

PERINGKASAN TEKS OTOMATIS UNTUK DOKUMEN BAHASA BALI BERBASIS METODE EKTRAKTIF

I Putu Gede Hendra Suputra

Program Studi Teknik Informatika, Jurusan Ilmu Komputer
Fakultas Matematika Dan Ilmu Pengetahuan Alam
Universitas Udayana
Email : hendra.suputra@unud.ac.id

ABSTRAK

Membaca menjadi salah satu kegiatan wajib yang dilakukan oleh manusia dalam memperoleh informasi, baik membaca buku, majalah maupun tulisan digital. Perkembangan teknologi internet menyebabkan perkembangan jumlah dokumen digital di internet semakin meningkat. Masalah terjadi ketika dokumen yang dibaca sangat panjang dan sangat banyak sehingga akan membutuhkan waktu yang sangat lama untuk mendapatkan informasi dan memahami isi dari dokumen tersebut. Salah satu cara untuk mendapatkan informasi dan memahami suatu dokumen secara cepat yaitu dengan membaca ringkasannya. Cara untuk mendapatkan ringkasan dokumen secara cepat yaitu dengan meringkasnya secara otomatis dengan menggunakan aplikasi peringkasan otomatis.

Peringkasan teks otomatis (Automatic Text Summarization) merupakan proses menghasilkan atau mengekstraksi sebuah teks yang berasal dari satu atau banyak teks dimana didalamnya terdapat informasi penting dari teks sumber, dan teks hasil tersebut tidak lebih panjang dari setengah teks sumber secara otomatis. Berbagai penelitian telah dilakukan dalam hal peringkasan otomatis untuk berbagai bahasa namun pada penelitian-penelitian sebelumnya belum pernah ada peringkasan otomatis untuk dokumen/tulisan berbahasa Bali yang sudah sangat banyak dijumpai di internet. Merujuk dari fakta tersebut sehingga pada saat ini merupakan momentum yang tepat untuk mengembangkan sistem peringkasan otomatis untuk dokumen berbahasa Bali yang tentunya memiliki karakteristik berbeda dengan kebanyakan bahasa lainnya.

Metode yang digunakan dalam peringkasan teks ini adalah metode ekstraktif yaitu mengekstraksi kalimat-kalimat ringkasan berdasarkan skor fitur-fitur penting dari kalimat yang dimiliki. Pada penelitian ini, fitur-fitur yang digunakan adalah Keyword positif ($f1$), Kemiripan Antar-Kalimat ($f2$) dan Kemiripan Kalimat dengan Judul dokumen ($f3$). Percobaan dilakukan terhadap kumpulan artikel Bahasa Bali yang diperoleh dari berbagai sumber. Dari hasil evaluasi nilai rata-rata ROUGE-1 yang terbaik adalah 0.52 dengan kombinasi bobot fitur $f1 = 0.3$, bobot $f2 = 0.5$ dan bobot $f3 = 0.2$. Terlihat fitur $f2$ adalah fitur yang memberikan pengaruh yang dominan dalam peringkasan dokumen Bahasa Bali.

Kata kunci : peringkasan teks otomatis, Bahasa bali, ekstraksi teks

ABSTRACT

Reading became one of the mandatory activities taken by humans in obtaining information, whether reading books, magazines and digital writing. The improvement of internet technology led to the development of the number of digital documents on the internet is increasing. Problems occur when reading the documents that are very long and very much so it will take a very long time to get information and understand the contents of the document. One way to get information and understand a document quickly is by reading the summary. How to get a quick summary of the document by automatically summarizing it using the automatic summary app.

Automatic text summary (Automatic Text Summarization) is the process of generating or extracting a text that comes from one or many texts in which there is important information from the source text, and the text result is no longer than half the source text automatically. Various studies have been done in terms of automatic summarization for different languages but in previous studies there has never been an automatic summary for documents / writings in Balinese language that has been very much encountered on the internet. Refers to that fact, this time is the right momentum to develop an automatic summary system for documents in the Balinese language which certainly has different characteristics from most other languages.

The method used in this text summary is the extractive method of extracting summary phrases based on the scores of important features of the sentences they possess. In this research, the features that used are Positive Keyword ($f1$), Similarity of Sentences ($f2$) and the Sentence Similarity with Document Title ($f3$). The experiment was conducted on a collection of Balinese language articles obtained from various sources. From the result of evaluation, the best average value of ROUGE-1 is 0.52 with the combination of feature weight, $f1 =$

0.3, the weight of $f_2 = 0.5$ and the weight of $f_3 = 0.2$. Visible feature f_2 is a feature that gives the dominant influence in the compilation of Balinese language documents.

Keywords: automatic text summarization, Balinese language, Text extraction

1. PENDAHULUAN

Salah satu kebutuhan yang tidak dapat lepas dari kehidupan manusia adalah kebutuhan akan informasi. Membaca menjadi salah satu kegiatan yang wajib yang dilakukan oleh manusia untuk memperoleh informasi, baik membaca buku, majalah maupun tulisan digital. Perkembangan teknologi internet menyebabkan perkembangan jumlah dokumen digital di internet semakin meningkat. Masalah muncul apabila dokumen yang dibaca sangat panjang dan sangat banyak karena membutuhkan waktu yang lama untuk dapat memahami isi dokumen tersebut. Salah satu cara untuk dapat memahami isi dokumen secara cepat adalah membaca ringkasannya. Cara untuk mendapatkan ringkasan dokumen secara cepat yaitu dengan meringkasnya secara otomatis dengan menggunakan aplikasi peringkasan otomatis. Peringkasan teks otomatis (Automatic Text Summarization) didefinisikan sebagai sebuah teks yang merupakan hasil dari satu atau banyak teks dimana teks hasil tersebut menyampaikan informasi penting dari teks sumber, dan teks hasil tidak lebih panjang dari setengah teks sumber secara otomatis (Zhu dan Li, 2012).

Berbagai penelitian telah dilakukan dalam hal peringkasan otomatis yaitu (Suanmali dkk, 2011), (Aristoteles dkk, 2012), (Zulkifli, dkk, 2012) dan (Silvia, et al, 2014). Dari semua penelitian tersebut telah dibangun sistem peringkasan teks otomatis berbasis metode ekstraktif, dengan konsep ekstraksi fitur kalimat dan optimasi pembobotan fitur. Sistem-sistem peringkasan yang telah dibangun tersebut ditujukan untuk berbagai bahasa seperti bahasa Inggris atau bahasa Indonesia namun belum pernah ada peringkasan otomatis untuk dokumen/tulisan berbahasa Bali yang kini sudah sangat banyak dijumpai di internet. Selain hal tersebut Google yang bekerja sama dengan Yayasan Dwijendra, BASAbali.org, mahasiswa dan ahli bahasa Universitas Udayana, serta Balai Bahasa pada tahun 2013 telah mengembangkan bahasa Bali sebagai salah satu bahasa yang ada di mesin pencarian google.co.id.

Merujuk pada dua fakta diatas sehingga pada saat ini merupakan momentum yang tepat untuk mengembangkan sistem peringkasan otomatis yang berbasis metode ekstraktif untuk dokumen berbahasa Bali yang tentunya memiliki karakteristik berbeda dengan kebanyakan bahasa lainnya. Aplikasi yang akan dibangun ini juga memiliki tujuan yaitu untuk melestarikan bahasa Bali di tengah kemajuan teknologi informasi.

2. TEORI, DESAIN, IMPLEMENTASI

2.1.1 Peringkasan Teks Otomatis

Peringkasan teks otomatis adalah proses mengurangi teks pada dokumen dengan menggunakan program komputer untuk membuat ringkasan yang berisikan poin-poin penting dimana hasil ringkasan tidak lebih dari setengah dokumen asli (Radev & McKeown, 2002). Terdapat dua arah dari penelitian-penelitian pada bidang peringkasan teks otomatis yaitu ekstraktif dan abstraktif. Metode peringkasan ekstraktif umumnya fokus terhadap konten-konten mana yang harus menjadi ringkasan. Metode ini sepenuhnya mengandalkan ekstraksi kalimat-kalimat yang ada pada dokumen sumber (Das dan Martin, 2007). Metode peringkasan abstraktif mengedepankan penekanan yang kuat pada bentuk ringkasan dan bertujuan untuk menghasilkan ringkasan sesuai dengan tata bahasa (Zhu dan Li, 2012). Umumnya pendekatan tersebut menggunakan teknik-teknik *language generation* (Das dan Martins, 2007). Namun pendekatan abstraktif memiliki keterbatasan yaitu sangat tergantung pada domain yang digunakan

2.1.2 Kalimat

Kalimat adalah bentuk satuan bahasa yang paling kecil dari suatu kesatuan pikiran (Widjono, 2007). Kalimat dibedakan menjadi dua jenis yaitu bahasa tulis dan bahasa lisan. Dalam bahasa tulis, kalimat merupakan kesatuan bahasa yang diawali oleh huruf kapital, kalimat dapat dikombinasi dengan berbagai tanda baca seperti diselingi atau tidak diselingi tanda koma, titik koma atau titik dua dan diakhiri dengan yaitu tanda titik (.), atau jika suatu kalimat memiliki maksud yang lain maka dapat diakhiri dengan tanda tanya (menunjukkan maksud bertanya) atau tanda seru (kalimat suruhan). Dalam bahasa lisan, kalimat didefinisikan sebagai suatu kesatuan bahasa yang dihasilkan dari penggabungan kata dengan kata, penggabungan kata dengan frasa, atau penggabungan frasa dengan frasa, yang minimal merupakan sebuah klausa bebas dan harus mengandung satu subjek dan predikat.

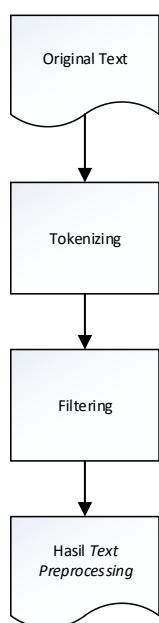
Pada penelitian ini, tiap-tiap kalimat yang ada pada dokumen akan dihitung skornya berdasarkan fitur ekstraksi. Skor dari kalimat tersebut akan menentukan apakah kalimat tersebut penting atau tidak. Kalimat yang memiliki skor tinggi kemungkinan merupakan kalimat yang penting dari sebuah dokumen.

Pada penelitian ini, kalimat dipisahkan berdasarkan tanda titik (.) dan kutipan langsung diasumsikan sebagai satu kalimat yang tidak dapat dipisahkan.

2.1.3 Text Preprocessing

Dataset yang didapat dari suatu sumber merupakan suatu data yang tidak terstruktur dalam hal pemenuhan data kebutuhan sistem peringkasan. Atas dasar tersebut maka sebelum diolah, data harus melalui tahapan *text preprocessing*. *Text Preprocessing* adalah pemrosesan data awal untuk mentransformasi data tekstual yang tidak terstruktur menjadi suatu data baru yang terstruktur. Data baru tersebut dapat mempermudah pemrosesan algoritma peringkasan agar menjadi efektif dan efisien.

Tahap *text preprocessing* terdiri dari beberapa tahap yaitu *tokenizing*, *filtering*, *tagging*, dan *stemming*. Pada penelitian ini hanya menggunakan proses *tokenizing*, dan *filtering*. Proses *tagging* tidak digunakan karena ketiga proses yang telah disebutkan sebelumnya sudah cukup untuk mendapatkan data yang terstruktur. Gambar 1 menunjukkan tahap *preprocessing text*:



Gambar 1 Proses Text Preprocessing

2.1.4 Fitur Ekstraksi Teks

Fitur-fitur yang digunakan dalam penelitian ini adalah fitur-fitur yang berasal dari ekstraksi teks dan dari fitur-fitur tersebut dikombinasikan untuk mendapatkan skor dari setiap kalimat. Nilai tiap fitur dinormalisasi sehingga selalu berada dalam range $[0,1]$. Normalisasi tersebut dilakukan agar nilai selisih dari masing-masing fitur ekstraksi tidak besar.

Fitur-fitur ekstraksi kalimat yang digunakan pada penelitian ini yaitu keyword positif pada kalimat (f_1), kemiripan antar kalimat (f_2), dan kemiripan kalimat dengan judul (f_3). Penjelasan dari tiap-tiap fitur adalah sebagai berikut ini:

2.1.5 Fitur Keyword positif (F_1)

Keyword positif adalah kata yang sering muncul pada sebuah paragraf (Marlina, 2012). Fitur ini dapat dihitung menggunakan rumus (2.1):

$$Score_{f_1}(s_i) = \frac{s_i(\text{positif keyword})}{\sum_{i=0}^n s_i(\text{positif keyword})} \quad (2.1)$$

Dengan s_i (*keyword positif*) adalah jumlah kata dalam suatu kalimat yang mengandung keyword dibagi dengan jumlah kata dalam seluruh kalimat yang mengandung keyword, dengan keyword merupakan banyaknya kata yang muncul dalam suatu dokumen.

2.1.6 Fitur Kemiripan Antar-Kalimat (F_2)

Kemiripan antar-kalimat dihitung berdasarkan jumlah kata-kata dalam suatu kalimat terhadap jumlah kata-kata yang sama (cocok) dengan kata tersebut yang ada pada kalimat lainnya.

Fitur ini dihitung dengan menggunakan rumus (2.2) (Aristoteles dkk, 2012).

$$Score_{f_2}(s) = \frac{|\text{Keyword dalam } s \cap \text{Keyword dalam kalimat lain}|}{|\text{Keyword dalam } s \cup \text{Keyword dalam kalimat lain}|} \quad (2.2)$$

2.1.7 Fitur Kemiripan Kalimat dengan Judul Dokumen (F_3)

Fitur ini dihitung berdasarkan frekuensi kata yang dicocok (sama) antara suatu kalimat dengan kata-kata yang muncul disuatu judul dokumen (Aristoteles dkk, 2012). Fitur ini dapat dihitung dengan menggunakan rumus (2.3).

$$Score_{f_3}(s) = \frac{|\text{Keyword dalam } s \cap \text{Keyword dalam judul}|}{|\text{Keyword dalam } s \cup \text{Keyword dalam judul}|} \quad (2.3)$$

2.1.8 Pembobotan Fitur Ekstraksi Teks

Pembobotan fitur ekstraksi teks adalah sebuah pendekatan yang dilakukan untuk menentukan kepentingan suatu fitur dari fitur-fitur yang akan diteliti dengan cara mengalikan bobot dengan skor fitur ekstraksi (Berker & Gungor, 2012). Pembobotan ini sangat berpengaruh terhadap akurasi hasil ringkasan sistem nantinya. Skor ($Score(S)$) untuk tiap kalimat dapat dihitung dengan menggunakan rumus (2.4):

$$Score(S) = \sum_{k=1}^n w_i * f_i \quad (2.4)$$

Diasumsikan w_i adalah bobot fitur ke- i dan f_i adalah fitur ekstraksi ke- i .

2.1.9 Evaluasi Hasil Ringkasan Sistem

Metode evaluasi yang sering dilakukan pada penelitian peringkasan teks adalah metode intrinsik. Pendekatan dalam evaluasi intrinsik yang

dilakukan adalah membandingkan hasil peringkasan sistem dengan ringkasan ideal. Ringkasan ideal dapat dibuat khusus oleh pakar atau dengan menggabungkan ringkasan-ringkasan yang dibuat oleh manusia menjadi sebuah ringkasan ideal.

Penelitian ini menggunakan metode evaluasi *Recall-Oriented Understudy for Gisting Evaluation* (ROUGE). ROUGE menghitung jumlah n -gram kata yang *overlap* antara ringkasan sistem dengan ringkasan referensi. Adapun teknik penghitungan ROUGE-N antara sebuah ringkasan sistem dan sekumpulan ringkasan manual terdapat pada persamaan 2.5.

$$ROUGE - N = \frac{\sum_{S \in \{Reference Summaries\}} \sum_{e \in S} Count_{match}(gram_n)}{\sum_{S \in \{Reference Summaries\}} \sum_{gram_n \in S} Count(gram_n)} \quad (2.5)$$

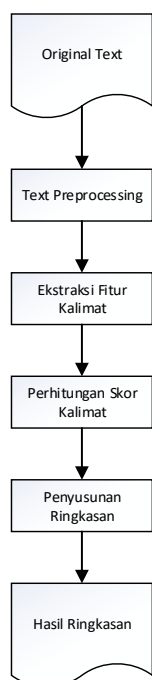
Dimana n adalah panjang dari n -gram, $Count_{match}(gram_n)$ adalah jumlah n -gram yang sama antara sebuah ringkasan sistem dan sebuah ringkasan referensi, $Count(gram_n)$ adalah jumlah n -gram dalam ringkasan referensi.

Studi dari Steinberger & Jezek, 2009 menunjukkan bahwa evaluasi otomatis menggunakan versi unigram dari ROUGE-N, yaitu ROUGE-1 berkorelasi baik dengan evaluasi manusia berdasarkan berbagai statistik. Oleh karena itu penelitian ini menggunakan evaluasi hasil ringkasan sistem dengan ROUGE-1.

2.3 Desain

2.3.1 Diagram Alir Proses Peringkasan Teks

Diagram alir proses peringkasan teks ditunjukkan pada Gambar 2 seperti berikut.



Gambar 2 Diagram Alir Sistem

Proses peringkasan teks otomatis dokumen Bahasa Bali dimulai dengan membaca judul, isi dokumen, dan rasio ringkasan. Masukan judul dan isi dokumen berupa teks. Sedangkan, masukan rasio ringkasan berupa angka presentase dari jumlah kalimat ringkasan yang diinginkan.

Setelah membaca masukan judul, isi dokumen dan rasio ringkasan, dilakukan proses *text preprocessing* yang. *Text preprocessing* dimulai dengan *tokenization* kalimat dan kata dari judul dan isi dokumen menjadi kata-kata (*token*). Setelah itu dilakukan proses pengecekan kata-kata (*token*) di dalam daftar *stop word*. Jika *token* yang terdapat dalam judul dan isi dokumen terdapat dalam daftar *stop word* maka proses yang dilakukan yaitu eliminasi kata. Kata atau *token* tersebut akan dibuang. Jika *token* tidak terdapat dalam daftar *stop word*, proses dilanjutkan yaitu proses *stemming*. Kata atau *token* akan dipecah kedalam bentuk kata dasarnya.

Setelah melewati proses *text preprocessing*, dilakukan proses perhitungan fitur kalimat. Proses yang dilakukan adalah perhitungan fitur keyword positif ($f1$) per kalimat, fitur kesamaan dengan kalimat lain ($f2$), fitur kemiripan kalimat dengan judul ($f3$).

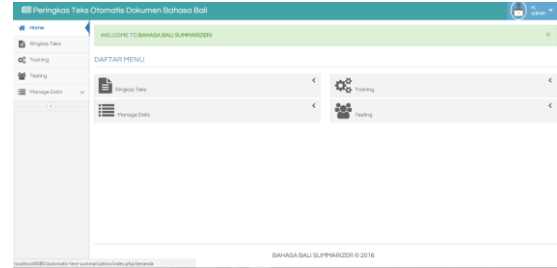
Proses yang selanjutnya adalah proses perhitungan total skor fitur per kalimat. Fitur-fitur yang sudah dihitung sebelumnya akan dikalikan dengan bobot tiap fitur yang sudah didapat dari proses *training* dengan membangkitkan secara random kemudian dijumlahkan menjadi total skor. Total skor akan diurutkan berdasarkan nilai tertinggi hingga terendah. Setelah mendapatkan urutan total skor, dilanjutkan dengan pemotongan jumlah kalimat dengan cara mengalikan rasio ringkasan sistem dengan jumlah kalimat. Kemudian sistem akan menampilkan hasil ringkasan.

2.4 Implementasi

Pada tahap implementasi ini, desain sistem diimplementasikan ke dalam kode program. Sistem ini akan dibangun menggunakan teknologi berbasis web menggunakan bahasa pemrograman PHP, kode HTML, CSS, dan JavaScript, serta menggunakan basis data MySQL untuk pengolahan dataset. Berikut adalah contoh data set yang akan diolah dan akan disimpan di dalam database seperti yang ditunjukkan oleh Gambar 3 berikut.

```
<?xml version="1.0"?>
<ROOT>
  <DOCUMENT>
    <TITLE>Parade Budaya Mepaketan Wanti Warsa Singaraja Sne Kaping 412</TITLE>
    <TEXT>
      Parade budaya mepaketan ring wanti warsa Kota Singaraja sane kaping 412 kagelar
      pinaka pangrayana sajeroning makudang-kudang acara sane sampun kalaksanayang. Acara
      sane kautanayang ring Tugu Singa Anbara Raja arep Kantor Bupati Buleleng puniki
      karahin olih Bupati Buleleng Putu Agus Suradnyana, Manggala miwah para angga DPRD
      Buleleng, Wakil Bupati Nyoman Sutjira, Sekda Ir. Dewa Ketut Puspaka, angga graja
      Kabupaten Buleleng, Forkopinda Buleleng, Manggala Tim Penggerak PKK Kabupaten
      Buleleng, Camat sa-Kabupaten Buleleng, Direktur BUMD, Direktur RSUD, Manggala MHP
      Buleleng, Manggala PHRI Buleleng, para seniman, budayawan, miwah para undangan
      siosan. Manggala Dinas Kebudayaan lan Pariwisata (Disbudpar) Kabupaten Buleleng Drs.
      Gede Suyasa, M.Pd sane pinaka manggala prawatarka mampang acara puniki kalaksanayang
      madasar antuk panglimbak jagate sane ngawinang kawentenan seni budaya Bali sayan
      henit karungu. Program sakadi puniki kalaksanayang mangda prasida ngukuhang
      kawentenan seni lan budaya mangda tetep ajeg. Acara sane kagelar anggen ngromyanin
      wanti warsa Kota Singaraja sane kaping 412 puniki matetujon mangda sami kawentenan
      seni budaya sane wenten ring soang-soang kecamatan utawi desa prasida kauningin lan
      para yowanane prasida sayan runggu ring seni budaya sane wenten ring Buleleng. Unteng
      sane kaambil inggih punika "Kukul Desa" sane madue teges mabinayan suara sakewanten
      tungguli ring sajeroning tetujon. "Tiising ngatiang kawentenan seni budaya sane
      pastika akeh lan mabinayan sumangdane prasida anggen ngwetuang rasa jengah lan
      samken mangda prasida ngamolihang tetujon ngardi Buleleng sane suhera samta",
      baosdane. Buda (38/3). Acara parade budaya puniki kawawin saking arep Kantor Bupati
      Buleleng ngantar ring Margi Veteran, Gajah Mada, Sutomo, Ahmad Yani, lan puput ring
      Margi Dewi Sartika sisi kaja. Acara puniki kamsletin olih duta saking soang-soang
      kecamatan sane wenten ring Kabupaten Buleleng. Jayanti kapertama (Adikara Nugraha)
      ngamolihang piala lan dana pembinaan akehyanane Rp. 16.000.000,-, Jayanti kaping kalih
```

Gambar 3 Contoh Dataset



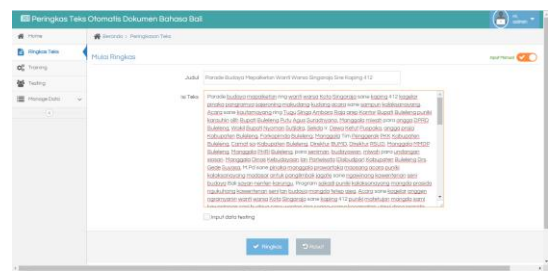
Gambar 4 Halaman Peringkasan Admin

3. UJI COBA DAN EVALUASI

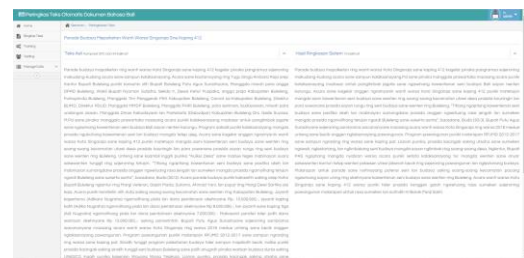
Ujicoba ringkasan dilakukan dengan mengubah-ubah nilai setiap bobot fitur w_i yang bergerak dari 0 hingga 1, dimana jumlah semua bobot adalah 1. Dalam setiap percobaan dicatat dan dicari hasil ROUGE-1 dari ringkasan yang dihasilkan. Nilai ROUGE adalah nilai yang didapat dari perbandingan hasil ringkasan sistem dengan hasil ringkasan manual oleh ahli Bahasa. Formula yang digunakan adalah formula ROUGE-1 sesuai dengan rumus 2.5.

4. HASIL UJI COBA

Sistem yang dibuat diimplementasikan dalam code bahasa pemrograman PHP dengan menggunakan database MySQL. konten yang dibangun adalah konten berbasis web. Sampai saat ini proses yang dilalui adalah proses tahapan text preprocessing, ekstraksi fitur kalimat, penyusunan ringkasan, dimana pada proses tersebut teks-teks inputan user telah diolah menjadi token-token dan difilter untuk menghilangkan kata-kata yang tidak berpengaruh. Proses selanjutnya adalah proses perhitungan nilai fitur kalimat hingga penyusunan ringkasan. Proses-proses tersebut digambarkan secara detail mulai dari Gambar 5 hingga 6. Sedangkan pada Gambar 4 terlihat halaman dashboard dari sistem dimana admin dapat melakukan proses peringkasan, proses training untuk mencari kombinasi bobot atau formula terbaik, proses testing untuk mengukur kinerja sistem, dan proses manage data/dokumen.



Gambar 5 Halaman Form Peringkasan



Gambar 6 Halaman Menampilkan Hasil Peringkasan Teks

Sebelum melakukan evaluasi terhadap sistem maka terlebih dahulu ditentukan berapa proporsi optimal untuk tiap data yang ada dalam proses training. Dataset yang diperoleh dibagi menjadi dua yaitu 60% buah untuk training dan 40% buah untuk testing. Dari hasil training didapatlah hasil optimal yang untuk ketiga buah fitur yaitu $w1 = 0.3$, $w2 = 0.5$ dan $w3 = 0.2$. Setelah mendapatkan bobot optimal tersebut langkah selanjutnya melakukan testing dengan dataset testing. Testing yang dilakukan dengan mengukur nilai ROUGE-1 berhasil mendapatkan nilai rata-rata terbaik yaitu 0.52

5. KESIMPULAN

1. Pembuatan proses pembentukan ringkasan otomatis dilakukan melalui beberapa tahapan yang harus dilalui yaitu text preprocessing, ekstraksi fitur

kalimat (f_1, f_2 , dan f_3), perhitungan skor kalimat dan penyusunan ringkasan berdasarkan ratio yang dihendaki.

2. Proses perhitungan evaluasi dengan ROUGE-1 rata-rata yang di raih adalah 0.52.
3. Perhitungan Nilai ROUGE-1 terbaik ditunjukkan oleh dominasi fitur kedekatan antar kalimat (f_2), dengan kata lain fitur tersebut memberikan pengaruh yang cukup besar terhadap peringkasan Bahasa Bali.

Widjono, HS. (2007). Bahasa Indonesia: Mata Kuliah Pengembangan Kepribadian di Perguruan Tinggi. Grasindo

Zhu T., dan Li, K. (2012), "The Similarit Measure based on LDA for Automatic Summarization", *International workshop on Information and Electronics Engineering (IWIEE)*, Vol. 29, hal. 2944-2949.

Zulkifli, Wibowo, A. T. dan Gia, S. (2012). Pembobotan Fitur Ekstraksi Pada Peringkasan Teks Bahasa Indonesia Menggunakan Algoritma Genetika, Telkom University

6. DAFTAR PUSTAKA

Aristoteles, Herdiyeni, Y., Ridha, A., & Adisantoso, J. (2012). Text Feature Weighting for Summarization of Documents in Bahasa Indonesia Using Genetic Algorithm. *International Journal of Computer Science Issues (IJCSI)*, 9 (3), 1-6.

Berker, M. and Gungor, T. 2012. Using Genetic Algorithms with Lexical Chains for Automatic Text Summarization. ICAART 2012: 595-600

Das, D. dan Martins, A. F. T. (2007), *A Survey on Automatic Text Summarization*, Technical Report Literature Survey for the Language and Statistics II course Carnegie Mellon University, Pittsburgh.

Jezek, K. Steinberger, Josef. 2009. Evaluation Measures For Text Summarization. *Computing and Informatics*, Vol 28: 1001-1026.

Marlina, M. (2012). Sistem Peringkasan Dokumen Berita Bahasa Indonesia menggunakan Metode Regresi Logistik Biner.

Radev, D. R., & McKeown, K. (2002). Introduction to the Special Issue on Summarization. *Computational Linguistics*, 28, 399-408.

Silvia, Rukmana, P., Aprilia, V.R., Suhartono, D., Wongso, R., dan Meiliana. 2014. Summarizing Text for Indonesian Language by Using Latent Dirichlet Allocation and Genetic Algorithm. *Proceeding EECISI 2014*: 148-153.

Suanmali, L., Salim, N., & Binwahlan, M. S. (2009). Genetic Algorithm Based Sentence Extraction For Text Summarization. *International Journal of Innovative Computing*, 1(1), 1-22.