499, Eid Awal: 80001, Jumlah Elemen: Semua</div> </div>			
<div> <div> <div> <div>No</div> <div>Id</div> <div>Entri</div> </div> <div> <div>Hasil Diagnosis Massal Terakhir</div> <div>Aksi</div> </div> </div> <div> <div> <div>1</div> <div>32</div> <div>mengaprit buih (Eid: 4914)</div> </div> <div> <div>Entri ini:</div> <div> <ul style="list-style-type: none"> mengaprit buih, jenis: gabungan, induk: 4912 </div> <div>memiliki kesalahan format berikut:</div> <div> <ol style="list-style-type: none"> Entri ini [mengaprit buih] bukan berjenis [varian], namun tidak memiliki potongan frasa yang mirip dengan [induk]nya, yaitu [apung-apung]. [Induk] entri ini mungkin seharusnya diubah </div> </div> <div> <div>Detail</div> <div>Cari</div> </div> </div> <div> <div> <div>2</div> <div>33</div> <div>asam alginat (Eid: 5368)</div> </div> <div> <div>Entri ini:</div> <div> <ul style="list-style-type: none"> asam alginat, jenis: gabungan, induk: 5362 </div> <div>memiliki kesalahan format berikut:</div> <div> <ol style="list-style-type: none"> Entri ini: <ul style="list-style-type: none"> [Entri: asam alginat, Eid: 5368, lafal:] memiliki 1 (satu) [homonim tak/berbeda bernomor] atau [duplikat] aktif berikut: <ul style="list-style-type: none"> [Entri: asam alginat, Eid: 103006, lafal:] </div> </div> <div> <div>Detail</div> <div>Cari</div> </div> </div> <div> <div> <div>3</div> <div>34</div> <div>asango (Eid: 5432)</div> </div> <div> <div>Makna #1 (Mid: 6491) Entri ini:</div> <div> <ul style="list-style-type: none"> W tempat hidangan makanan pejabat kerajaan </div> <div>memiliki kesalahan format berikut:</div> <div> <ol style="list-style-type: none"> Makna ini berasal dari entri berjenis [dasar] atau [turunan], yaitu [asango, jenis: dasar], namun tidak memiliki [kelas kata] </div> </div> <div> <div>Detail</div> <div>Cari</div> </div> </div> </div> <div> <div>List of entries containing known errors</div> <div>Diagnostic Results</div> </div>			

Figure 15: Mass diagnostic results

The base version of the KBBI printed version generated by the printing tool is a near ready-to-print version with the formatting effects specified by Badan Bahasa – except for the word hyphenation which occasionally needs to be corrected by human editors. Unlike L^AT_EX, the Microsoft Word 2013 tool used to generate the auto-hyphenation does not have a list of exceptions for correcting wrongly generated hyphenations.⁷ Word is used due to its popularity and its ease of integration with Adobe InDesign file (.indd format), which is used by Badan Bahasa for the finishing touch on the document and is occasionally required by major printing companies in Indonesia.

3.6 Others

Besides the features mentioned above, KBBI Daring is also equipped with a customized security system to protect the data from web crawlers.

4. KBBI Daring Impacts and Receptions

KBBI Daring made the editorial work more efficient. Automatic notifications of errors are shown by the machine and the editorial staff can focus only on the notified errors. For some formatting errors, the changes can even be done solely by the machine. The changes allow the team to focus more on the substantial issues, such as the accuracy of definitions. The conversion process from the database to the near ready-to-print version is automatically done, including the header on each page. However, formatting issues such

⁷ As of May 2017, L^AT_EX-based printing tool is still being built as an alternative version of the currently used Word-based printing tool because it can handle hyphenation better and saves more time.

as the ‘widow’ and ‘orphan’ line as well, as an adjustment of the columns on the last page of each letter section, need to be handled manually.

Dissemination was held to introduce KBBI Daring to the public (students, teachers and lecturers, journalists, translators and writers) for the Indonesian vocabulary enrichment program. The dissemination event has been held in 15 provinces in Indonesia in 2016. As of May 2017, it was held in 15 other provinces in 2017. In the event, the participants are encouraged to use KBBI Daring and to make proposals. The number of proposals received from the participants varies, ranging from the fewest (fewer than 10) to the largest (more than 50). We expect that the participants will continue using KBBI Daring.

As of 26 May 2017, KBBI Daring has been used to search entries more than 3.4 million times and has accumulated more than 9,800 proposals for the dictionary’s enrichments and corrections. Alexa site⁸ shows the domain as the 81st most searched domain in Indonesia and the first (most searched) among all the domains ending with `go.id`.

5. Conclusion and Future Works

We have described our work in building KBBI Daring which revolutionized the lexicographical workflow, helping the editorial staff work more efficiently, and involves more public participations in enriching and improving the dictionary. It also minimizes formatting errors in the printed version which are inevitable in the previous editions. In the future, we will add etymological information and connect KBBI Daring to corpora and lexical databases such as Wordnet Bahasa (Bond et al., 2014). We plan to publish supplements for the printed version every six months in order to provide the users with the current lexicon and other lexicographical information which reflect the language used by the society over time.

6. Acknowledgements

Thanks to Ardianto Suhendar for creating some base code for the printing tools using `Elistia.DotNetRtfWriter`. Thanks to elistia⁹ and Serg-Norseman¹⁰ for creating and making it available for public use in the GitHub. Thanks to Ivan Lanin for his valuable examples of the main page and search results for KBBI Daring. Thanks to Francis Bond for his ideas for KBBI Daring features, especially the deactivation feature and the feature to create new proposals based on the already existing items.

7. References

- Alwi, H., Dardjowidjojo, S., Lapoliwa, H. & Moeliono, A.M. (2014). *Tata bahasa baku bahasa Indonesia*. Jakarta: Balai Pustaka, 3 edition.
- Amalia, D. (ed.) (2016). *Kamus Besar Bahasa Indonesia*. Jakarta: Badan Pengembangan dan Pembinaan Bahasa, 5 edition.
- Atkins, B. & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

⁸ www.alexacom

⁹ <https://github.com/elistia>

¹⁰ <https://github.com/Serg-Norseman>

- Bond, F., Lim, L.T., Tang, E.K. & Riza, H. (2014). The combined Wordnet Bahasa. *NUSA: Linguistic studies of languages in and around Indonesia*, 57, pp. 83–100.
- Lewis, M.P. (ed.) (2009). *Ethnologue: languages of the world*. Dallas: SIL International, 16 edition.
- Moeljadi, D., Kamajaya, I. & Amalia, D. (2017). Building the Kamus Besar Bahasa Indonesia (KBBI) Database and Its Applications. In *Proceedings of The 11th International Conference of the Asian Association for Lexicography*. Guangzhou, pp. 64–80.
- Sugono, D. (ed.) (2008). *Kamus Besar Bahasa Indonesia Pusat Bahasa*. Jakarta: PT Gramedia Pustaka Utama, 4 edition.

This work is licensed under the Creative Commons Attribution ShareAlike 4.0 International License.

<http://creativecommons.org/licenses/by-sa/4.0/>

