

سَلُوهُرْنَ سَاكِرْنَا فِرَاعُ ۲ عِلْمُ كَرْنِ دَاوُهُنْ سَفَالِهِنَا عُلْمَاءُ ۱
الْصَرْفُ أَمُّ الْعُلُومِ وَالنَّحْوُ أَبُوهَا هَرْتَنَا ارِي صَرْفُ آيْتِ اِنْدُوغْ
فِرَاعُ ۲ عِلْمُ وِجْ اَرِي شَوَايْتِ بَفَانَا فِرَاعُ ۲ عِلْمُ ۱ (۵) وَنِسْبَتُهُ



Nanang Ismail & Abdul Kodir

Kompresi Naskah Bahasa Arab dan Bahasa Arab Pegon Sunda untuk Komunikasi OFDM

ISBN 978-623-7633-05-1



Pusat Penelitian dan Penerbitan
UIN Sunan Gunung Djati Bandung

Kompresi Naskah Bahasa Arab dan Bahasa Arab Pegon
Sunda untuk Komunikasi OFDM pada Next Generation
Telecommunication

Penulis

Nanang Ismail
Abdul Kodir

PUSAT PENELITIAN DAN PENERBITAN
UIN SUNAN GUNUNG DJATI BANDUNG
2019

Kompresi Naskah Bahasa Arab dan Bahasa Arab Pegon Sunda untuk Komunikasi OFDM pada Next Generation Telecommunication

Penulis:

Nanang Ismail
Abdul Kodir

ISBN: xxx – xxx – xxxx – xx – x

Penyunting:

Mufid Ridlo Effendi

Desain Sampul dan Tata letak:

Ahmad Sujana

Penerbit:

Pusat Penelitian Dan Penerbitan UIN SGD Bandung

Jl. H.A. Nasution No. 105 Bandung
Tlp. (022) 7800525, Fax. (022) 7800525
<http://lp2m.uinsgd.ac.id>

Cetakan pertama, Oktober 2019

Hak cipta dilindungi undang-undang

Dilarang memperbanyak karya tulis ini dalam bentuk dan dengan cara apapun tanpa ijin tertulis dari penerbit.

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Alhamdulillahirabbil'alamin, segala puji dan syukur kami panjatkan kepada Allah SWT yang telah melimpahkan nikmat dan rahmat-Nya sehingga penulis dapat menyelesaikan penelitian ini. Shalawat dan salam semoga selalu tercurah bagi nabi kita Muhammad SAW.

“Kompresi Naskah Bahasa Arab dan Bahasa Arab Pegon Sunda untuk Komunikasi OFDM pada Next Generation Telecommunication” merupakan penelitian Pengembangan Pendidikan Tinggi yang didanai dari BOPTN UIN Bandung.

Dalam penelitian ini kami mengucapkan terima kasih yang tulus serta penghargaan yang setinggi-tingginya penulis sampaikan kepada:

1. Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Gunung Djati Bandung.
2. Kepala Pusat Penelitian dan Penerbitan UIN SGD Bandung
3. Bapak Khoirul Anwar, Ph.D selaku Direktur Adwitech, Telkom University
Semoga Allah membalas dengan kebaikan yang tak terhingga serta senantiasa dilancarkan setiap urusannya.
4. Dan semua pihak yang terlibat dalam penelitian ini.

Kami menyadari bahwa laporan penelitian ini masih banyak kekurangan yang perlu diperbaiki. Oleh karena itu, segala saran dan kritik dari semua pihak sangat diharapkan demi kesempurnaannya.

Akhirnya penulis berharap, semoga penelitian ini dapat bermafaat bagi semua pihak serta diridhoi Allah SWT. Amin.

Wassalamu'alaikum Wr. Wb.

Peneliti

DAFTAR ISI

	Halaman
KATA PENGANTAR	ii
DAFTAR ISI.....	iii
DAFTAR GAMBAR	v
DAFTAR TABEL.....	vi
BAB I PENDAHULUAN.....	7
1.1 Latar Belakang	7
1.2 <i>State of The Art</i>	9
1.3 Rumusan Masalah	11
1.4 Tujuan	11
1.5 Batasan Masalah.....	12
BAB II KAJIAN TEORI.....	13
2.1 Teori Informasi.....	13
2.2 <i>Source Code</i>	13
2.3 Kompresi Data	14
2.3.1. <i>Lossless Compression</i>	14
2.3.2. <i>Lossy Compression</i>	15
2.3.3. <i>Huffman Codes</i>	15
2.4 Teks File.....	17
2.5 Probabilitas.....	18
2.6 <i>Entropy</i>	18
2.7 <i>Expected Code Length</i>	20
2.8 <i>Efisiensi</i>	20
2.9 <i>Compression Rates</i>	21
2.10 <i>Outage Probability</i>	21
BAB III METODOLOGI PENELITIAN.....	22
3.1 Alur Metodologi Penelitian.....	22
3.1.1 Pengumpulan dan Pengolahan Data	22
3.1.2 Pembuatan <i>Binary Tree</i> dan <i>Codeword</i>	23
3.1.3 Menghitung <i>Expected Code Length</i> , <i>Efisiensi</i> dan <i>Compression Rates</i>	23
3.1.4 Pengujian Teoritis Menggunakan <i>Outage Probability</i>	23
3.1.5 Analisis Hasil	24
BAB IV ANALISIS STATISTIK UNTUK BAHASA ARAB DAN PEGON.....	25
4.1 Bahasa Arab	25
4.1.1 Objek Yang Dianalisis	25

4.1.2	Pengumpulan dan Pengolahan Data	26
4.1.3	Pembuatan <i>Huffman Tree</i> dan <i>Codeword</i>	31
4.2	Arab Pegon.....	37
4.2.1	Objek Yang Dianalisis	37
4.2.2	Pengumpulan dan Pengolahan Data.....	38
4.2.3	Pembuatan <i>Binary Tree</i> dan <i>Codeword</i>	44
BAB V TINGKAT KOMPRESI DAN KINERJA BAHASA ARAB DAN ARAB PEGON UNTUK TRANSMISI.....		50
5.1	Menghitung <i>Expected Code Length</i>	50
5.1.1	Bahasa Arab	50
5.1.2	Arab Pegon.....	54
5.2	Menghitung Efisiensi	58
5.3	Menghitung <i>Compression Rates</i>	59
5.4	Hasil Penelitian Keseluruhan	60
5.5	Pengujian Kinerja Secara Teoritis Menggunakan <i>Outage Probability</i>	61
BAB VI KESIMPULAN DAN SARAN		65
6.1	Kesimpulan	65
6.2	Saran.....	65
DAFTAR PUSTAKA		66
INDEKS		68

DAFTAR GAMBAR

Gambar 1.1 Ilustrasi Penelitian	9
Gambar 2.1 Contoh Binary Tree	16
Gambar 2.2 Algoritma Huffman Codes [11].....	17
Gambar 4.1 Huffman Tree bahasa Arab	32
Gambar 5.1 Grafik Hasil Simulasi Uji Kinerja Bahasa Arab dan Arab Pegon dengan Teoritis Menggunakan Outage Probability.	62

DAFTAR TABEL

Tabel 2.1 Contoh Pemberian Codeword	17
Tabel 4.1 Pengumpulan dan Pengolahan data Bahasa Arab	26
Tabel 5.1 Expected code length bahasa Arab.....	51

BAB I PENDAHULUAN

1.1 Latar Belakang

Kompresi data merupakan seni yang menampilkan informasi dalam bentuk yang pendek dari data aslinya atau bentuk yang belum dikompresi[1]. Dengan kata lain kompresi data sangat penting karena dapat memperkecil data tanpa menghilangkan informasi yang terkandung di dalamnya dan mempercepat proses pengiriman data. Berdasarkan hal tersebut, semakin berkembangnya teknologi informasi diketahui bahwa kompresi data sangat diperlukan dalam mengirim data (audio, gambar, teks, dan video). Karena dalam komunikasi data, pesan yang dikirim seringkali ukurannya sangat besar sehingga waktu pengirimannya lama[2]. Masalah tersebut dapat diatasi dengan mengkodekan data atau pesan sesingkat mungkin, sehingga waktu pengiriman data relatif cepat dan ruang penyimpanan yang digunakan sedikit.

Kompresi data memegang peranan penting dalam pengembangan dan pembangunan pada Generasi kelima telekomunikasi (5G). Pada tahun 2020 jaringan 5G diharapkan memiliki banyak aplikasi dengan berbagai keunggulan. Kebutuhan jaringan telekomunikasi pada umumnya akan naik, dengan laju pertumbuhan yang meningkat. Teknologi 5G diharapkan bisa memberikan solusi dari berbagai tantangan yang dihadapi termasuk volume downlink dan uplink pada telekomunikasi[3]. Semakin terlihat jelas bahwa harus ada suatu gagasan baru mengenai perkembangan teknologi informasi dan rumusan dengan efisiensi yang maksimal. Salah satu solusinya, misalnya dengan menggunakan kompresi data.

Teknik kompresi data dapat diklarifikasikan sebagai *lossy* dan *loseless*. Kompresi data *lossy* tidak dapat membangun lagi data asli dari data yang telah dikompresi[4]. Ada beberapa detail penting yang mungkin hilang selama

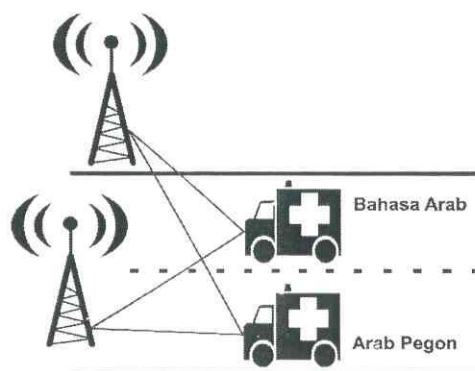
proses kompresi. Kompresi data *lossy* akan menjadi efektif apabila diaplikasikan pada gambar, film dan suara digital. Sedangkan kompresi data *loseless* menggunakan teknik-teknik yang menjamin sebuah duplikat yang sama persis dengan data yang diinput, setelah data tersebut dikompresi. Dalam proses dekompresi tidak ada informasi yang hilang. Kompresi data *loseless* disebut juga kompresi reversibel karena data asli dapat dikembalikan dengan sempurna dengan dekompresi[5]. Teknik kompresi data *loseless* digunakan ketika sumber data asli sangat penting sehingga tidak akan terjadi kehilangan data. Jenis kompresi ini digunakan untuk menyimpan database, spreadsheet, atau untuk memproses file teks.

Bahasa berperan penting bagi manusia, terbukti dengan penggunaannya sehari-hari untuk berkomunikasi dan menyampaikan maksud baik secara lisan maupun tulisan. Pengkaji bahasa telah melakukan beberapa penelitian tentang bahasa-bahasa yang digunakan manusia di dunia. Terdapat lebih dari 5000 wujud bahasa yang ada di dunia, tidak termasuk dialek yang jauh lebih banyak dibanding bahasa induk[6]. Belum ada ketepatan yang dikemukakan para peneliti mengenai jumlah bahasa yang digunakan seluruh umat manusia.

Adapun penelitian mengenai kompresi data teks bahasa Arab dan bahasa daerah Indonesia belum banyak dilakukan baik di Indonesia maupun di institusi ini khususnya di jurusan teknik elektro. Penelitian ini melakukan penelitian *Source Compression* dari bahasa Arab dan bahasa Sunda beraksara Arab Pegon. Penelitian ini mengusulkan pola pelabelan *codeword* untuk kompresi data teks bahasa Arab dan bahasa Sunda beraksara Arab Pegon sehingga dapat berguna untuk pengkodean teknologi komunikasi masa depan, misalnya untuk teknologi generasi kelima (5G). Pola pelabelan yang diusulkan menggunakan algoritma *Huffman Codes* yang memiliki *Expected Code*

Length, yang mendekati *Entropy* Bahasa Arab dan Sunda beraksara Arab Pegon.

Penelitian ini diilustrasikan dengan adanya dua buah *user* atau mobil (Arab dan Arab Pegon) yang sedang mengakses suatu *Base Transceiver Station* (BTS) pada waktu yang bersamaan, dimana kondisi *user* akan mempengaruhi *performance* komunikasi. Seperti dapat dilihat pada Gambar 1.1.



Gambar 1.1 Ilustrasi Penelitian

Dengan demikian, kompresi data pada dasarnya adalah pemampatan data untuk memperkecil ukuran data dari data aslinya tanpa mengubah isi data tersebut, sehingga dapat menghemat penyimpanan data dan mempercepat pengiriman informasi data pada jaringan. Kompresi data teks dapat dilakukan melalui berbagai alternatif. Dalam penelitian ini dilakukan serangkaian penelitian dengan menggunakan *Huffman Codes* untuk mendapatkan nilai *Compression Rates* yang optimal saat digunakan dibidang telekomunikasi[7].

1.2 State of The Art

Pada penelitian sebelumnya, dengan judul *Lossless Compression of Fluoroscopy Medical Images using Correlation and the Combination of Run-length and Huffman Coding* oleh Sarina Mansor, Hezrul Abdul Karim dan Rajasvaran Logeswaran pada tahun 2012 [8]. Melakukan kompresi gambar

medis sekuensial untuk pemeriksaan seperti CT, MRI dan *Fluoroscopy*. metode *lossless compression* gambar *Pharynx* dan *Esofagus Fluoroscopy* dengan menggunakan korelasi dan kombinasi *Run Length* dan *Huffman* yang mengkodekan perbedaan pasangan gambar yang dikelompokkan berdasarkan korelasi. Sedangkan penelitian dengan judul *Development of Word-Based Text Compression Algorithm for Indonesian Language Document* yang dilakukan oleh Ardiles Sinaga, Adiwijaya, and Hertog Nugroho melakukan kompresi berbasis kata terhadap file dokumen bahasa Indonesia, penelitian yang dilakukan lebih menekankan kompresi pada satu metode dan hanya mengkompresi file teks dengan ukuran tertentu [3]. Selain itu penelitian yang dilakukan oleh K. Parish Venkata Kumar, Sk. Khayum, K. Anji Reddy dan B. Lakshmi pada tahun 2016 dengan judul *Data Compression for Lossless Data File*. Paper ini melakukan penelitian kompresi data menggunakan *Lossless Compression* seperti konten, video, suara, rekaman gambar dengan metode *Huffman Codes* dan *Arithmetic Codes* [4]. Selain itu penelitian dengan *Source Coding-based Compressions of Indonesian Local Languages for 5G Potential Applications* oleh K. Anwar, R. Fais, and Y. Julian. Dalam penelitiannya, melakukan penelitian *source compression* untuk bahasa lokal Indonesia yaitu Sunda, Jawa, dan Bali. Penelitian memperoleh probabilitas *outage* menggunakan teorema pengkodean *Slepian-Wolf* untuk mengamati aplikasi potensial untuk layanan *voice* generasi kelima (5G). Perbandingan probabilitas *outage* dilakukan terhadap bahasa Perancis dan Inggris sebagai representatif bahasa asing untuk mengetahui tingkat kompresi bahasa lokal Indonesia. Pelabelan simbol atau *codeword* berdasarkan metode *Huffman Code* hanya untuk alfabet saja. Tidak untuk angka dan simbol lainnya. Hasil yang diperoleh, bahasa lokal Indonesia memiliki tingkat kompresi yang tinggi dan probabilitas *outage* yang rendah[5].

Berdasarkan tabel referensi di atas, sudah banyak peneliti yang penelitiannya tentang kompresi data teks. Namun, pada penelitian ini, dilakukan penelitian yang berjudul *Analisis Tingkat Kompresi pada Bahasa Arab dan Sunda beraksara Arab Pegon Menggunakan Huffman Codes*. Penelitian ini hampir sama dengan penelitian referensi pertama pada Tabel 1.1. Namun yang membedakan dalam penelitian ini semua simbol yang ada diberikan *codeword*-nya masing-masing. Selain itu, tujuan utama yang ingin dicapai dalam penelitian ini adalah kompresi teks. Tetapi dapat juga dikombinasikan antara kompresi teks dengan teks, teks dengan *voice* dan *voice* dengan *voice*. Sehingga penelitian ini dilakukan untuk kombinasi dan juga untuk memperkuat bahwa layanan 5G di masa depan khususnya di Indonesia lebih baik menggunakan bahasa daerah. Penelitian ini nantinya lebih menekankan teknik kompresi data pada bahasa Arab dan Sunda beraksara Arab Pegon sehingga menghasilkan data yang lebih optimal menggunakan metode *Huffman Code*. Hal tersebut menjadikan penelitian ini termasuk ke dalam kategori penelitian baru.

1.3 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah:

1. Bagaimana tingkat kompresi bahasa Arab dan bahasa Sunda beraksara Arab Pegon menggunakan *Huffman Codes*.
2. Bagaimana tingkat kinerja bahasa Arab dan Sunda beraksara Arab Pegon menggunakan uji teoritis *outage probability*.

1.4 Tujuan

Tujuan yang ingin dicapai dari penelitian ini adalah:

1. Menganalisis tingkat kompresi bahasa Arab dan Sunda beraksara Arab Pegon menggunakan algoritma *Huffman Codes*.

BAB II KAJIAN TEORI

2.1 Teori Informasi

Teori informasi (*information theory*) adalah disiplin ilmu dalam bidang matematika terapan yang berkaitan dengan kuantisasi data sehingga data atau informasi itu dapat disimpan dan dikirimkan tanpa kesalahan (*error*) melalui suatu kanal komunikasi [9]. Entropi informasi (*information entropy*) sering dipakai sebagai alat untuk maksud ini dan biasanya dinyatakan sebagai banyaknya bit rerata yang diperlukan untuk penyimpanan dan pengiriman informasi tersebut.

Aplikasi dari topik dasar dalam teori informasi meliputi kompresi data sama dengan aslinya (*lossless data compression*, misalnya pada file ZIP), kompresi data tidak sama dengan aslinya (*lossy data compression*, misalnya pada file MP3) dan pengkodean kanal (*channel coding*, pada saluran DSL, ADSL, dll) [9].

2.2 Source Code

Suatu *source code* (C) menentukan aturan pemetaan dari sebuah variabel acak $x \in X$ yang diambil dari definisinya (S) untuk panjang string yang berhingga $C(x)$ dari X . Panjang string yang dibutuhkan untuk mengekspresikan realisasi x dari variabel acak dilambangkan dengan $l(x)$. Jadi, *source code* dapat diartikan sebagai sebuah variabel acak yang didalamnya mengandung deklarasi yang dijadikan sebagai sampel [10].

Contoh: $\{C(\text{Red}) = 00, C(\text{Blue}) = 11\}$ adalah suatu *source code* untuk $X = \{\text{Red}, \text{Blue}\}$.

2.3 Kompresi Data

Konsep pemikiran mengenai kompresi mulai dikembangkan sehubungan dengan keterbatasan media penyimpanan terhadap ukuran file data yang akan disimpan dan juga untuk kebutuhan dalam pengiriman data agar lebih lebih efektif dan efisien. Kompresi teks mempunyai tujuan untuk membuat file baru yang berisikan pemampatan teks dari suatu file teks. Kompresi ini tidak akan menghilangkan informasi yang terkandung pada teks yang disimpan dalam file awal sehingga pada saat proses dekompresi, teks tersebut akan dikembalikan ke bentuk teks semula [11].

Didalam teks yang ditulis dalam suatu bahasa akan terdapat ke-*redundant*-an dalam hal kemunculan suatu simbol. Beberapa karakter atau pasangan karakter muncul lebih sering dibanding yang lainnya. Didalam teknik kompresi pengulangan-pengulangan ini akan digantikan dengan kode tertentu sehingga terbentuk file baru dengan ukuran yang lebih kecil. Ukuran file yang lebih kecil dari file awal diharapkan dapat mengatasi masalah mengenai media penyimpanan yang terbatas dan juga untuk proses pengiriman data agar lebih efektif dan efisien [11].

2.3.1. *Lossless Compression*

Lossless Compression adalah suatu teknik kompresi data tanpa hilangnya informasi atau sama dengan data aslinya [1]. Jika data telah dikompresi dengan kompresi *lossless*, maka data asli dapat dipulihkan persis dari data yang dikompres. *Lossless Compression* umumnya digunakan untuk aplikasi yang tidak dapat mentolerir perbedaan antara data asli dan data yang direkonstruksi.

Kompresi teks merupakan area penting untuk *lossless compression*. Sangat penting bahwa rekonstruksi identik dengan teks asli, karena perbedaan yang sangat kecil dapat menghasilkan pernyataan dengan makna yang sangat berbeda [1].

2.3.2. *Lossy Compression*

Lossy Compression adalah suatu teknik kompresi data yang melibatkan hilangnya beberapa informasi atau tidak sama dengan aslinya [1]. Jika data telah dikompres dengan teknik *lossy* umumnya tidak dapat dipulihkan atau direkonstruksi dengan tepat. Sebagai gantinya untuk menerima distorsi ini dalam rekonstruksi, umumnya kita dapat memperoleh rasio kompresi yang jauh lebih tinggi daripada *lossless compression*.

Dalam banyak aplikasi, kurangnya rekonstruksi pasti tidak menjadi masalah. Sebagai contoh, saat menyimpan atau mentransmisikan ucapan, nilai pasti setiap sampel ucapan tidak diperlukan [1].

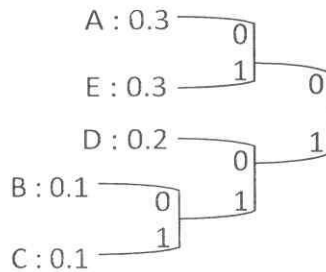
2.3.3. *Huffman Codes*

Huffman Codes adalah kode prefiks optimal (*shortest expected length*) untuk alfabet yang diberikan dimana setiap simbol dalam alfabet telah memiliki probabilitas kemunculan masing-masing. Kode tersebut dirancang oleh *Huffman* sebagai bagian dari tugas kursus di MIT pada awal 1950an [12]. Metode *Huffman Codes* merupakan salah satu teknik kompresi yang menggunakan peluang kemunculan suatu simbol (satu karakter atau grup karakter) untuk menghasilkan suatu kode yang unik. Metode ini menggunakan peluang kemunculan, sehingga metode ini dikategorikan sebagai pengkodean secara statistik (*statistical encoding*). Dengan metode *Huffman Codes*, semakin sering suatu simbol muncul maka panjang kode yang dihasilkan akan semakin pendek, sebaliknya jika kemunculan suatu simbol semakin jarang maka kode yang dihasilkan akan semakin panjang [11].

Kode yang dihasilkan metode *Huffman Codes* merupakan kode yang terbaik karena kode yang dihasilkan menghasilkan panjang kode rata-rata (*average code length*) yang paling pendek daripada kode yang dihasilkan oleh teknik pengkodean statistik yang lain. Kode yang dihasilkan oleh metode *Huffman*

Codes merupakan kode biner yaitu kode yang hanya terdiri dari angka satu dan nol. Sifat khusus yang dimiliki oleh kode Huffman ini yaitu kode yang dihasilkan tidak akan merupakan awalan (*prefix*) dari kode lain. Sebagai contoh, jika kode dari suatu simbol direpresentasikan sebagai 101 maka 10101 atau 10100 tidak dapat dijadikan kode untuk simbol lain [11].

Huffman Codes dibangun dengan menggunakan struktur pohon (*Tree*) dengan dua cabang [11] yang disandarkan sesuai dengan algoritma *Binary Huffman Codes* seperti yang terlihat pada Gambar 2.1. Bentuk struktur pohon seperti ini disebut dengan *Binary Tree*.



Gambar 2.1 Contoh *Binary Tree*

Mula-mula data peluang dari teks diurutkan mulai dari yang terbesar sampai yang terkecil. Dua simbol yang memiliki peluang terkecil kemudian digabungkan untuk membentuk satu *node* baru. *Node* baru ini kemudian diurutkan kembali dengan data sebelumnya. Proses dilanjutkan dengan mengulang penggabungan tadi sehingga akhirnya akan terbentuk pohon dengan nilai kumulatif sama dengan 1. Setiap cabang atas diberi nilai 0 dan cabang bawah diberi nilai 1 [11]. Pemberian *Codeword* untuk masing-masing simbol dilakukan dengan mengambil urutan bit dari akar menuju daun seperti yang terlihat pada Tabel 2.1.

Tabel 2.1 Contoh Pemberian *Codeword*

Simbol	Probabilitas	Codeword
A	0.3	00
E	0.3	01
D	0.2	10
B	0.1	110
C	0.1	111

Agar lebih jelas, algoritma *Huffman Codes* dapat dilihat pada Gambar 2.2.

1. Buatlah suatu barisan terurut yang terdiri dari data yang berisi simbol serta probabilitas simbol dimulai dari probabilitas simbol tertinggi menuju probabilitas simbol terendah.
2. Ambil 2 data dengan probabilitas yang paling rendah dan jumlahkan probabilitasnya.
3. Buatlah satu *node* baru dengan cabang 2 data tersebut.
4. *Node* tersebut disisipkan ke dalam barisan dengan posisi yang sesuai agar barisan tetap dalam keadaan yang terurut.
5. Jika didalam barisan terdapat data dengan probabilitas yang sama dengan probabilitas hasil penjumlahan 2 data tadi, maka *node* baru tersebut disisipkan pada posisi yang lebih tinggi dari data yang sama tadi. Hal ini berguna untuk mendapatkan kode yang lebih efektif dan efisien.
6. Ulangi langkah 2, 3, 4 dan 5 sampai semua simbol diproses dan didapatkan bentuk pohon biner yang utuh dengan jumlah probabilitas 1.
7. Pembentukan kode untuk setiap simbol dilakukan dengan memberikan nilai 0 untuk cabang atas dan memberikan nilai 1 untuk cabang bawah.

Gambar 2.2 Algoritma *Huffman Codes* [11]

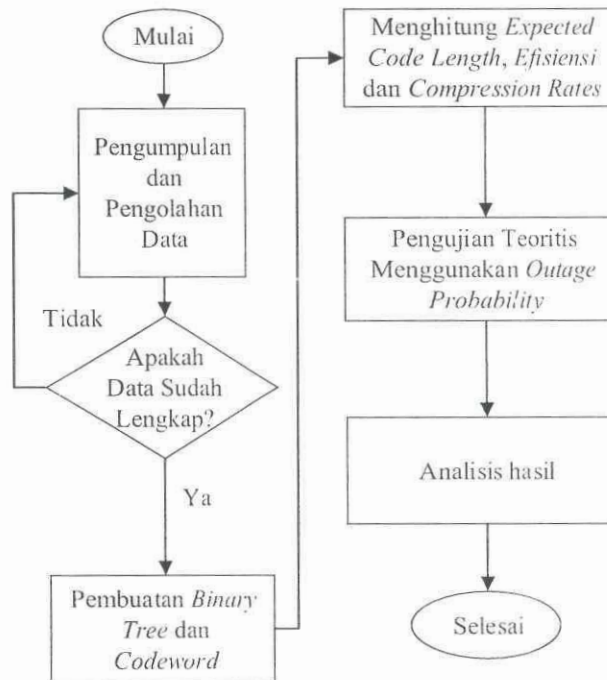
2.4 Teks File

Teks adalah kumpulan dari karakter-karakter atau string yang menjadi satu kesatuan. Teks yang memuat banyak karakter didalamnya selalu menimbulkan masalah pada media penyimpanan dan kecepatan waktu pada saat transmisi

BAB III METODOLOGI PENELITIAN

3.1 Alur Metodologi Penelitian

Berbagai tahapan yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 3.1.



Gambar 3.1 Diagram Alir Penelitian Kompresi Teks

3.1.1 Pengumpulan dan Pengolahan Data

Pengumpulan data meliputi pengambilan sampel untuk penelitian, sampel penelitian dalam penelitian ini diambil dari beberapa artikel bahasa Arab dan Sunda beraksara Arab Pegon dengan ketentuan semua simbol pernah muncul. Semua simbol yang muncul dalam artikel tersebut dihitung frekuensi kemunculan dan probabilitasnya masing-masing. Kemudian, dihitung *Entropy* untuk bahasa Arab dan Sunda beraksara Arab Pegon menggunakan *software* pengolah data *spreadsheet*.

3.1.2 Pembuatan *Binary Tree* dan *Codeword*

Nilai probabilitas dari masing-masing simbol diurutkan mulai dari yang terbesar hingga yang terkecil. Proses pembuatan *Binary Tree* mengacu pada algoritma *Huffman Codes* yang terdapat pada Tinjauan Pustaka. Dua simbol dengan probabilitas terendah dijumlahkan sehingga membentuk node baru. Selanjutnya, node baru tersebut dijumlahkan dengan nilai probabilitas yang ada di atasnya. Langkah tersebut dilakukan sampai didapatkan *Binary Tree* yang utuh dengan nilai probabilitas 1. Pembuatan *codeword* untuk setiap simbol dilakukan dengan memberikan nilai 0 untuk cabang atas dan memberikan nilai 1 untuk cabang bawah. Pemberian *codeword* untuk setiap simbol dilakukan dengan mengambil urutan bit dari akar menuju daun.

3.1.3 Menghitung *Expected Code Length*, *Efisiensi* dan *Compression Rates*

Setelah semua simbol memperoleh *codeword*-nya masing-masing, maka akan diperoleh *length code* untuk setiap simbol tersebut dengan menghitung banyaknya bit pada *codeword* untuk setiap simbol. Kemudian, *Expected Code Length* dihitung dengan cara melakukan perkalian antara nilai probabilitas setiap simbol dengan *length code*-nya masing-masing. *Efisiensi* dihitung dengan cara membandingkan *Entropy* dengan *Expected Code Length*. *Compression Rates* dihitung dengan cara membandingkan nilai *Expected Code Length Huffman Codes* dengan *Expected Code Length* file asli.

3.1.4 Pengujian Teoritis Menggunakan *Outage Probability*

Setelah semua parameter dihitung dan didapatkan hasilnya, selanjutnya dilakukan pengujian teoritis menggunakan *outage probability* untuk mengetahui tingkat kompresi dan kinerja transmisi kedua bahasa tersebut, baik bahasa Arab maupun Arab Pegon. Pengujian dilakukan dengan cara simulasi komputer pada *software* komputasi.

3.1.5 Analisis Hasil

Menentukan hasil yang lebih baik antara bahasa Arab dan bahasa Sunda beraksara Arab Pegon ketika dilakukan uji teoritis menggunakan *outage probability*, tujuannya untuk mengetahui tingkat kompresi dan kinerja transmisi antara kedua bahasa tersebut.

BAB IV ANALISIS STATISTIK UNTUK BAHASA ARAB DAN PEGON

Penelitian ini melakukan penelitian mengenai analisis bahasa Arab dan Arab Pegon yang ada di Indonesia. Objek penelitian ini adalah bahasa bahasa Arab dan Sunda beraksara Arab Pegon.

4.1 Bahasa Arab

Bahasa Arab digunakan sebagai bahasa resmi PBB (Perserikatan Bangsa Bangsa) dan digunakan lebih dari 25 negara, dan merupakan bahasa peribadatan dalam agama Islam karena merupakan bahasa yang dipakai oleh *Al-Qur'an*. Berdasarkan penyebaran geografisnya, bahasa Arab memiliki banyak variasi (dialek), dan dipandang sebagai bahasa yang sangat orisinal tidak memiliki masa kanak-kanak sekaligus masa renta[6].

Sebagai subsistem budaya kekampuan bahasa Arab merupakan salah satu bahasa semit yang paling tua dan tetap eksis hingga sekarang. Bahasa Arab modern telah diklasifikasikan sebagai satu makrobahasa dengan 27 sub-bahasa dalam *ISO 639-3*[19]. Bahasa Arab Baku (kadang-kadang disebut Bahasa Arab Sastra) diajarkan secara luas di sekolah dan universitas, serta digunakan di tempat kerja, pemerintahan, dan media massa.

4.1.1 Objek Yang Dianalisis

Objek yang dianalisis dalam penelitian ini diambil dari beberapa sampel artikel bahasa Arab. Artikel ini diperoleh dari situs web bahasa Arab <http://saudied.com/الدروس-الخصوصية-تقديم-الاختبارات-ي> dan buku bahasa Arab.

التعليم السعودي : على الرغم أن هدفها “تعليمي”، ويرى البعض أن هدفها هو الرفع من المستوى العلمي للطلاب، إلا أن آخرين يرونها مضيعة للمال وإجهادا للأبناء، فيما ذهب البعض إلى أبعد من ذلك من خلال أن هؤلاء المعلمين يستغلون مهنتهم للمتاجرة بهم وجمع أموال مقابل تعليم أبنائهم في المنازل.

تبقى قضية الدروس الخصوصية مشكلة موسمية تُوْرَق الأسر ما بين معارض، ومؤيد يريد استغلال فرصة الدروس الخصوصية لرفع كفاءة أبنائه التعليمية، وسد نقص عدم اهتمامه بمراجعة الحصة الدراسية التي تعطى للطالب في المدرسة خلال العام الدراسي.

Teks artikel hanya sampel bagian dari paragraf satu dan dua saja. Semua sampel artikel yang dijadikan objek penelitian, disajikan dalam lampiran. Jumlah seluruh karakter yang dijadikan sampel penelitian ini berjumlah 2565 karakter dengan kemuculan 64 macam karakter dan 21 macam kemunculan karakter inisialisasi, nilai inisialisasi kemunculan mengikuti jumlah kemunculan frekuensi terkecil. Beberapa karakter yang diinisialisasikan kemunculannya adalah “,” (dibaca tanpa tanda petik), *, &, ^, %, \$, #, @, !, “” لا, }, {, \, |, <, >, ؛, ×, ÷, ‘ dan “~”. Semua karakter yang muncul dihitung frekuensi dan probabilitas kemunculannya. Sehingga setiap karakter mendapatkan *codeword*-nya masing-masing.

4.1.2 Pengumpulan dan Pengolahan Data

Pengumpulan data meliputi pengambilan beberapa sampel teks dari artikel Bahasa Arab. Hasil dari sampel teks artikel tersebut dilakukan pengolahan data meliputi perhitungan frekuensi simbol dan probabilitas kemunculan setiap karakter serta *Entropy* Bahasa Arab. Dan membuat *Binary Tree* yang mengacu pada algoritma *Huffman Codes*, dan akan didapatkan *Codeword* dari tiap karakter serta *Length Code* dari *Codeword* yang sudah didapatkan. Hasil perhitungannya dapat dilihat pada Tabel 4.1.

Tabel 4.1 Pengumpulan dan Pengolahan data Bahasa Arab

Karakter	Frekuensi	Probabilitas	Entropi
spasi	438	0,1708	0,435
ا	281	0,1096	0,350
ل	208	0,0811	0,294
م	162	0,0632	0,252

Karakter	Frekuensi	Probabilitas	Entropi
ي	124	0,0483	0,211
ر	102	0,0398	0,185
نا	96	0,0374	0,177
بـ	82	0,0320	0,159
و	81	0,0316	0,157
د	79	0,0308	0,155
ع	79	0,0308	0,155
ن	77	0,0300	0,152
س	58	0,0226	0,124
هـ	54	0,0211	0,117
فا	53	0,0207	0,116
ة	52	0,0203	0,114
ق	36	0,0140	0,086
،	31	0,0121	0,077
ص	28	0,0109	0,071
خ	28	0,0109	0,071
لا	27	0,0105	0,069
ك	26	0,0101	0,067
ج	25	0,0097	0,065
أ	23	0,0090	0,061
ى	21	0,0082	0,057
.	21	0,0082	0,057
ح	19	0,0074	0,052
ء	18	0,0070	0,050
ض	18	0,0070	0,050

Karakter	Frekuensi	Probabilitas	Entropi
ذ	15	0,0058	0,043
لا	12	0,0047	0,036
!	12	0,0047	0,036
ط	12	0,0047	0,036
-	10	0,0039	0,031
ش	9	0,0035	0,029
:	9	0,0035	0,029
غ	8	0,0031	0,026
/	8	0,0031	0,026
ز	7	0,0027	0,023
ا	7	0,0027	0,023
لا	7	0,0027	0,023
-	7	0,0027	0,023
ث	7	0,0027	0,023
"	6	0,0023	0,020
ئ	6	0,0023	0,020
٩	6	0,0023	0,020
)	6	0,0023	0,020
(6	0,0023	0,020
و	4	0,0016	0,015
=	4	0,0016	0,015
+	4	0,0016	0,015
ا	3	0,0012	0,011
٨	3	0,0012	0,011
؟	3	0,0012	0,011

Karakter	Frekuensi	Probabilitas	Entropi
۳	3	0,0012	0,011
ظ	2	0,0008	0,008
٤	2	0,0008	0,008
٥	2	0,0008	0,008
۷	2	0,0008	0,008
٠	1	0,0004	0,004
۲	1	0,0004	0,004
٦	1	0,0004	0,004
!	1	0,0004	0,004
@	1	0,0004	0,004
#	1	0,0004	0,004
\$	1	0,0004	0,004
%	1	0,0004	0,004
^	1	0,0004	0,004
&	1	0,0004	0,004
*	1	0,0004	0,004
'	1	0,0004	0,004
÷	1	0,0004	0,004
×	1	0,0004	0,004
:	1	0,0004	0,004
<	1	0,0004	0,004
>	1	0,0004	0,004
	1	0,0004	0,004
\	1	0,0004	0,004
]	1	0,0004	0,004

BAB V TINGKAT KOMPRESI DAN KINERJA BAHASA ARAB DAN ARAB PEGON UNTUK TRANSMISI

5.1 Menghitung *Expected Code Length*

Perhitungan *Expected Code Length* dilakukan untuk mendapatkan nilai rata-rata panjang bit dari data sampel teks. Nilai rata-rata panjang kode pada *Huffman Codes* akan mendekati nilai *Entropy*-nya. Semakin kecil perbedaan nilai rata-rata panjang kode dengan nilai *Entropy*-nya akan membuat kode tersebut semakin efisien. Nilai *Expected Code Length* didapatkan dari perkalian antara probabilitas tiap karakter dengan *length code* pada tiap karakter yang sudah didapatkan dari proses pembuatan *binary tree*. Selanjutnya menghitung nilai *Expected Code Length* dengan menggunakan Persamaan 2.3.

5.1.1 Bahasa Arab

Perhitungan *Expected Code Length* bahasa Arab :

$$\begin{aligned} L(C) &= \sum_{x \in X} l(x)p(x) \\ &= (3*0.1708) + (3*0.1096) + (4*0.0811) + (4*0.0632) + \dots + \\ &(11*0.0004) \\ &= 4,767 \text{ bits/simbol} \end{aligned}$$

Dari hasil *Expected Code Length* dapat diketahui apakah proses pengkompresian efektif atau tidak, yaitu dilihat dari nilai *Expected Code Length* yang mendekati nilai *Entropy*. Jika nilai *Expected Code Length* mendekati nilai *Entropy* maka proses pengkompresian bisa dikatakan sudah efektif, dan sebaliknya jika nilai *Expected Code Length* kurang dari nilai *Entropy* maka dapat dikatakan ada yang salah dalam proses pengkompresian

atau perhitungan, karna nilai *Expected Code Length* tidak boleh kurang dari nilai *Entropy*. Hasil perhitungan *expected code length* bahasa Arab dapat dilihat pada tabel 5.1.

Tabel 5.1 *Expected code length* bahasa Arab

Karakter	Probabilitas	Length code	<i>Expected code length</i>
space	0,1708	3	0,512
ا	0,1096	3	0,329
ل	0,0811	4	0,324
م	0,0632	4	0,253
ي	0,0483	4	0,193
ر	0,0398	5	0,199
ت	0,0374	5	0,187
ب	0,0320	5	0,160
و	0,0316	5	0,158
د	0,0308	5	0,154
ع	0,0308	5	0,154
ن	0,0300	5	0,150
ف	0,0226	5	0,113
ه	0,0211	5	0,105
ظ	0,0207	5	0,103
ة	0,0203	6	0,122
ق	0,0140	6	0,084
ـ	0,0121	6	0,073
ك	0,0109	6	0,065
خ	0,0109	6	0,065
لا	0,0105	6	0,063
ط	0,0101	7	0,071
ج	0,0097	7	0,068
أ	0,0090	7	0,063
ى	0,0082	7	0,057
.	0,0082	7	0,057
ح	0,0074	7	0,052
ء	0,0070	7	0,049
ض	0,0070	7	0,049

Karakter	Probabilitas	Length code	Expected code length
ذ	0,0058	7	0,041
لا	0,0047	8	0,037
ا	0,0047	8	0,037
ط	0,0047	8	0,037
-	0,0039	8	0,031
ش	0,0035	8	0,028
:	0,0035	8	0,028
غ	0,0031	8	0,025
/	0,0031	8	0,025
ز	0,0027	8	0,022
ا	0,0027	8	0,022
لا	0,0027	8	0,022
_	0,0027	8	0,022
ث	0,0027	8	0,022
"	0,0023	9	0,021
ئ	0,0023	9	0,021
٩	0,0023	9	0,021
)	0,0023	9	0,021
(0,0023	9	0,021
و	0,0016	9	0,014
=	0,0016	9	0,014
+	0,0016	9	0,014
آ	0,0012	9	0,011
٨	0,0012	9	0,011
؟	0,0012	9	0,011
٣	0,0012	10	0,012
ظ	0,0008	10	0,008
٤	0,0008	10	0,008
٥	0,0008	10	0,008
٧	0,0008	10	0,008
٠	0,0004	11	0,004
٢	0,0004	11	0,004
٦	0,0004	11	0,004
!	0,0004	11	0,004
@	0,0004	11	0,004

Karakter	Probabilitas	Length code	<i>Expected code length</i>
#	0,0004	11	0,004
\$	0,0004	11	0,004
%	0,0004	11	0,004
^	0,0004	11	0,004
&	0,0004	11	0,004
*	0,0004	11	0,004
'	0,0004	11	0,004
÷	0,0004	11	0,004
×	0,0004	11	0,004
!	0,0004	11	0,004
<	0,0004	11	0,004
>	0,0004	11	0,004
	0,0004	11	0,004
\	0,0004	11	0,004
]	0,0004	11	0,004
[0,0004	11	0,004
~	0,0004	11	0,004
}	0,0004	11	0,004
{	0,0004	11	0,004
~	0,0004	11	0,004
,	0,0004	11	0,004
Jumlah	1	-	4,767

Dari hasil *Expected Code Length* yang sudah didapatkan, dapat dilihat bahwa nilai *Expected Code Length* mendekati nilai *Entropy*, dimana nilai *Expected Code Length* bahasa Arab adalah sebesar 4,767 bit, sedangkan nilai *Entropy* sebesar 4.733 bit. Maka dengan hasil tersebut proses pengkompresian yang sudah dilakukan bisa dikatakan sudah efektif jika dilihat dari nilai *Expected Code Length*-nya.

5.1.2 Arab Pegon

Perhitungan *Expected Code Length* bahasa Sunda beraksara Arab Pegon :

$$\begin{aligned}
 L(C) &= \sum_{x \in X} l(x)p(x) \\
 &= (2*0.1938) + (4*0.0882) + (4*0.0776) + (4*0.0532) + \dots + \\
 &\quad (12*0.0004) \\
 &= 4,651 \text{ bits/simbol}
 \end{aligned}$$

Dari hasil *Expected Code Length* dapat diketahui apakah proses pengkompresian efektif atau tidak, yaitu dilihat dari nilai *Expected Code Length* yang mendekati nilai *Entropy*. Jika nilai *Expected Code Length* mendekati nilai *Entropy* maka proses pengkompresian bisa dikatakan sudah efektif, dan sebaliknya jika nilai *Expected Code Length* kurang dari nilai *Entropy* maka dapat dikatakan ada yang salah dalam proses pengkompresian atau perhitungan, karna nilai *Expected Code Length* tidak boleh kurang dari nilai *Entropy*. Hasil perhitungan *expected code length* bahasa Sunda beraksara Arab Pegon dapat dilihat pada tabel 5.2.

Tabel 5.2 *Expected code length* Arab Pegon

Karakter	Probabilitas	Length code	Expected code length
spasi	0,1938	2	0,388
ا	0,0882	4	0,353
ن	0,0776	4	0,310
ل	0,0532	4	0,213
و	0,0484	4	0,194
ت	0,0481	4	0,192
م	0,0425	5	0,213

Karakter	Probabilitas	Length code	Expected code length
ر	0,0410	5	0,205
ي	0,0390	5	0,195
هـ	0,0311	5	0,156
تـ	0,0291	5	0,146
فـ	0,0244	4	0,098
عـ	0,0236	4	0,095
مـ	0,0232	6	0,139
بـ	0,0224	6	0,135
ثـ	0,0213	6	0,128
دـ	0,0189	6	0,113
جـ	0,0130	6	0,078
ى	0,0122	7	0,085
حـ	0,0114	7	0,080
ـ	0,0114	7	0,080
)	0,0091	7	0,063
(0,0091	7	0,063
ة	0,0079	7	0,055
"	0,0079	7	0,055
ا	0,0067	7	0,047
٢	0,0063	8	0,050
ي	0,0059	8	0,047
ـ	0,0059	8	0,047
لـ	0,0059	8	0,047
ـ	0,0051	8	0,041
ـ	0,0035	8	0,028

BAB VI KESIMPULAN DAN SARAN

6.1 Kesimpulan

Berdasarkan penelitian yang dilakukan dapat disimpulkan:

1. Penelitian ini memprediksikan bahwa bahasa Sunda beraksara Arab Pegon memiliki tingkat kompresi yang lebih rendah yaitu sebesar 0.664 dengan efisiensi sebesar 99.9%. Sedangkan, bahasa Arab memiliki tingkat kompresi sebesar 0.681 dengan efisiensi sebesar 99.3%.
2. Pengujian dengan uji teoritis menggunakan *outage probability* menunjukkan bahwa, tingkat kinerja komunikasi bahasa Arab dan bahasa Sunda beraksara Arab Pegon tidak akan terdapat *error floor* jika kedua *user* berada dalam kondisi bergerak secara bersamaan. Namun, jika salah satu *user* berada dalam kondisi diam, maka akan terdapat *error floor*.

6.2 Saran

Penelitian ini masih perlu pengembangan demi mencapai kesempurnaan.

1. Karakter serta simbol yang diuji lebih banyak untuk menguji kinerja kompresi bahasa menggunakan huffman codes.
2. Dapat dikembangkan dengan melakukan simulasi secara real atau nyata untuk menganalisis keinerja kedua bahasa tersebut, baik Bahasa Arab maupun Bahasa Sunda beraksara Arab Pegon. Karena dalam penelitian ini, pengujian baru dilakukan dengan simulasi secara teori saja.
3. Penelitian dapat dikembangkan dengan menambahkan bahasa dan sampel objek penelitian.

DAFTAR PUSTAKA

- [1] K. Sayood, Ed., *Introduction to Data Compression*, 4th ed. USA, 2012.
- [2] M. Program *et al.*, "Analisa dan Perbandingan Algoritma Run Length Encoding dan Algoritma LZW (LEMPEL ZIV WECH)," *J. Ilm. INFOTEK*, vol. 1, no. 2, 2016.
- [3] K. Anwar, R. Fais, and Y. Julian, "Source Coding-based Compressions of Indonesian Local Languages for 5G Potential Applications," *Int. Symp. Electron. Smart Devices*, pp. 2–6, 2017.
- [4] I. Akman, H. Bayindir, S. Ozleme, Z. Akin, and S. Misra, "Lossless text compression technique using syllable based morphology," *Int. Arab J. Inf. Technol.*, vol. 8, no. 1, pp. 66–74, 2011.
- [5] S. R. Kodituwakku and U. S. Amarasinghe, "Comparison of lossless data compression algorithms for text data," *Indian J. Comput. Sci. Eng.*, vol. 1, no. 4, pp. 416–425, 2010.
- [6] A. Bin Muhammad, "Beberapa Aspek Keunikan Dan Keistimewaan Bahasa Arab Sebagai Bahasa Al-Quran," *J. Teknol.*, vol. 42, no. 1, pp. 61–76, 2005.
- [7] A. Sinaga, Adiwijaya, and H. Nugroho, "Development of word-based text compression algorithm for Indonesian language document," *2015 3rd Int. Conf. Inf. Commun. Technol. ICoICT 2015*, pp. 450–454, 2015.
- [8] A. S. Arif, S. Mansor, H. Abdul Karim, and R. Logeswaran, "Lossless Compression of Fluoroscopy Medical Images using Correlation and the Combination of Run-Length and Huffman Coding," *IEEE-EMBS Int. Conf. Biomed. Eng. Sci.*, pp. 759–762, 2012.
- [9] I. Made and A. D. Suarjaya, "A New Algorithm for Data Compression Optimization," *Int. J. Adv. Comput. Sci. Appl.*, vol. 3, no. 8, pp. 14–17, 2012.
- [10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. 2005.
- [11] Y. E. Hutasoit, "Huffman Coding untuk Kompresi Data Teks Berbahasa Indonesia," 2001.
- [12] M. Kuruvila and D. P. Gopinath, "Entropy of Malayalam Language and Text Compression Using Huffman Coding," *2014 1st Int. Conf. Comput. Syst. Commun. ICCSC 2014*, no. December, pp. 150–155, 2003.
- [13] T. John Pattiasina, "Analisa Kode Huffman Untuk Kompresi Data Teks," *Ikado-Tech*, pp. 1–12, 1952.
- [14] Y. Christyono, "Kompresi Data Berdasarkan Perhitungan Distribusi Probabilitas Kemunculan Karakter Orde Dua Dalam Teks Bahasa Indonesia," pp. 33–37.
- [15] Sudjana, *Metoda Statistika*, 7th ed. Bandung: PT. Tarsito Bandung, 2013.
- [16] D. A. Huffman, "Minimum Redundancy Codes." *Proc. I.R.E.*, vol. 27, 1951.
- [17] "Arti Kata - Kamus Besar Bahasa Indonesia (KBBI)) Online." [Online]. Available: <https://kbbi.web.id>. [Accessed: 04-Jul-2018].

- [18] K. P. V. Kumar, S. Khayum, K. A. Reddy, and B. Lakshmi, "Data Compression for Lossless Data File," *Int. J. Mech. Eng. Comput. Appl.*, vol. 4, no. 2, pp. 223–225, 2016.
- [19] M. A. Wahab, "Peran Bahasa Arab Dalam Pengembangan Ilmu Dan Peradaban Islam," *Arab. J. Pendidik. Bhs. Arab dan Kebahasaaraban*, vol. 1, no. 1, 2014.
- [20] "Aksara Arab Pegon di Kancan Internasional | NU Online." [Online]. Available: <http://www.nu.or.id/post/read/86920/aksara-arab-pegon-di-kancan-internasional>. [Accessed: 04-Jul-2018].
- [21] "Mengenal Asal Arab Pegon Serta Kaidah Metode Cara Menulis Arab Pegon Untuk Pemula Dan Penemu Huruf Pegon | Santri-yai." [Online]. Available: <http://www.santriyai.com/2018/01/mengenal-asal-arab-pegon-serta-kaidah.html>. [Accessed: 04-Jul-2018].

INDEKS

- 5G, 1, 2, 4, 5, 61
- Arab Pegon, 1, ii, iv, v, 2, 3, 5, 6, 15, 16, 17, 18, 19, 31, 32, 33, 38, 39, 41, 48, 52, 53, 54, 56, 59, 62
- bahasa Arab, v, vi, 2, 5, 6, 15, 16, 17, 18, 19, 26, 28, 44, 45, 47, 52, 53, 54, 55, 57, 59
- Base Transceiver Station*, 3, 56
- Binary Tree, iii, iv, v, 10, 17, 20, 32, 38
- codeword*, 2, 4, 5, 17, 20, 25, 27, 32, 38, 40
- Compression Rates*, iii, iv, 15, 17, 53, 54
- data, vi, 1, 2, 3, 4, 5, 7, 8, 9, 10, 12, 15, 16, 20, 32, 44, 52, 53, 54, 61
- dialek, 2, 19
- Efisiensi*, iii, iv, 6, 14, 17, 52, 53, 54
- Entropy*, iii, 3, 6, 12, 13, 14, 16, 17, 20, 25, 32, 37, 44, 47, 48, 52, 53, 54, 61
- Expected Code Length*, iii, iv, 3, 6, 14, 17, 44, 47, 48, 52, 54
- expected length*, 9, 14
- Huffman Codes*, iii, v, 2, 3, 4, 5, 6, 9, 10, 11, 13, 14, 15, 17, 20, 24, 25, 27, 32, 36, 37, 38, 40, 44
- Huffman Tree*, iv, v, 25, 26, 27, 38, 39, 40
- informasi, 1, 2, 3, 7, 8, 9, 12, 13
- Kompresi, 1, ii, iii, 1, 3, 5, 8, 16, 61
- komunikasi, 1, 2, 3, 7, 15, 56, 57, 59
- loseless*, 1
- lossy*, 1, 7, 9
- Node*, 10
- outage*, 4, 5, 6, 17, 18, 57, 59
- Outage Probability*, iii, iv, v, 15, 17, 55, 56
- performance*, 3, 56
- probabilitas, 4, 9, 12, 13, 14, 17, 20, 24, 25, 27, 32, 36, 37, 40, 44
- Rayleigh fading*, 15
- redundant*, 8
- reversibel, 2
- Run Length*, 4, 61
- simbol, 4, 5, 6, 8, 9, 10, 13, 16, 17, 20, 25, 32, 37, 44, 48, 59
- Slepian-Wolf*, 4
- source code*, 7, 14
- Source Compression*, 2
- user*, 3, 56, 57, 59
- voice*, 4, 5

سَلُوهُمْ زَنْ سَاكِرْنَا فِرَاعُ ۲ عِلْمُ كَرْنِ دَلُوْهُنْ سَفَالِهِنَا عِلْمَاءُ ۲
الْصَرْفُ اَمْرُ الْعُلُوْمِ وَالنَّحْوُ اَبُوْهَا هَرْتِنَا اِرِيْ صَرْفُ اَيْتِ اِنْذِرْ
فِرَاعُ ۲ عِلْمُ جَعَّ اِرِيْ نَحْوُ اَيْتِ بَفَانَا فِرَاعُ ۲ عِلْمُ (۵) وَنَسَبَتْ



Nanang Ismail & Abdul Kodir

Kompresi Naskah Bahasa Arab dan Bahasa Arab Pegon Sunda untuk Komunikasi OFDM



Pusat Penelitian dan Penerbitan
UIN Sunan Gunung Djati Bandung