

Development of English-Indonesia Idiom Mobile Dictionary Using Optical Character Recognition

Robby Kurniawan Budhi¹, Dwi Taufik Hidayat², Jonathan Putra Pangapul³

¹Fakultas Teknik Universitas Widya Kartika
Surabaya, Indonesia, e-mail: robbi@widyakartika.ac.id

² Fakultas Teknik Universitas Widya Kartika
Surabaya, Indonesia, e-mail: dwitaufighidayat@widyakartika.ac.id

³ Fakultas Teknik Universitas Widya Kartika
Surabaya, Indonesia, e-mail: jonathanpangapul@gmail.com

ARTICLE INFO

Article history:

Received 23 July 2018

Received in revised form 30 July 2018

Accepted 3 Agustus 2018

Available online 17 Agustus 2018

ABSTRACT

Idioms are phrases that often have different meanings with their literal meaning. The combination of words used, will have different meaning if translated one by one, especially for English words. Currently there are many idiom dictionary applications that exist, but the users have to type the words as the input manually. This requires more precision and time, especially for users who are less fluent in English. To speed up the search process and reduce errors that may occur, optical character recognition methods are used in this study. Through testing and observation to the user, it can be concluded that this application is able to help users to find translations of idioms in English more quickly and precisely.

Keywords: idiom, dictionary, optical character recognition

1. Pendahuluan

Bahasa merupakan salah satu cara manusia dalam berkomunikasi. Selain kata-kata yang memiliki arti harafiah, seringkali ditemukan penggunaan frase atau kumpulan kata yang apabila diartikan kata per kata menjadi tidak sesuai pengertiannya. Frase tersebut biasanya disebut sebagai ungkapan atau idiom. Menurut Kamus Besar Bahasa Indonesia, kata idiom memiliki arti konstruksi yang maknanya tidak sama dengan gabungan makna unsurnya [1].

Beberapa contoh idiom dalam bahasa Indonesia misalnya “membanting tulang” yang artinya bekerja keras, “meja hijau” yang berarti pengadilan, dan “gulung tikar” yang artinya bangkrut. Ungkapan tersebut apabila diartikan per kata, akan memiliki arti yang berbeda. “Meja hijau” bukan berarti meja yang berwarna hijau. Ungkapan ini berdasarkan analogi bahwa pada ruang pengadilan, meja yang digunakan berwarna hijau. Demikian juga untuk “gulung tikar” yang bukan berarti kegiatan menggulung tikar, namun digunakan untuk mengungkapkan kegiatan seperti menutup usaha.

Ungkapan atau idiom juga terdapat dalam bahasa Inggris. Ungkapan “*cross your finger*” bukan berarti secara harafiah kegiatan untuk menyilangkan jari, namun memiliki arti tanda untuk mengharapkan keberuntungan. Ungkapan “*feel blue*” jika diartikan kata per kata berarti “merasa biru”, tapi yang dimaksudkan adalah perasaan sedih.

Bagi masyarakat yang masih awam terhadap penggunaan idiom, hal tersebut tentunya lebih sulit untuk dipahami. Umumnya untuk mencari arti kata dalam bahasa Inggris, kata per kata akan diterjemahkan satu per satu. Apabila hal ini dilakukan untuk idiom, maka artinya berbeda. Kamus idiom merupakan salah satu alat bantu yang dapat digunakan untuk mengetahui arti ungkapan dalam bahasa Inggris tersebut.

Saat ini, terdapat banyak buku maupun aplikasi kamus idiom. Penggunaan buku kamus dinilai kurang praktis karena umumnya ukurannya tebal dan berat. Sedangkan untuk penggunaan aplikasi, umumnya untuk mencari arti idiom tertentu, pengguna dapat mencari dengan memasukkan kata yang diperlukan. Bagi idiom yang tidak terlalu panjang, hal ini dapat dilakukan dengan mudah. Namun apabila idiom yang dimasukkan terdiri dari beberapa kata, hal ini kurang praktis dan kemungkinan terjadinya kesalahan inputan kata lebih besar, terutama bagi masyarakat yang kurang fasih dalam berbahasa Inggris. Idiom seperti “*don't count your chickens until they hatch*” atau “*put wool over other people's eyes*” membutuhkan waktu dan ketelitian dalam pengetikan kata. Sedangkan penggunaan Google Translate tidak dapat mengenali idiom tersebut dengan baik.

Kemampuan telepon genggam yang saat ini berbasis Android dan memiliki fitur kamera mampu dioptimalkan untuk mengatasi masalah tersebut. Pengenalan karakter optik (*Optical Character Recognition (OCR)*) merupakan sebuah aplikasi yang digunakan untuk mengenali teks dari gambar yang dideteksi, baik berupa hasil pindai maupun dari pengambilan gambar langsung melalui kamera [2]. Teknologi tersebut digunakan untuk memasukkan kata yang dicari artinya dalam penelitian ini.

Melalui penelitian ini, dihasilkan aplikasi yang mampu mencari arti terjemahan kalimat idiom bahasa Inggris ke dalam bahasa Indonesia secara lebih praktis dan lebih cepat.

2. Metode Penelitian

2.1. Metode Pengumpulan Data

Data yang digunakan merupakan kata-kata idiom yang diambil dari berbagai sumber untuk mengatasi keterbaruan data yang sering digunakan saat ini melalui studi literatur. Sedangkan untuk data pengukuran umpan balik dari pengguna digunakan kuesioner yang kemudian diolah secara statistik.

2.2. Jenis Data

Data yang diolah berupa gambar yang diperoleh langsung melalui kamera maupun hasil pemindaian. Untuk data pembandingan, digunakan data kata-kata idiom yang sebelumnya telah dikumpulkan sebagai basis data kamus idiom. Sedangkan untuk pengukuran hasil, digunakan data hasil survey kepada pengguna untuk mengetahui umpan balik terhadap sistem yang dibuat.

2.3. Metode Pengembangan Sistem

Pengembangan sistem dalam penelitian ini menggunakan System Development Life Cycle dengan tahapan meliputi [3]:

- a. Analisa Kebutuhan: berupa pengumpulan data yang diperlukan dalam pengembangan sistem, beserta proses menentukan kebutuhan perangkat keras, perangkat lunak, dan pengguna. Termasuk dalam proses ini, studi terhadap penelitian sejenis yang sebelumnya pernah tersedia. Tahap ini juga akan menentukan alur sistem yang dibuat berdasarkan tahapan proses *Optical Character Recognition (OCR)* yang sudah ada. Hasil pengumpulan data tersebut kemudian dianalisis untuk selanjutnya digunakan dalam tahapan berikutnya yaitu perancangan sistem.
- b. Perancangan Sistem: berdasarkan hasil pengamatan yang dilakukan pada tahapan pengumpulan data, serta hasil analisisnya, disusun suatu rancangan sistem yang digunakan sebagai acuan dalam pengembangan sistem. Rancangan sistem yang dibuat meliputi alur kerja sistem, perancangan basis data yang diperlukan, serta perancangan antar muka pengguna. Termasuk dalam alur kerja sistem adalah prinsip kerja OCR yang diterapkan ke dalam aplikasi, serta pencarian informasi terjemahan kata yang terdapat dalam basis data kamus idiom.
- c. Implementasi Sistem: tahapan penerapan desain yang sudah dibuat sebelumnya ke dalam program. Pada penelitian ini, aplikasi dibuat menggunakan Android Studio.

- d. Pengujian Sistem: yaitu tahapan untuk menguji sistem yang telah terbangun untuk mengetahui kemungkinan terjadinya kesalahan alur kerja sistem, serta mengukur respon pengguna. Pengujian internal dilakukan oleh Tim, sedangkan pengujian eksternal dilakukan terhadap pengguna umum.

3. Hasil dan Pembahasan

Berdasarkan hasil analisa kebutuhan, pada sistem yang akan dibangun akan menampilkan beberapa informasi yang disajikan ke dalam beberapa menu sebagai berikut:

- a. Pencarian terjemahan kata idiom berdasarkan masukan melalui pengetikan. Fasilitas ini tetap diberikan bagi pengguna yang ingin melakukan pencarian dengan mengetikkan kata idiom yang dicari melalui papan ketik.
- b. Pencarian berdasarkan inputan berupa foto yang diambil dari kamera. Fasilitas ini merupakan fitur utama dari sistem yang dibuat ini. Teks langsung diambil melalui kamera dan akan dilakukan pendeteksian kata serta pencarian terjemahannya dalam bahasa Indonesia apabila kata tersebut dikenali sebagai idiom.
- c. Pencarian berdasarkan gambar yang sudah tersimpan di memori. Hampir sama dengan fitur sebelumnya, namun inputan gambar dibedakan, karena yang diproses pada bagian ini adalah gambar yang sudah tersimpan sebelumnya.

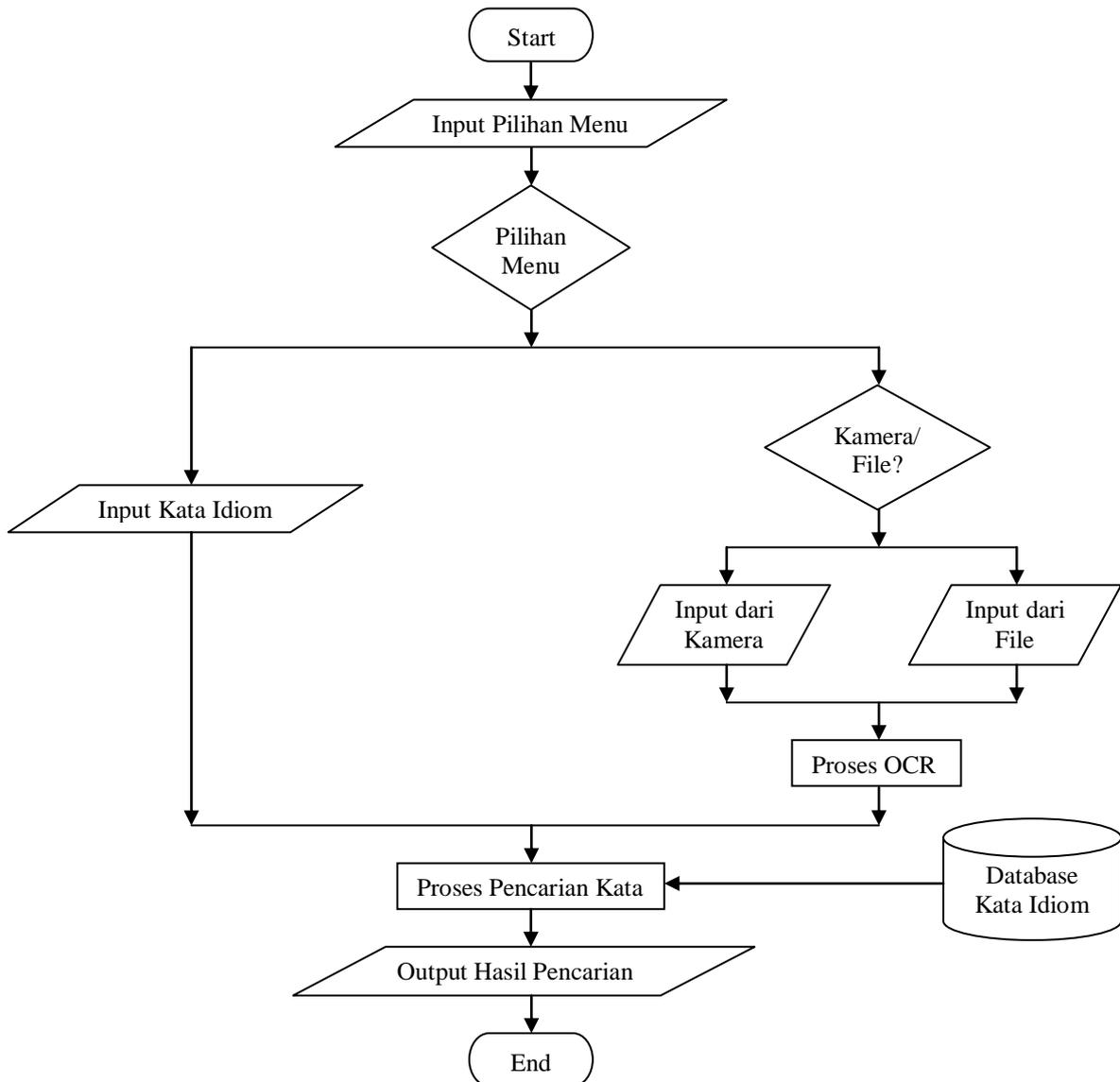
3.1. Alur Sistem

Pada bagian ini akan dijelaskan alur kerja sistem yang telah dibuat ke dalam flowchart sistem seperti pada gambar 1. Pada saat pengguna menjalankan aplikasi Kamus Idiom, akan ditampilkan pilihan menu. Menu ini digunakan untuk memberikan pilihan kepada pengguna tentang cara pencarian Kata Idiom yang dibutuhkan.

Pilihan menu pertama yaitu dengan memasukkan huruf awal serta kata idiom yang diperlukan melalui papan ketik. Cara ini tetap diberikan untuk memberikan fasilitas pencarian bagi pengguna yang ingin memasukkan kata idiom yang ingin dicari artinya melalui ketikan langsung. Ketika memasukkan huruf awal, pengguna kemudian akan dapat memilih kata idiom yang huruf depannya adalah huruf yang sudah dipilih tadi. Selanjutnya pengguna dapat melanjutkan pencarian dengan memasukkan kata idiom melalui papan ketik. Melalui pilihan menu ini, tidak diperlukan proses OCR karena yang dilakukan adalah pencarian langsung berdasarkan masukan yang kemudian akan dicocokkan dengan basis data yang tersedia.

Pilihan menu kedua adalah apabila pengguna ingin mencari kata idiom dari gambar. Pada pilihan ini, terdapat dua alternatif yang dimungkinkan sebagai masukan gambar, yaitu melalui hasil foto dari kamera secara langsung, ataupun dengan mencari file gambar yang sebelumnya sudah tersimpan di dalam memori. Apabila pengguna memasukkan gambar dari kamera, maka kamera piranti akan diaktifkan untuk dapat menangkap gambar. Sedangkan apabila ingin mencari dari file, pengguna dapat memilih file yang diinginkan dari folder yang sesuai di dalam kartu memori atau memori internal dari piranti.

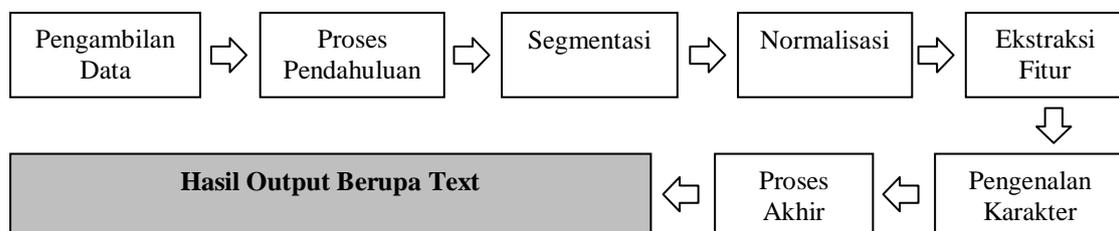
Pada pilihan kedua inilah, proses OCR akan dijalankan. Secara garis besar, proses OCR yang dilakukan meliputi beberapa tahap, yaitu penerimaan masukan, *preprocessing*, segmentasi, normalisasi, dan pengenalan. Gambar yang didapatkan melalui kamera ataupun hasil dari file yang sudah ada, akan diproses pada tahap *preprocessing*. Tahap ini digunakan untuk menyamakan kualitas dan warna dari gambar yang akan diproses. Gambar yang dibutuhkan adalah gambar *grayscale*. Gambar tersebut kemudian akan disegmentasi untuk mengenali area mana yang terdeteksi huruf ataupun kata. Selanjutnya karakter tersebut akan dinormalisasi agar ketebalan garis dapat diseragamkan. Proses selanjutnya adalah pengenalan karakter berdasarkan basis data yang ada. Kemudian dari huruf yang terbaca tersebut, dirangkai kata dengan memperhatikan spasi antar kata. Hasil proses inilah yang nantinya digunakan untuk pencarian dalam basis data kata idiom. Apabila terdapat kesamaan, hasil pencarian arti akan ditampilkan. Sedangkan apabila tidak dikenali dalam basis data, kata tersebut dianggap bukan idiom.



Gambar 1. Flowchart sistem aplikasi kamus idiom

3.2. Proses *Optical Character Recognition* (OCR)

Setelah sebelumnya dijelaskan secara singkat, proses OCR[4] secara detail ditampilkan pada gambar 2 berikut.

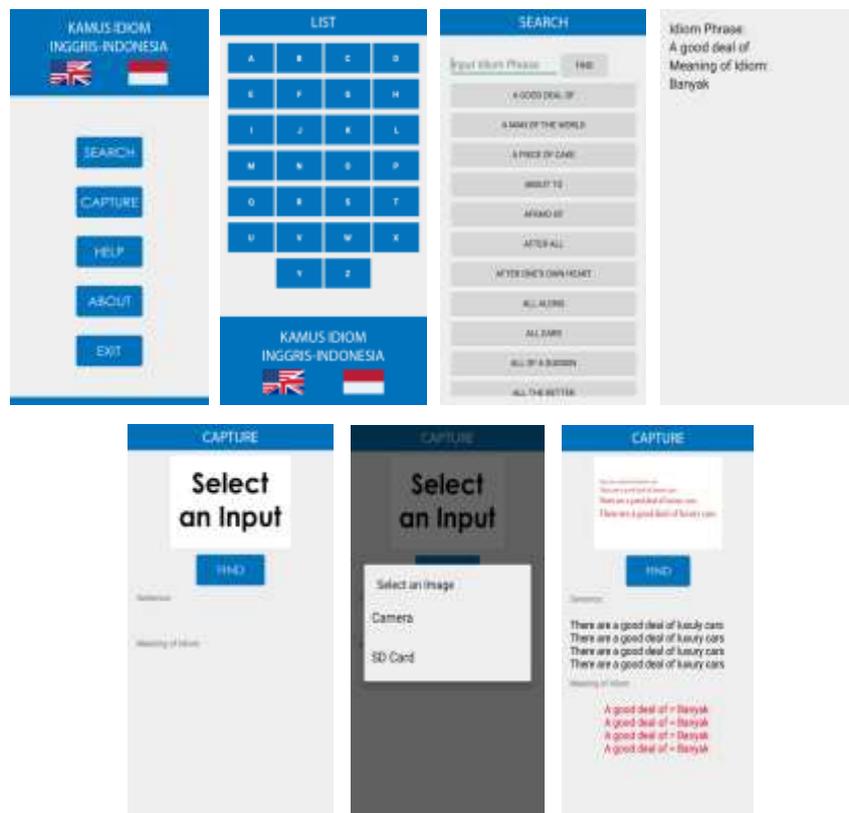


Gambar 2. Proses *Optical Character Recognition*

- a. Tahap pengambilan data merupakan tahapan untuk mendapatkan gambar yang akan diproses.
- b. Proses pendahuluan meliputi beberapa proses yaitu:
 - *Greyscale* dan *thresholding*, yaitu perubahan warna gambar dari RGB menjadi gambar keabuan, kemudian dilakukan langkah pengambungan (*thresholding*) untuk mengubah gambar menjadi biner (*black and white*)[5].
 - *Smoothing*, digunakan untuk mengurangi derau (*noise*).
 - *Scaling*, proses perubahan ukuran gambar agar tidak terlalu membebani proses.
 - *Stroke thinning*, merupakan proses untuk menandai garis batas karakter.
- c. Tahap Segmentasi, untuk membatasi area gambar yang akan diproses.
- d. Tahap Normalisasi, untuk mengubah ketebalan karakter yang dideteksi.
- e. Tahap Ekstraksi Fitur, merupakan proses pengubahan data menjadi kumpulan fitur yang mendeskripsikan properti gambar yang dianalisa. Termasuk dalam proses ini adalah untuk mengetahui ukuran rata-rata tinggi karakter serta penggunaan garis tulisan untuk mengenali karakter.
- f. Tahap Pengenalan Karakter, merupakan tahapan perbandingan antara hasil karakter yang terdeteksi dengan basis data karakter yang sudah tersimpan.
- g. Proses Akhir, umumnya untuk koreksi ejaan dari kata yang terdeteksi.
 Hasil dari keseluruhan proses ini adalah teks yang kemudian dapat diproses untuk pencarian arti kata idiom dari basis data kamus idiom yang tersedia.

3.3. Hasil Implementasi

Berdasarkan hasil analisa dan perancangan pada proses sebelumnya, maka dibangunlah aplikasi dengan tampilan seperti pada gambar 3 berikut.



Gambar 3. Tampilan Aplikasi Kamus Idiom

3.4. Hasil Pengujian dan Evaluasi

Proses setelah terbentuknya aplikasi adalah melakukan tahapan pengujian dan evaluasi. Menu “Search” digunakan untuk pengguna yang ingin memasukkan kata idiom yang dicari melalui papan ketik. Akan ditampilkan daftar huruf yang mewakili kelompok kata idiom yang diawali oleh huruf tersebut. Pengguna kemudian dapat memilih huruf yang sesuai, dan dapat melakukan pencarian kata idiom dengan memasukkan kata yang sesuai, yang setelah itu akan ditampilkan artinya.

Pada menu “Capture”, pengguna memilih gambar sebagai masukan. Gambar dapat diperoleh melalui kamera atau mengambil dari memori piranti. Setelah gambar terakses, aplikasi akan menganalisa dan berusaha mengenali kata idiom yang terdapat pada gambar tersebut. Apabila kata tersebut terdapat dalam basis data kamus, maka akan ditampilkan artinya. Pada menu inilah proses OCR berjalan.

Proses uji coba dilakukan dengan menggunakan kalimat “*there are a good deal of luxury cars*” dengan kata idiom yang seharusnya terdeteksi “*a good deal of*” yang memiliki arti “banyak”. Digunakan beberapa variasi tipe huruf (*font*), ukuran huruf, warna huruf, serta warna latar belakang. Hasil uji coba apabila masukan berupa file gambar dari memori piranti, ditampilkan pada tabel 1 berikut.

Tabel 1. Hasil Uji Coba Aplikasi Kamus Idiom dengan Masukan File dari Memori Piranti

Tipe Huruf	Ukuran	Warna Teks	Latar Belakang			
			Putih	Merah	Hijau	Biru
Times New Roman	6	Hitam	√	√	√	√
	8		√	√	√	√
	10		√	√	√	√
	12		√	√	√	√
Arial	6		√	√	√	√
	8		√	√	√	√
	10		√	√	√	√
	12		√	√	√	√
Calibri	6		√	√	√	√
	8		√	√	√	√
	10		√	√	√	√
	12		√	√	√	√
Comic Sans	6		√	√	√	√
	8		√	√	√	√
	10		√	√	√	√
	12		√	√	√	√

Keterangan: √ = berhasil, X = gagal

Seluruh hasil uji coba dengan masukan berupa file dari memori adalah berhasil. Aplikasi dapat mendeteksi dengan baik, kata idiom yang terdapat pada gambar, dan memberikan arti yang sesuai.

Pada uji coba dengan masukan dari kamera, langkah pengujian sedikit berbeda, dengan menambahkan ukuran huruf yang lebih besar serta warna font yang berbeda. Hasil uji coba ditampilkan pada tabel 2 berikut.

Tabel 2. Hasil Uji Coba Aplikasi Kamus Idiom dengan Masukan File dari Kamera

Tipe Huruf	Ukuran	Warna Latar Belakang	Warna Huruf			
			Hitam	Merah	Hijau	Biru
Times New Roman	6	Putih	X	X	X	X
	8		√	√	X	√
	10		√	√	√	√
	12		√	√	√	√
	14		√	√	√	√
Arial	6		√	√	X	√
	8		√	√	√	√
	10		√	√	√	√
	12		√	√	√	√
	14		√	√	√	√
Calibri	6		√	√	√	√
	8		√	√	√	√
	10		√	√	√	√
	12		√	√	√	√
	14		√	√	√	√
Comic Sans	6		√	√	√	X
	8		√	√	√	X
	10		√	√	√	√
	12		√	√	√	√
	14		√	√	√	√

Keterangan: √ = berhasil, X = gagal

Pada hasil uji coba menggunakan kamera, terdapat beberapa kegagalan. Saat menggunakan jenis huruf Times New Roman ukuran 6 (enam), aplikasi tidak dapat mendeteksi dengan baik. Sedangkan pada jenis huruf yang lain, hal tersebut dapat diatasi. Hal ini disebabkan oleh gagalnya aplikasi mengenali karakter huruf Serif yang memiliki sirip di tiap hurufnya apabila ukurannya terlalu kecil. Hasil pengambilan gambar dari kamera juga kurang optimal untuk ukuran huruf yang terlalu kecil ini. Selain itu, faktor derau (*noise*) pada gambar juga mempengaruhi hasil pendeteksian karakter.

Uji coba juga dilakukan untuk mengukur waktu yang digunakan oleh pengguna dalam menemukan arti kata idiom yang dibutuhkan. Perbandingan antara penggunaan buku kamus idiom dan aplikasi, menunjukkan bahwa aplikasi dapat lebih cepat menemukan arti kata idiom tersebut.

4. Kesimpulan

Berdasarkan hasil uji coba, aplikasi kamus idiom ini dapat mengenali frase idiom yang terdapat pada kalimat dengan baik. Fitur pengenalan frase melalui gambar juga dapat berjalan dengan baik, hanya dengan sedikit catatan untuk jenis huruf Serif yang kurang terdeteksi dengan baik apabila ukuran huruf terlalu kecil. Selain itu, aplikasi juga membantu pengguna untuk dapat lebih cepat menemukan arti frase idiom yang dibutuhkan.

5. Acknowledgement

Penelitian ini didanai oleh Kementerian Riset, Teknologi, dan Pendidikan Tinggi melalui skema Hibah Penelitian Dosen Pemula tahun pelaksanaan 2018.

References

- [1] Depdiknas PB. Kamus besar bahasa Indonesia. Edisi Ketiga. Jakarta: Balai Pustaka. 2002.
- [2] Mithe R, Indalkar S, Divekar N. Optical Character Recognition. International Journal of Recent Technology and Engineering (IJRTE). 2013 Mar;2(1):72-5.
- [3] Pressman RS. Software engineering: a practitioner's approach. Palgrave Macmillan; 2005.
- [4] Mollah AF, Majumder N, Basu S, Nasipuri M. Design of an optical character recognition system for camera-based handheld devices. arXiv preprint arXiv:1109.3317. 2011 Sep 15.
- [5] Pangestu P. Penerapan Histogram Equalization pada Optical Character Recognition Preprocessing. ULTIMATICS. 2015 Jun 1;7(1).