

Building the *Kamus Besar Bahasa Indonesia* (KBBI) Database and Its Online Application

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

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 Ministry of Education and Culture, Republic of Indonesia. KBBI Fourth Edition (KBBI IV) (Sugono 2008) which was published in 2008, had its data in Microsoft Excel and Word files. KBBI official online dictionary before October 2016 was made based on the data from the third edition (KBBI III) and was only meant for simple word search by root. For example, if we want to search the word *kebaikan* “goodness”, we should type its root *baik* “good” in the search field, click the search button, and we will get all information (derived words, compounds, proverbs, senses) under the root *baik*, exactly in the same format as the one in the printed dictionary. There was no way we could search derived words or compounds directly by typing them in the search field because the data was not structured and there was no database for it. In this talk, I will describe my efforts in building the KBBI database and the online KBBI application (*Kamus Besar Bahasa Indonesia Dalam Jaringan* or *KBBI Daring*) which was developed using the database.

Using Python programming language (www.python.org), the KBBI IV data in the Excel and Word files was broken down by entries, subentries, types of subentries (derived words, compounds, proverbs, and idioms), labels (parts-of-speech, styles, languages, domains, and abbreviations), pronunciations, senses, examples, scientific names, and chemical formulas. After the data was broken down, cleaned, and given unique IDs, the first database for KBBI was built. Afterwards, the database was expanded with more than 30,000 new entries for the fifth edition (KBBI V) (Amalia 2016) which was launched on 28 October 2016. Its online version, KBBI Daring (kbbi.kemdikbud.go.id), was developed to enable lexicographers to compile and edit dictionary text, as well as to facilitate project management, typesetting, and output to printed or electronic media. It is categorized as Dictionary Writing System (DWS) (Atkins and Rundell 2008, p. 113-117). Its user interface and the system were made using ASP.NET (www.asp.net).

KBBI Daring has revolutionized both the way people use a dictionary and the lexicographical workflow of the editorial staffs. Through KBBI Daring, we invite online public participation to make proposals to add and to edit entries (words), senses, and examples. As of 15 February 2017, it has already been used to search entries more than 1.4 million times and has accumulated more than 6,400 proposals for the dictionary’s corrections. Alexa site (www.alexa.com) shows the domain as the 57th most searched domain in Indonesia and the first (most searched) among all the domains ended with `go.id`. In the future, etymological information will be added, corpora and lexical databases such as Wordnet Bahasa (Bond et al. 2014) will be connected to KBBI.

Participants at ISMIL will be invited to register and make proposals (adding and/or editing entries, subentries, senses, or examples) in KBBI Daring.

References

- Amalia, Dora, ed. (2016). *Kamus Besar Bahasa Indonesia*. 5th ed. Jakarta: Badan Pengembangan dan Pembinaan Bahasa.
- Atkins, B. S. and M. Rundell (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Bond, Francis et al. (2014). “The combined Wordnet Bahasa”. In: *NUSA: Linguistic studies of languages in and around Indonesia* 57, pp. 83–100.
- Sugono, Dendy, ed. (2008). *Kamus Besar Bahasa Indonesia Pusat Bahasa*. 4th ed. Jakarta: PT Gramedia Pustaka Utama.