

PENYUSUNAN KAMUS ELEKTRONIS

Mirna Adriani

Fakultas Ilmu Komputer - Universitas Indonesia

1 Latar Belakang

Kamus merupakan buku pelengkap yang penting dan bermanfaat bagi masyarakat. Saat ini kamus dalam pelbagai bidang telah tersedia untuk membantu masyarakat memperoleh informasi yang sering diperlukan namun tidak diketahui arti atau maknanya. Kamus merupakan acuan yang dapat diandalkan bila kita mengalami kesulitan dalam mempelajari suatu bahasa, ingin mengetahui arti suatu istilah dalam suatu bidang secara mendalam, atau hanya ingin memperoleh padanan kata bahasa asing dalam Bahasa Indonesia.

Sebagai orang awam di bidang bahasa, kami juga menggunakan kamus bahasa sebagai pedoman bila kami merasa tidak yakin mengenai definisi ataupun penulisan kata-kata yang kami perlukan. Kebutuhan akan adanya pedoman bahasa yang benar makin kami perlukan ketika kami mulai mengembangkan fasilitas pemeriksa ejaan elektronis untuk Bahasa Indonesia karena kami memerlukan kosa kata yang benar sebagai pembanding kata-kata yang tertulis di dalam dokumen elektronis. Sebagai kamus pembanding kami menggunakan Kamus Besar Bahasa Indonesia (KBBI) karena sepengetahuan kami KBBI merupakan kamus standar bagi bahasa Indonesia.

Setelah menggunakan KBBI beberapa lama, kami mulai merasa bingung karena adanya beberapa kesulitan yang kami temui, misalnya adanya sejumlah entri penting yang tidak kami temukan pada KBBI 1991 namun kami temukan pada KBBI 1988 yaitu entri *cenderung* sampai dengan *cenduai*, entri *gelepot* sampai dengan entri *gelindung*. Selain itu kami juga menemukan adanya dua bentuk penulisan kata yang berbeda pada KBBI seperti *andaikata* dengan *andai kata*, *mengkup* dengan *mengekup*, *acap kali* dengan *acapkali*. Selain masalah yang kami temui dalam menggunakan KBBI, kami juga merasa enggan menggunakan KBBI

karena banyaknya kata yang harus kami teliti penulisannya telah menyita waktu kami untuk membuka-buka KBBI yang berukuran tebal.

Sebagai pengguna komputer, kami mulai menginginkan adanya suatu bentuk kamus elektronis sehingga informasi yang kami perlukan dapat tersedia dengan cepat. Keinginan ini makin bertambah setelah kami melihat adanya kamus elektronis *Webster English Dictionary* yang tersedia di tempat kami. Kami sangat berharap agar kamus bahasa Indonesia juga tersedia dalam bentuk elektronis.

Kamus elektronis akan memudahkan pemakai komputer mencari arti, bentuk dasar, pemenggalan, ataupun penggunaan suatu kata yang tepat. Bagi penulis atau penyunting dokumen yang telah terbiasa bekerja dengan komputer, keberadaan suatu kamus elektronis tersebut akan sangat dirasakan manfaatnya dalam mempercepat pekerjaan mereka. Mereka tidak perlu lagi membalik-balik halaman kamus untuk mencari informasi tentang suatu kata, tetapi tetap dapat berkonsentrasi pada layar monitornya mencari informasi yang diperlukan dari kamus elektronis. Kamus elektronis dapat menyediakan informasi yang ada dengan cepat melalui fasilitas peninjauan (*browsing*) yang memudahkan pengguna komputer untuk memperoleh informasi hanya yang diperlukannya. Adanya kemudahan dan kecepatan dalam mendapatkan informasi melalui komputer telah mendorong kami untuk mengembangkan kamus elektronis untuk Bahasa Indonesia.

2 Pengembangan Kamus Elektronis

2.1 Konversi Teks

Sebagai langkah awal pengembangan kamus elektronis, ketersediaan data teks dalam bentuk elektronis merupakan satu syarat yang mutlak. Dengan makin murah nya harga komputer, maka makin bertambah pula penggunaan komputer sebagai alat bantu untuk mempermudah penyelesaian pekerjaannya. Akibatnya makin banyak pula tersedia dokumen elektronis yang dihasilkan oleh perangkat lunak pemroses kata atau pengolah data lainnya. Namun bila data teks yang ada belum

berbentuk elektronik, seperti masih berupa tulisan dalam buku atau majalah, maka untuk menghindari pengetikan yang menyita banyak waktu dapat digunakan pemindai (*scanner*) yang disertai dengan perangkat lunak pengenalan karakter secara optis (*Optical Character Recognition*). Pemindai dapat mengubah bahan cetakan menjadi citra yang berisi teks sedangkan perangkat lunak pengenalan karakter dapat menerjemahkan citra teks menjadi data teks elektronik. Selanjutnya data elektronik dapat diolah sesuai dengan keperluannya.

Dalam mengembangkan kamus elektronik ini, kami bekerja sama dengan Pusat Pengembangan dan Pembinaan Bahasa sehingga KBBI akan dikembangkan menjadi kamus elektronik. Data teks KBBI yang tersedia telah berbentuk elektronik sehingga kami dapat langsung melakukan pemrosesan lebih lanjut.

2.2 Basis Data Tekstual

Data elektronik yang dihasilkan perangkat lunak pemroses kata hanyalah data teks yang dapat dibaca, sehingga untuk memperoleh informasi tertentu yang ada pada teks tersebut tidak dapat dilakukan dengan mudah. Untuk mengelola data teks yang ada maka diperlukan suatu sistem basis data yang menghimpun data-data yang ada sehingga penyimpanan dan pengambilan informasi yang diperlukan dapat dilakukan dengan mudah.

Pada basis data yang kita kembangkan data-data teks yang ada perlu dikelompokkan sesuai dengan kebutuhan pemakai data. Untuk mengelola data tekstual diperlukan suatu basis data yang dapat menangani bentuk teks karena sifat setiap data teks bisa mempunyai panjang, format, serta isi yang berbeda. Misalnya informasi mengenai definisi atau contoh lema ada yang sangat panjang namun juga ada yang sangat singkat. Pemakaian basis data yang terstruktur sangat memboroskan tempat penyimpanan karena tidak semua tempat digunakan semuanya. Oleh karena itu penggunaan basis data tekstual diperlukan untuk menangani data tekstual yang bentuknya tidak terstruktur. Sementara basis data tekstual tidak tersedia, penggunaan basis data terstruktur tetap kami lakukan untuk memenuhi kriteria pencarian informasi di atas.

Selain untuk keperluan pengolahan data, sistem basis data dapat memberi keuntungan lain yaitu menjaga konsistensi data yang ada. Saat ini data elektronik yang ada di Pusat Bahasa kemungkinan berbeda dengan yang dimiliki Balai Pustaka karena bahan yang telah diketik oleh beberapa orang di Pusat Bahasa telah diketik kembali di Balai Pustaka untuk keperluan pencetakan KBBI. Oleh karena itu terjadinya kesalahan pengetikan yang membingungkan pengguna KBBI dapat dimengerti. Dengan adanya satu sistem basis data kamus maka konsistensi data dapat dijaga, karena semua perubahan pada data yang ada hanya dilakukan pada satu sistem basis data.

Data teks yang ada perlu dikelompokkan sesuai dengan keperluan ahli bahasa, sehingga data elektronik tidak hanya bisa dibaca di layar monitor tetapi juga dapat diolah lebih lanjut. Data elektronik yang ada perlu dipilah-pilah terlebih dahulu ke dalam klasifikasi yang telah dibuat sesuai dengan kebutuhannya. Pemilahan data tekstual harus dilakukan dengan program yang dikembangkan untuk mengenali batas-batas bagian data tekstual.

Data tekstual dari KBBI yang berbentuk seperti contoh di bawah ini:

flam.bo.yan *n* 1 tumbuhan pelindung, berdaun rindang, bunganya indah berwarna merah; *Poinciana regia*; 2 ki gemerlapan; serba megah; *dia termasuk tokoh yg berpredikat -- di di dunia mode dan musik*

fla.mi.ngo *n* Zool burung yg hidup di daerah tropis (di Afrika dan Amerika Selatan), berkaki dan berleher panjang, berparuh lengkung, tepi paruhnya bagian dalam bergerigi spt paruh itik, berbulu putih dng sayap hitam yg pangkalnya ditutupi bulu halus berwarna merah anggur; *Phoenicopteridae*

fla.nel /flanel/ *n* kain panas yg berkabu-kabu

(Kamus Besar Bahasa Indonesia, h. 278)

perlu dipilah-pilah sesuai dengan penggolongan yang diperlukan. Data tekstual di atas dikelompokkan pada basis data yang telah dirancang seperti data di bawah ini:

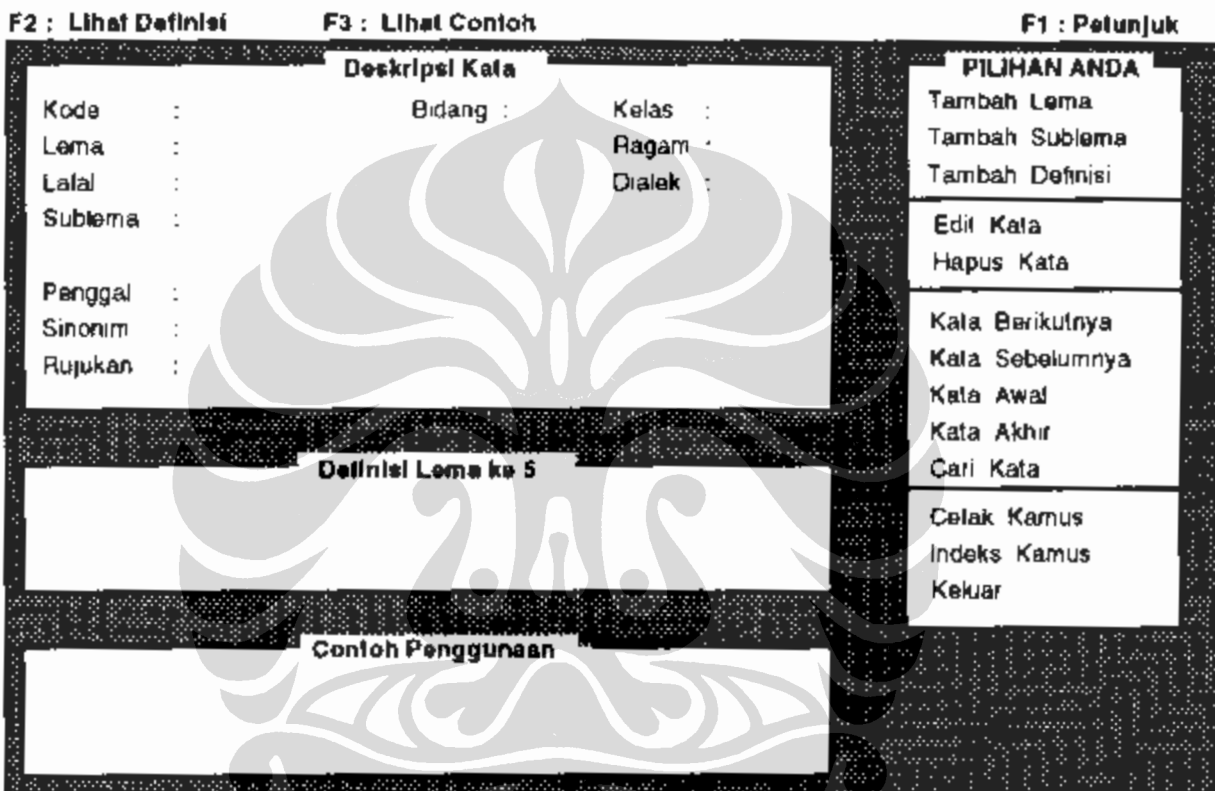
Lema	Penggalan	Lafal	Kelas Kata	...
flamboyan	flam.bo.yan		n	...
flamingo	fla.mi.ngo		n	...
flanel	fla.nel	/flanel/	n	...

Data yang telah tersusun pada basis data dapat ditampilkan sesuai dengan keinginan pemakai atau diolah lebih lanjut. Pemindahan data elektronik ke dalam basis data yang telah dirancang sesuai dengan keperluan akan mempermudah dan mempercepat perolehan informasi yang diinginkan, misalnya mencari lema yang mempunyai kelas kata *n*.

2.3 Penyuntingan Entri Pada Basis Data

Dengan program yang kami kembangkan, kami berusaha memilah-milah data elektronik yang ada sesuai dengan pengelompokan pada basis data. Pemilahan ini tidak mudah dilakukan karena ketidak konsistenan dalam pengetikan data KBBI telah menyebabkan data tekstual tidak dapat dikenali sesuai dengan penggolongan yang telah dibuat. Hasil dari proses pemilahan teks ini tentu saja tidak sempurna. Oleh sebab itu kami mengembangkan suatu fasilitas penyuntingan yang khusus dibuat untuk menyunting data-data KBBI yang telah dimasukkan ke dalam basis data.

Program penyunting data teks KBBI ini menyediakan fasilitas untuk mencari, menambah atau menghilangkan lema; melakukan perubahan data pada definisi, contoh, kelas kata dan sebagainya. Selain itu data KBBI yang ada dapat langsung dilihat di layar komputer atau dicetak sesuai dengan format yang terdapat pada KBBI.



Tahap penyuntingan ini merupakan tahap yang menyita waktu dan tenaga karena perlunya tenaga manusia untuk melakukan pemeriksaan data yang ada sesuai dengan isi KBBI, kemudian melakukan penyuntingan di komputer dengan program yang telah tersedia. Namun hasil dari tahap penyuntingan ini akan sangat bermanfaat untuk pemrosesan selanjutnya karena informasi yang benar dapat diolah menjadi data yang berguna.

2.4 Strategi Pelacakan Kata

Sebagai bagian dari pengembangan kamus elektronis, maka perolehan informasi yang telah tersimpan dengan baik pada basis data juga menjadi

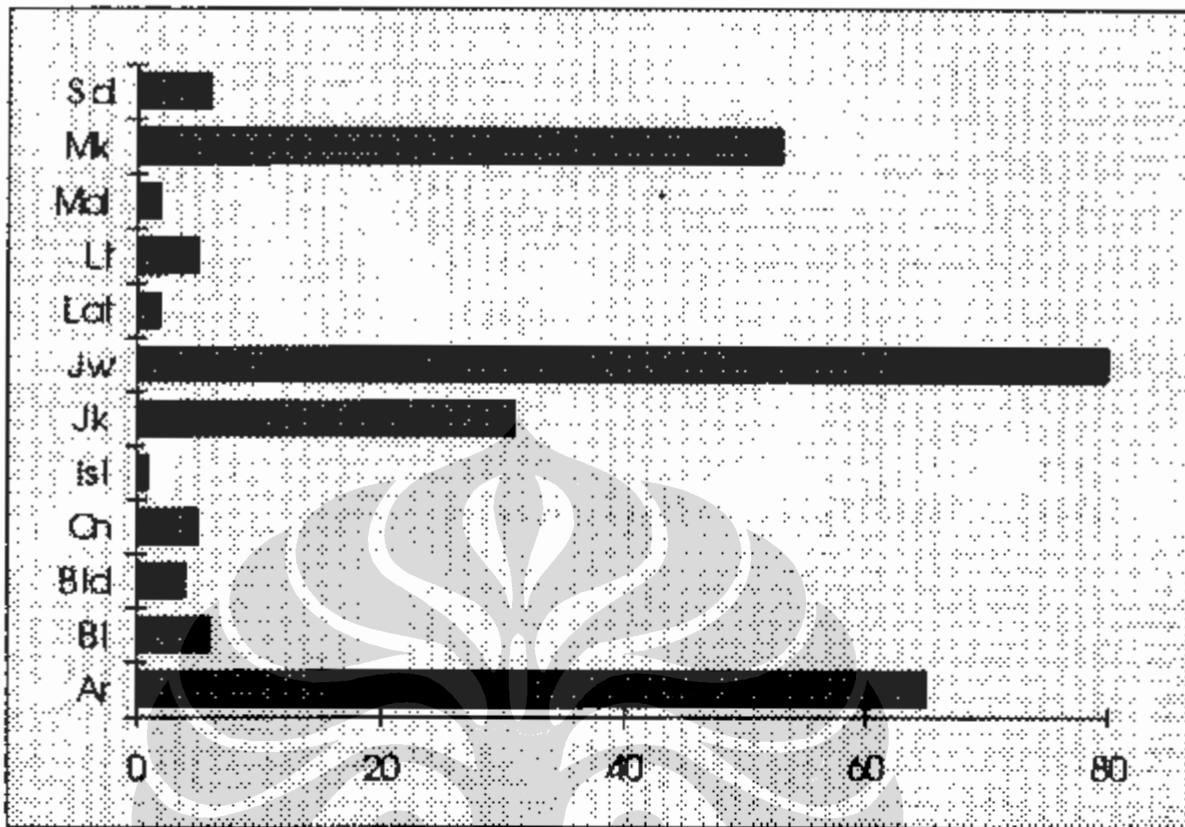
masalah yang penting. Besarnya data yang disimpan, contohnya KBBI mempunyai kurang lebih 72.000 lema beserta keterangan penting lainnya, menyebabkan perlunya suatu teknik yang memungkinkan perolehan informasi yang diperlukan secara cepat dan tepat. Oleh karena itu pencarian informasi tekstual merupakan topik yang menarik pada Ilmu Komputer. Teknik-teknik pencarian kata berdasarkan kesamaan kata, asal kata, tesaurus, dan lain sebagainya merupakan bagian dari upaya yang dilakukan untuk mencari suatu kata pada data yang berskala besar.

2.5 Analisis Statistik Terhadap Isi Kamus

Ahli bahasa tentunya memerlukan data linguistik yang ada pada kamus yang dikembangkannya. Setelah data kamus tersedia pada basis data maka pengolahan data dapat dilakukan dengan mudah dan cepat. Dengan bantuan program yang telah dikembangkan sesuai dengan keperluan ahli bahasa, maka informasi yang diinginkan dapat tersedia dengan cepat. Misalnya seorang ahli bahasa ingin mengetahui jumlah atau persentase dari

- lema atau sublema yang ada,
- lema yang mempunyai label kelas kata tertentu,
- lema yang mempunyai label bidang ilmu tertentu.

Berikut ini merupakan contoh data statistik dari lema yang mempunyai label dialek pada abjad A dari KBBI.



Data-data statistik ini dapat tersedia dalam sekejap sehingga mempersingkat pekerjaan penghitungan yang biasanya dilakukan oleh manusia.

3 Penyusunan Bank Kosa Kata Elektronik

Untuk menyusun kamus tentunya perlu mempunyai acuan yang luas sebagai dasarnya. Berbagai sumber pengetahuan yang akan digunakan untuk bahan penyusunan harus dikumpulkan sehingga hasilnya mencerminkan penggunaan kata yang ada di masyarakat atau lingkungan tertentu. Data-data yang akan digunakan sebagai bahan penyusunan kamus dapat dihimpun pada suatu basis data berskala besar yang dapat disebut sebagai bank kosa kata.

Bank kosa kata ini nantinya harus dapat merepresentasikan kata-kata yang digunakan oleh masyarakat pada suatu periode waktu tertentu baik yang digunakan dalam tulisan maupun dalam percakapan sehari-hari. Agar dapat menjadi representasi yang baik, data ini harus dapat mencakup kata-kata yang digunakan di media yang berbeda, oleh kalangan yang berbeda, dan pada lokasi yang berbeda pula. Luasnya cakupan yang diinginkan ini akan menuntut data dalam jumlah yang sangat besar. Selanjutnya, data yang dimasukkan ke dalam bank kosa kata tersebut tentunya harus diberi informasi yang lengkap seperti tempat penggunaannya, penulisnya, waktu, tempat, serta informasi linguistik yang diperlukan. Informasi ini kelak diperlukan pada saat pakar-pakar linguistik mengolahnya di kemudian hari.

Akibatnya bank kosa kata ini nantinya akan berisi kumpulan kata-kata dalam skala amat besar ditambah dengan informasi rinci untuk masing-masing kata. Volume data dalam skala yang besar menyebabkan penggunaan teknologi komputer tidak lagi dapat dihindarkan sehingga bank kosa kata harus berbentuk data elektronik. Bila pengolahan data seluruhnya dilakukan oleh manusia maka waktu yang diperlukan akan sangat lama sehingga mungkin tidak lagi dapat mengikuti perkembangan penggunaan kata-kata itu sendiri.

Untuk melakukan analisis penggunaan suatu kata, maka ahli bahasa dengan mudah dapat mencari kalimat-kalimat yang menggunakan kata tersebut. Data-data yang telah ditemukan dapat dianalisis sehingga penggunaan suatu kata dapat diketahui. Pemilihan kata yang akan dimuat pada kamus akan dengan mudah diidentifikasi dari bank kosa kata elektronik dengan mencari frekuensi pemunculannya serta dapat diketahui konteks penggunaan kata tersebut. Penggunaan kata dalam konteks yang berbeda akan dapat digunakan untuk menyusun definisi dan contoh dalam kamus sehingga penjelasan yang ada memang mencerminkan penggunaan kata tersebut saat ini. Selain itu data tersebut dapat digunakan untuk menghindari kesalahan makna serta mengetahui perkembangan kosa kata yang ada pada saat ini di masyarakat.

Sebagai contoh pengembangan bank kosa kata elektronik kami, Kelompok Penelitian Pemrosesan Teks - FIKUI dan Jurusan Linguistik FSUI, saat ini mempunyai bank kosa kata elektronik yang akan kami gunakan untuk penyusunan kamus elektronik untuk perguruan tinggi. Untuk mengetahui kosa kata pada berbagai bidang yang ada di perguruan tinggi maka kami mengumpulkan berbagai macam dokumen elektronik seperti tesis, makalah seminar, ataupun bahan pengajaran di perguruan tinggi. Dokumen elektronik tersebut kami simpan dalam bank kosa kata elektronik disertai dengan data-data penulisan seperti nama penulis, jenis tulisan, waktu penulisan dan sebagainya.

Bila penulisan-penulisan di waktu lalu dimasukkan juga ke bank kosa kata elektronik maka data yang ada dapat digunakan untuk menyusun etimologi suatu lema dengan mudah. Berdasarkan tahun penggunaan lema tersebut maka peneliti bahasa dengan cepat dapat mengetahui asal-usul suatu lema, konteks penggunaan lema pada suatu periode, ataupun perkembangan makna suatu lema.

Kritik terhadap KBBI yang dimuat pada harian Kompas menyebutkan bahwa KBBI mempunyai kekurangan yaitu mengenai banyaknya penjelasan lema yang keliru, masalah Jawanisasi, serta belum adanya etimologi. Kritikan yang dikemukakan ini dapat diatasi seandainya Pusat Pembinaan dan Pengembangan Bahasa telah mempunyai bank kosa kata elektronik. Penjelasan untuk suatu lema dapat disusun berdasarkan data penggunaan kata tersebut pada masyarakat. Lemma yang berasal dari bahasa Jawa dapat dihilangkan bila data penggunaannya hanya sedikit atau bila frekuensi penggunaannya banyak maka pemuatan lema tersebut pada KBBI dapat dijelaskan. Sedangkan etimologi juga dapat disusun bila data yang terdapat pada bank kosa kata elektronik mencakup kurun waktu yang lama.

Bank kosa kata elektronik juga memungkinkan pengembangan kamus-kamus yang lebih spesifik seperti misalnya kamus khusus untuk bidang Ilmu Komputer atau Sastra. Kamus khusus dengan cepat dapat dikembangkan melalui pemilihan lema dari bank kosa kata elektronik yang

digunakan pada konteks pendidikan tertentu. Berdasarkan penjelasan di atas dapat dilihat bahwa manfaat pengembangan bank kosa kata elektronik sangat besar bagi perkembangan Bahasa Indonesia.

4 Pengembangan Kamus Elektronik Lainnya

Salah satu subtema pada Kongres Bahasa Indonesia VI yaitu "Peningkatan Mutu Bahasa Indonesia Memperlancar Pengembangan Ilmu Pengetahuan Dan Teknologi" menyiratkan bahwa perhatian yang besar perlu diberikan kepada ilmu pengetahuan dan teknologi. Dengan makin pentingnya peran ilmu pengetahuan dan teknologi pada waktu-waktu mendatang, subtema di atas secara nyata menunjukkan perlunya peningkatan mutu Bahasa Indonesia untuk mendukung pengembangan ilmu pengetahuan dan teknologi di negeri kita.

Selain tersedianya kamus elektronik untuk Bahasa Indonesia, kami juga berpendapat bahwa kamus istilah perlu juga tersedia dalam bentuk elektronik. Cepatnya kemajuan ilmu pengetahuan dan teknologi menyebabkan banyaknya istilah asing yang digunakan untuk membahas topik baru di bidang tersebut karena mereka tidak mengetahui padanan kata yang tepat dalam bahasa Indonesia. Untuk mengurangi penggunaan istilah asing ini maka alangkah baiknya bila tersedia glosarium elektronik dalam pelbagai bidang sehingga masyarakat yang memerlukannya dapat dengan mudah dan cepat memperoleh padanan kata asing yang diperlukannya dalam Bahasa Indonesia. Saat ini kami telah mempunyai glosarium elektronik di bidang Ilmu Komputer walaupun kami masih memerlukan ahli bahasa untuk meneliti padanan kata Bahasa Indonesia yang telah dihasilkan.

Dengan tersedianya glosarium elektronik maka diharapkan dapat membantu penggunaan bahasa Indonesia di lingkungan pendidikan tinggi karena untuk menjelaskan suatu istilah telah dapat digunakan bahasa Indonesia yang baik dan benar.

5 Media Penyimpanan

Besarnya data teks seperti yang terdapat pada kamus tentunya memerlukan media penyimpanan berkapasitas besar. Untuk mengembangkan kamus elektronis yang dapat digunakan oleh masyarakat, maka pemilihan tempat penyimpanan data menjadi pemikiran selanjutnya. Data elektronis serta aplikasi perangkat lunak yang memerlukan tempat penyimpanan yang besar sudah bukan masalah yang perlu dirisaukan. Media penyimpanan berbentuk disket tampaknya akan segera disaingi oleh *CD-ROM* karena kemampuan penyimpanan datanya yang besar.

Data kamus elektronis yang berukuran besar sangatlah tepat bila disimpan dalam *CD-ROM* sehingga kamus tidak saja dapat dibeli dalam bentuk cetakan namun juga dalam bentuk elektronis. Data yang terdapat pada *CD-ROM* sifatnya hanya dapat dibaca oleh pengguna komputer tanpa dapat mengubahnya. Hal ini tentu saja baik bagi kamus elektronis karena kita tidak mengharapkan bahwa setiap orang dapat mengubah isi kamus yang ada.

6 Penutup

Untuk mendukung penggunaan Bahasa Indonesia sebagai alat komunikasi yang utama maka kami mengharapkan agar pengembangan kamus elektronis, glosarium elektronis, serta bank kosa kata elektronis dapat segera diwujudkan. Selanjutnya pengembangan kamus elektronis untuk lingkungan pendidikan juga dapat dihasilkan dan segera dimanfaatkan. Beragamnya bahasa daerah di tanah air kita juga akan dapat bertahan di masa yang akan datang seandainya kita dapat menyusun kamus bahasa daerah elektronis mulai sekarang. Teknologi komputer dapat juga digunakan untuk menyediakan informasi kebahasaan seperti kata-kata baru yang belum terdapat pada kamus ataupun glosarium sehingga masyarakat yang memerlukannya dapat memperoleh informasi dengan

cepat. Keprihatinan masyarakat terhadap perkembangan penggunaan bahasa Indonesia saat ini diharapkan dapat dikurangi dengan adanya tindakan pembinaan yang menggunakan kemajuan teknologi. Oleh karena itu kerja sama antara bidang komputer dan bahasa sangat diperlukan untuk mewujudkan pembinaan serta pengembangan Bahasa Indonesia melalui teknologi komputer.

7 Bahan Bacaan

Adriani, Mirna dan Bobby Nazief. "Menyempurnakan KBBI Dengan Teknologi Komputer." *Kompas*, 2 Des. 1994, h. 4.

Dewire, Dawna Travis. *Text Management*. New York: McGraw-Hill, Inc., 1994.

Nazief, Bobby dan Mirna Adriani. "Pemanfaatan Teknologi Informasi Dalam Pengembangan Bahasa Indonesia." *Suara Pembaruan*, 6 Des. 1993, h. 10.

"Politik Kebahasaan Kita, Tiga Kritik Terhadap KBBI, dan Kehadiran Pusat Bahasa." *Kompas*, 24 Okt. 1994, h. 9.

Pusat Pembinaan Dan Pengembangan Bahasa. *Laporan Komputerisasi Perkamusan Tahap I: Proyek Penelitian Dan Pembinaan Bahasa Dan Sastra Indonesia Dan Duerah*. Jakarta: 1993.

Tim Penyusun Kamus, Pusat Pembinaan Dan Pengembangan Bahasa. *Kamus Besar Bahasa Indonesia*. ed. 2. Jakarta: Balai Pustaka, 1993.

Raymond, Darrell R. dan Frank WM. Tompa. "Hypertext And The Oxford English Dictionary". *Communications of the ACM*. 31 (July 1988). 871-879.