

Pengelompokan Kata Berdasarkan Kemiripan Ucapan Pada Kamus Menggunakan Algoritma Metaphone Pada Sistem Operasi Android

Aditya Dwi Maryanto^{*1}, Yuda Munarko², Yufis Azhar³

^{1,2,3}Teknik Informatika Universitas Muhammadiyah Malang
rhaditya05@gmail.com^{*1}, yuda@umm.ac.id², yufis@umm.ac.id³

Abstrak

Indonesia merupakan negara kepulauan yang terdiri dari berbagai macam suku dan budaya serta bahasa yang berbeda-beda. Salah satunya adalah bahasa Sumbawa yang digunakan oleh masyarakat bagian barat pulau Sumbawa. Bahasa Sumbawa terdapat 4 dialek yaitu Samawa, Taliwang, Jereweh, dan Tongo. Dalam bahasa Sumbawa juga terdapat kata-kata yang bersifat homofon, homonim, dan homograf. Pengelompokan kata berdasarkan kemiripan ucapan pada kamus Sumbawa dengan algoritma Metaphone pada operating sistem android menjadi pembahasan pada penelitian ini. Algoritma Metaphone dapat diterapkan pada berbagai bahasa dengan aturan yang telah dimodifikasi sesuai dengan karakteristik bahasa yang diinginkan. Berdasarkan hasil pengujian recall dan precision pada aturan baru algoritma Metaphone dapat disimpulkan bahwa pengelompokan kata berdasarkan kemiripan ucapan dalam bahasa Sumbawa dapat dikatakan efektif. Indeks persentase rata-rata pengujian recall 98,97% dan precision 78,52%.

Kata Kunci: Sumbawa, Kamus, Recall, Precision, Algoritma Metaphone

Abstract

Indonesia is an archipelagic country consisting of various tribes and cultures and different languages. One of them is the Sumbawa language used by the people of the western part of the island of Sumbawa. Sumbawa language has 4 dialects namely Samawa, Taliwang, Jereweh, and Tongo. In Sumbawa there are also homophonic, homonim, and homograph words. Grouping words based on the similarity of speech to the Sumbawa dictionary used Metaphone algorithm on the android operating system is discussed in this study. Metaphone algorithms can be applied to multiple languages with rules that have been modified according to the desired language characteristic. Based on the results of testing the recall and precision on the new rules Metaphone algorithm can be concluded that the grouping of words based on similarity of speech in Sumbawa language can be said to be effective. The average percentage index of the recall test is 98,97% and the precision is 78,52%.

Keywords: Sumbawa, Dictionary, Recall, Precision, Metaphone Algorithm

1. Pendahuluan

Indonesia merupakan negara kepulauan yang terdiri dari berbagai macam suku dan budaya. Setiap suku bangsa di Indonesia memiliki bahasa daerah masing-masing. Salah satu bahasa daerah yang dimaksud adalah bahasa Sumbawa. Seperti kebanyakan bahasa daerah lainnya, bahasa Sumbawa juga mempunyai fungsi-fungsi sebagai lambang kebanggaan daerah, sebagai lambang identitas daerah [1]. Dan sebagai sarana komunikasi dalam masyarakat daerah [2]. Bahasa Sumbawa merupakan bahasa yang dituturkan di wilayah bagian barat pulau Sumbawa, Indonesia. Bahasa Sumbawa merupakan anggota kelompok bahasa Melayu-Polinesia Barat, family Austronesia [3]. Bahasa Sumbawa, dikenal memiliki beberapa dialek atau variasi bahasa berdasarkan daerah penyebarannya yaitu Samawa, Taliwang, Jereweh, dan Tongo [4]. Dalam bahasa Sumbawa juga terdapat kata-kata yang bersifat homofon, homonim, dan homograf. Sehingga ketika melakukan pengetikan kata yang bersifat homofon terkadang terjadi kesalahan kata yang dimaksud dikarenakan sebagian besar masyarakatnya tidak mengenal sistem tulis dari bahasa mereka sendiri [5].

Untuk menyelesaikan masalah pengelompokan kata homofon (kata yang memiliki ejaan dan

makna yang berbeda tetapi memiliki lafal yang sama) serta meminimalisir keasalahan pengetikan, dapat diimplementasikan algoritma seperti Metaphone. Algoritma metaphone merupakan algoritma pencocokan string berdasarkan kemiripan ucapan. Algoritma ini dapat diterapkan pada berbagai bahasa [6]. Karena setiap bahasa memiliki caranya masing-masing dalam pengucapannya, sehingga algoritma ini harus dimodifikasi sesuai dengan karakteristik bahasa yang diinginkan. Adapun tujuan dari penelitian ini adalah membuat sebuah aplikasi pengelompokkan kata Sumbawa berdasarkan kemiripan ucapan pada kamus menggunakan algoritma Metaphone pada operating sistem android.

2. Metodologi Penelitian

2.1. Pengumpulan Data

Data dalam tugas akhir ini bersumber dari Kamus Sumbawa – Indonesia yang diterbitkan oleh Departemen Pendidikan Nasional Pusat Bahasa Provinsi Nusa Tenggara Barat tahun 2009. Referensi lainnya didapat dari jurnal, karya tulis ilmiah, media cetak atau internet

2.2. Analisa Kebutuhan

Aplikasi yang dibangun akan memenuhi kebutuhan fungsional yaitu sistem mampu menampilkan kata yang diinginkan oleh pengguna, menampilkan detil kata, dan menampilkan pilihan algoritma pencarian dari aplikasi. Sedangkan kebutuhan non-fungsionalitas yaitu system dapat berjalan pada smartphone android versi 4.0 dan seterusnya.

2.3. Perancangan Sistem

Perancangan perangkat lunak dilakukan setelah semua kebutuhan sistem didapatkan melalui tahap analisis kebutuhan. Perancangan sistem yang dilakukan meliputi beberapa tahap yaitu perancangan perancangan database, perancangan diagram *activity* yang digunakan untuk merancang alur atau aktifitas yang dilakukan oleh pengguna aplikasi, perancangan antarmuka aplikasi, dan perancangan aturan baru algoritma metaphone untuk mengelompokkan kata sumbawa.

2.4. Implementasi

Tahap ini seluruh analisis dan perancangan sistem akan diimplementasikan. Implementasi perangkat lunak mengacu kepada perancangan perangkat lunak. Implementasi perangkat lunak dilakukan dengan menggunakan Anroid Studio sebagai code editor dan DB Browser for Sqlite untuk penyimpanan data.

2.5. Pengujian dan Analisis

Pengujian perangkat lunak dilakukan untuk mengetahui apakah kinerja dan performa aplikasi telah memenuhi spesifikasi kebutuhan pada tahap sebelumnya dan untuk mengetahui keefektifan dari aturan baru metaphone yang telah disesuaikan. Pengujian perangkat lunak yang akan dilakukan yaitu pengujian fungsional dengan menggunakan metode *black-box testing*. Sedangkan pengujian untuk algoritma metaphone yaitu dengan menghitung nilai rata-rata *recall* dan *precision* yang diperoleh.

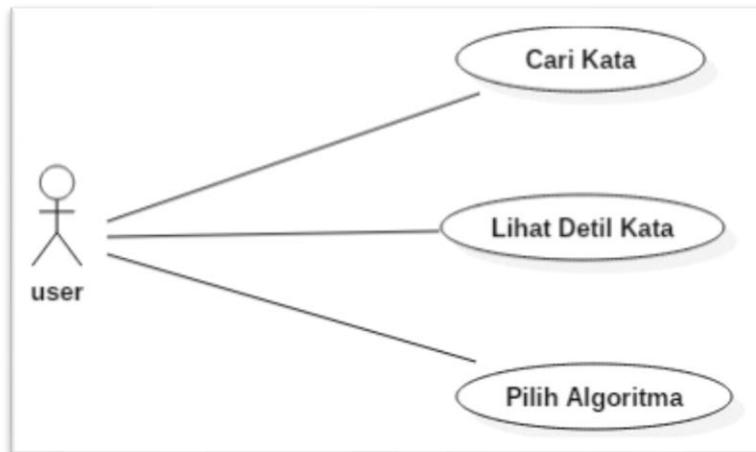
3. Analisis dan Perancangan

3.1. Analisa Masalah

Bahasa Sumbawa merupakan bahasa yang dituturkan di wilayah bagian barat pulau Sumbawa, Indonesia. Dalam bahasa Sumbawa, dikenal beberapa dialek atau variasi bahasa berdasarkan daerah penyebarannya. Bahasa Sumbawa mempunyai kata-kata yang homofon, homograf, dan homonim. Sebagian besar masyarakat penutur dari bahasa Sumbawa tidak mengenal sistem tulis dari bahasa mereka sendiri, itu disebabkan karena bahasa Sumbawa diturunkan melalui tradisi lisan [5]. Sehingga saat melakukan pencarian kata yang homofon terkadang terjadi kesalahan. Maka dari itu dalam Tugas Akhir ini akan menggunakan algoritma metaphone untuk mempermudah dalam mencari atau mengelompokkan kata yang bersifat homofon. Hal ini tentu saja membuat proses penyajian informasi menjadi lebih cepat dan efisien daripada mencari secara manual di buku kamus.

3.2. Analisa Kebutuhan

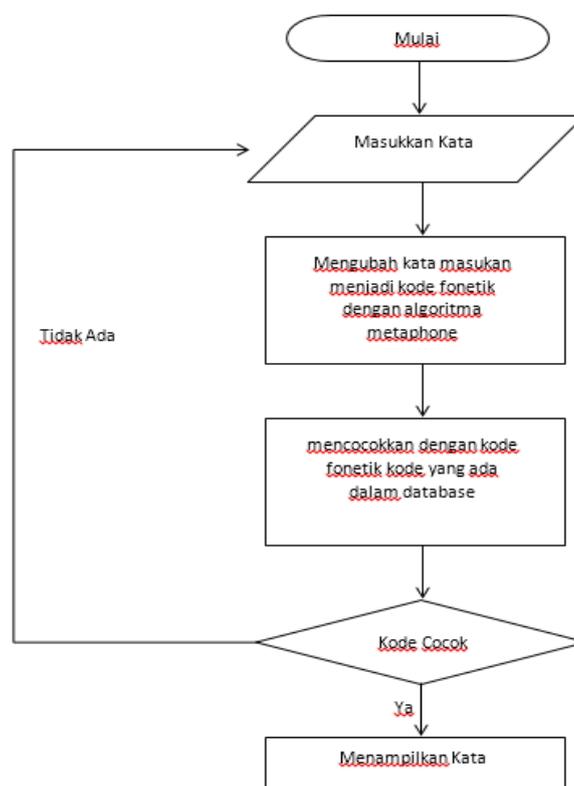
Tahap analisa kebutuhan aplikasi ini meliputi analisis data, kebutuhan fungsional dan non fungsional. Analisis data dilakukan dengan cara mengumpulkan data-data dan literature pendukung yang berhubungan dengan tugas akhir ini. Kebutuhan fungsional terdiri dari spesifikasi kebutuhan perangkat lunak / sistem yaitu sistem dapat menampilkan kata yang dicari oleh pengguna, sistem dapat menampilkan detail kata, dan sistem dapat menampilkan pilihan algoritma pencarian, sebagaimana ditunjukkan Gambar 1. Sedangkan Kebutuhan non fungsional yaitu aplikasi dapat berjalan pada operating sistem android version 4.0.3 – 4.0.4 keatas.



Gambar 1. Use Case Diagram Kebutuhan Fungsional

3.3. Alur Sistem

Pada Gambar 2 dapat dilihat alur dari sistem pengelompokan kata berdasarkan kemiripan ucapan pada kamus dengan algoritma metaphone pada operating sistem android.



Gambar 2. Flowchart Pencarian

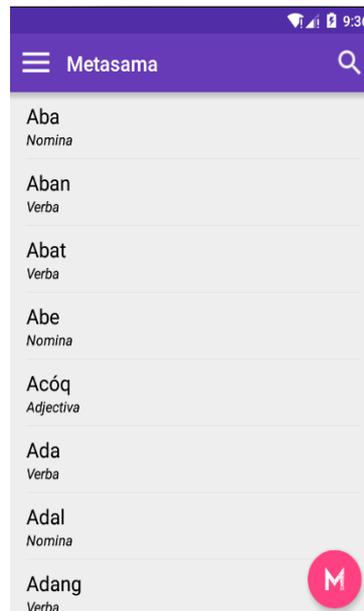
4. Hasil dan Pembahasan

Implementasi perangkat lunak dilakukan dengan menggunakan Anroid Studio sebagai code editor dan DB Browser for Sqlite untuk penyimpanan data.

4.1. Implementasi

- Beranda

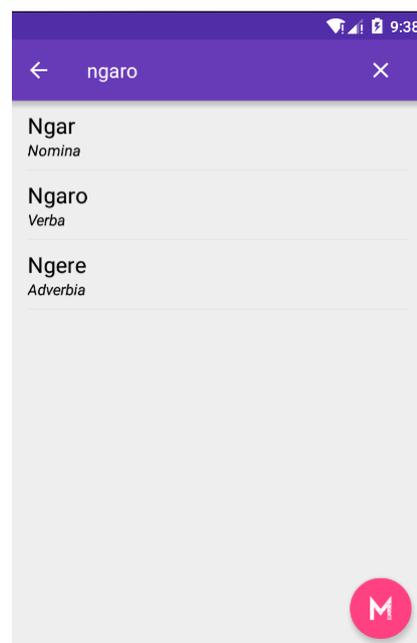
Beranda merupakan tampilan awal yang akan muncul ketika aplikasi pertama kali dibuka, sebagaimana ditunjukkan Gambar 3.



Gambar 3. Tampilan Beranda

- Pencarian Kata

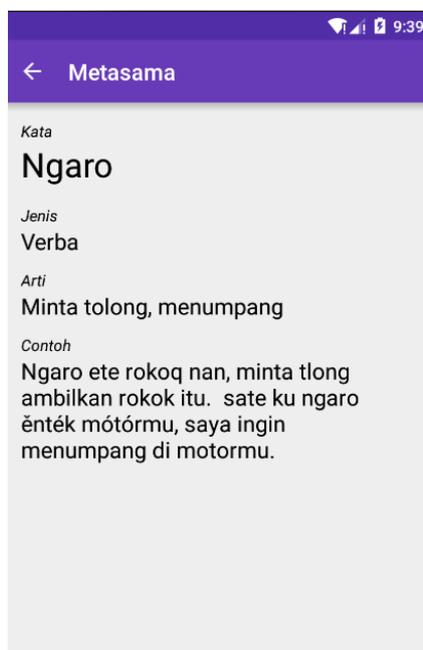
Tampilan pencarian kata akan muncul saat pengguna menggunakan fitur *search*, sebagaimana ditunjukkan Gambar 4.



Gambar 4. Tampilan Pencarian Kata

- Detail Kata

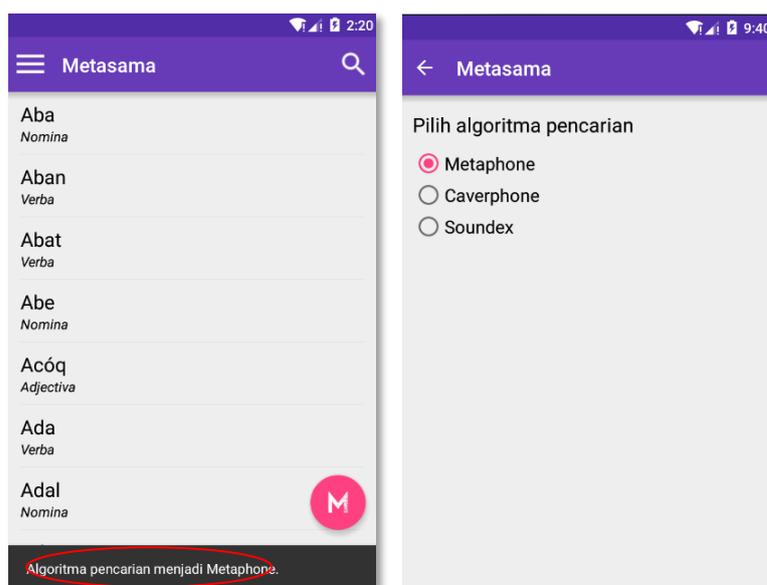
Detail kata merupakan tampilan yang berisi tentang keterangan dari kata yang dicari oleh pengguna. Keterangan yang ditampilkan pada detail kata meliputi arti kata, jenis atau sifat kata, dan contoh penggunaan kata tersebut, sebagaimana ditunjukkan Gambar 5.



Gambar 5. Tampilan Detail Kata

- Pilih Algoritma

Tampilan untuk pilih algoritma dapat diakses dengan menggunakan icon “M” yang terletak pada pojok kanan bawah saat pengguna berada ditampilan awal (beranda). Beberapa algoritma yang dapat digunakan adalah Metaphone, Caverphone, dan Soundex. Fitur ini untuk mempermudah dalam proses pengujian keefektifan masing-masing algoritma dalam pengelompokan kata, sebagaimana ditunjukkan Gambar 6.



Gambar 6. Tampilan Pilih Algoritma

- Algoritma Metaphone

Agar dapat diterapkan pada bahasa sumbawa diperlukan aturan metaphone yang sesuai dengan penuturan bahasa sumbawa agar pengelompokkan kata dapat dilakukan dengan optimal. Berikut merupakan aturan algoritma metaphone yang telah disesuaikan untuk mengelompokkan kata untuk bahasa sumbawa.

1. Hilangkan huruf yang memiliki duplikat, kecuali G [7].
2. Biarkan G yang berurutan (ganda), jika diikuti oleh vocal dan jika huruf sebelum G adalah N (contoh: BANGGGA >> BANGGA, ABANGGG >> ABANG) [7].
3. Ubah A menjadi W jika diikuti oleh U atau O, dan ubah menjadi Y jika diikuti oleh I atau E.
4. Ubah E menjadi W jika diikuti oleh U, dan ubah menjadi Y jika diikuti oleh A, I, atau O.
5. Hilangkan E jika diikuti oleh L atau M [8].
6. Jika I diikuti oleh vocal (kecuali vocal I) ubah menjadi Y.
7. Jika O diikuti oleh vocal (kecuali vocal O) ubah menjadi W.
8. Jika U diikuti oleh vocal (kecuali vocal U) ubah menjadi W.
9. Jika B berada diakhir ubah menjadi P (contoh: MAGRIB >> MAGRIP).
10. Jika D berada diakhir ubah menjadi T (contoh: ABAD >> ABAT).
11. Hilangkan H jika berada diakhir (contoh: BASUH >> BASU).
12. Ubah F menjadi P.
13. Ubah V menjadi P.
14. Ubah Q menjadi K.
15. Ubah X menjadi K [8].
16. Ubah Z menjadi S [8].
17. Hapus semua vocal kecuali jika berada diawal.

Contohnya pada kata baruak jika diubah dengan algoritma metaphone secara berurutan akan menjadi baruak > BARUAK > BRWK. Sehingga kode untuk kata baruak adalah BRWK.

4.2. Pengujian

Pengujian yang dilakukan untuk mengetahui keefektifan aturan baru metaphone yaitu pengujian nilai recall dan precision. Recall diuji berdasarkan perbandingan antara jumlah kata relevan yang ditampilkan oleh sistem dengan jumlah kata relevan seluruhnya. Sedangkan precision diuji berdasarkan perbandingan antara kata yang diinginkan oleh pengguna dengan kata relevan yang dapat ditampilkan oleh sistem [9]. Pemilihan kata yang diujikan berdasarkan kesamaan penyebutan dari beberapa huruf seperti k,q,b,p,v,d,t dan huruf vocal ganda yang membentuk huruf w ataupun y tergantung bunyi bentukan dari kedua vocal tersebut, seperti pada kata Rua dan Siup.

Tabel 1. Nilai Rata-Rata Pengujian Recall Dan Precision

Algoritma	Recall	Precision
Metaphone	98,97	78,52
Soundex	98,98	44,66
Caverphone	98,59	79,01

Pada Tabel 1 merupakan hasil dari pengujian recall dan precision dari masing-masing algoritma. Dapat dilihat pada tabel diatas, nilai recall Metaphone dan Soundex memiliki hasil rata-rata yang lebih tinggi jika dibandingkan dengan Caverphone. Nilai recall tersebut menunjukkan bahwa algoritma Metaphone dan Soundex mampu menampilkan kata yang mungkin relevan lebih banyak dibandingkan dengan algoritma Caverphone. Hal ini disebabkan karena adanya beberapa kesamaan penyebutan dari beberapa huruf saat melakukan pencarian yang tidak ditangani oleh algoritma Caverphone seperti pada kata ahat dan magrep. Algoritma Caverphone tidak melakukan perubahan huruf yang memiliki penyebutan yang sama seperti pada huruf d dan t kemudian b dan p. Namun hasil rata-rata pengujian dari precision menunjukkan, Caverphone dan Metaphone memiliki rata-rata nilai tertinggi jika dibandingkan dengan Soundex.

Suatu algoritma dikatakan efektif apabila nilai rasio recall dan precision sama besarnya (1:1) [10]. Akan tetapi dalam kenyataannya, hasil yang diperoleh menunjukkan nilai recall lebih tinggi dibandingkan nilai precision. Selain itu, algoritma dinyatakan efektif apabila hasil penelusuran mampu menunjukkan *precision* yang tinggi meskipun *recall* rendah. Sehingga untuk penarikan kesimpulan dari tugas akhir ini mengacu pada hasil nilai rata-rata precision yang telah diuji. Hasil nilai rata-rata pengujian precision menunjukkan algoritma Metaphone dan Caverphone

mencapai 78,52 dan 79,01% sedangkan algoritma Soundex hanya memperoleh 44,66%. Algoritma Metaphone yang baru menunjukkan rata-rata lebih baik dari dari soundex dan sama baiknya dengan Caverphone. Jadi aturan baru untuk algoritma Metaphone dalam mengelompokkan kata Sumbawa dapat dikatakan efektif.

5. Kesimpulan

Berdasarkan hasil analisa, perancangan, implementasi dan pengujian yang dilakukan, maka dapat diambil kesimpulan sebagai berikut:

1. Perancangan aplikasi pengelompokan kata berdasarkan kemiripan ucapan pada kamus menggunakan algoritma metaphone pada operating sistem android telah dibuat sesuai dengan spesifikasi dan dapat menjalankan kebutuhan fungsional dengan baik seperti mencari kata, menampilkan detil kata, dan menampilkan pemilihan algoritma pencarian.
2. Berdasarkan hasil pengujian algoritma Metaphone yang baru untuk mengelompokkan kata dalam bahasa Sumbawa menunjukkan rata-rata lebih baik dari dari soundex dan sama baiknya dengan caverphone dengan nilai *recall* sebesar 98,97% dan *precision* sebesar 78,52%. Jadi aturan baru Metaphone yang dibuat untuk mengelompokkan kata dalam bahasa Sumbawa dapat dikatakan efektif.

Referensi

1. Tondo H. *Kepunahan Bahasa-Bahasa Daerah: Faktor Penyebab dan Implikasi Etnolinguistik*. Jurnal masyarakat dan budaya, volume 11.
2. Asako, S. 2013. Voice in the Sumbawa Besar dialect of Sumbawa. *Nusa* 54, 145-158.
3. Mahsun. 1997. "Kekerabatan Bahasa-Bahasa Di Nusa Tenggara Barat: Kajian Tanah Asal Penutur-Penuturnya". Makalah pada Bulan Apresiasi Budaya IV NTB, Mataram.
4. Prasyas O. 2015. "RANCANG BANGUN KAMUS BAHASA SUMBAWA - INDONESIA BERBASIS ANDROID". Pemrograman Teknologi Informasi dan Ilmu Komputer, Informatika, Universitas Brawijaya
5. Rahmattullah S. 2015. "Rancang Bangun Kamus Besar Bahasa Daerah Sumbawa Berbasis Android Dengan Algoritma Caverphone". Teknik Informatika, Universitas Muhammadiyah Malang.
6. Adiwiguna G. 2015. "*Analisis dan Implementasi Pencocokan String Berdasarkan Kemiripan Penguacapan (Phonetic String Matching) Menggunakan Algoritma Metaphone Dalam Pencarian Ayat Al-Quran*". Teknik Informatika, Fakultas Informatika, Universitas Telkom.
7. Lawrence Philips, "Hanging on the Metaphone", *Computer Language* v7 n12, pp39-43.
8. Rahadian, R. "Phonetic String Matching Termodifikasi". Departemen Teknik Informatika, Sekolah Tinggi Teknologi Telkom Bandung.
9. Putung D. 2016. "*Penerapan Sistem Temu Kembali Informasi Pada Kumpulan Dokumen Skripsi*". Teknik Informatika, Universitas Sam Ratulangi Manado.
10. Hasugian, J. 2006. "Penelusuran Informasi Ilmiah Secara Online: Perlakuan Terhadap Seorang Pencari Informasi Sebagai Real User". *Pustaka* Vol.2 (1).

