

## **PENERAPAN FILTER KATA MENGGUNAKAN METODE *STEMMING* PADA APLIKASI *CHATTING* BERBASIS WEB**

**Nurul Chafid<sup>1)</sup>, Luqman Mujiyanto<sup>2)</sup>, Ichsan Nuur Irmansyah<sup>3)</sup>**  
<sup>1,2,3)</sup> Teknik Informatika Fakultas Teknik Universitas Satya Negara Indonesia  
Email: chafid09@gmail.com<sup>1)</sup>, luqman.mujiyanto@gmail.com<sup>2)</sup>,  
ichsanni99@gmail.com<sup>3)</sup>

### **ABSTRAK**

Berawal dari dunia era millennium pada tahun 2000 pada masa teknologi yang semakin canggih mulai dari *sms*, *telephone* dan percakapan melalui media online merupakan sarana terpenting dalam kebutuhan hidup sehari-hari apalagi di zaman sekarang ini, media percakapan elektronik yang diciptakan oleh para ahli teknologi tersebut salah satunya adalah MIRC, Yahoo Messenger, WhatsApp dan lain sebagainya. Ada banyak macam informasi yang dapat diterima oleh seseorang dalam kegiatan sehari-hari yang bersumber dari medsos atau sarana *chatting*, tentunya informasi yang diinginkan adalah informasi yang berkualitas bagi si penerimanya, informasi berkualitas itu sendiri merupakan informasi yang mengandung unsur-unsur yang dapat enak dibaca akan tetapi faktanya telah terbukti bahwa, informasi yang didapat kenyataannya kurang baik, dari beberapa aplikasi *chat* yang dipakai dimasyarakat khususnya para remaja dan orang dewasa setelah penulis melakukan analisa dari berbagai sekian banyak orang, saya selaku penulis dan kebetulan termasuk salah satu pengguna aplikasi *chat* sebelum era masa android sekarang ini banyak sekali kata-kata atau gambar yang kurang layak dilihat atau dibaca karena khawatir, misalkan saja informasi yang mengarah pada informasi yang mengandung kata seronok, mengirim gambar porno dan lain sebagainya, hal ini sangat mengganggu bagi seseorang khususnya anak-anak yang belum siap menerimanya secara mental, kalangan remaja atau bahkan dewasa sekalipun sebetulnya buat mereka suatu hal yang biasa akan tetapi ini dapat merusak generasi muda yang akan datang umumnya bisa saja terjadi pada kalangan anak sekolah dasar, menengah, atau sekolah tinggi atau kalangan umumnya yang dapat terjadi dimasyarakat. Tujuan dari penelitian ini adalah membuat suatu Aplikasi berbasis *web* guna memfilter kata pada aplikasi *chatting*, Metode yang digunakan adalah Metode *Stemming* yaitu metode yang dapat menghilangkan *kata* dengan porter guna meniadakan *suffix* dan *order prefix* dari suatu kata menjadi tidak bermunculan. Hasil yang didapatkan adalah Filter kata yang tidak seharusnya disebutkan pada aplikasi *chatting* sehingga pengguna tidak lepas kontrol dalam penggunaannya, kemudian luaran yang diharapkan oleh selaku penulis kedepannya aplikasi ini dapat diterapkan di masyarakat dengan pengembangan ke arah android dan dapat dipublish dalam sebuah jurnal ilmiah (*proceeding*).

**Kata Kunci** : Filter Kata, *Chatting*, *Stemming*.

### **I Pendahuluan**

Berawal dari dunia era millennium pada tahun 2000 pada masa teknologi yang semakin canggih mulai dari *sms*, *telephone* dan percakapan melalui media online merupakan sarana terpenting dalam kebutuhan hidup sehari-hari apalagi dizaman sekarang ini, media percakapan elektronik yang diciptakan oleh para ahli teknologi tersebut salah satunya adalah MIRC, Yahoo Messenger, WhatsApp dan lain sebagainya. Ada banyak macam informasi yang dapat diterima oleh seseorang dalam kegiatan sehari-hari yang bersumber dari medsos atau sarana *chatting*, tentunya informasi yang diinginkan adalah informasi yang berkualitas bagi si penerimanya, informasi berkualitas itu sendiri merupakan informasi yang mengandung unsur-unsur yang dapat enak dibaca akan tetapi faktanya telah terbukti bahwa, informasi yang didapat kenyataannya kurang baik, dari beberapa aplikasi *chat* yang dipakai dimasyarakat khususnya para remaja dan orang dewasa setelah penulis melakukan analisa dari berbagai sekian banyak orang, saya selaku penulis dan kebetulan termasuk salah satu pengguna aplikasi *chat* sebelum era masa android sekarang ini banyak sekali kata-kata atau gambar yang kurang layak dilihat atau dibaca karena khawatir, misalkan saja informasi yang mengarah pada informasi yang mengandung kata seronok, mengirim gambar porno dan lain sebagainya, hal ini sangat mengganggu bagi seseorang khususnya anak-anak yang belum siap menerimanya secara mental, kalangan remaja atau bahkan dewasa sekalipun sebetulnya buat mereka suatu hal yang biasa akan tetapi ini dapat merusak generasi muda yang akan datang umumnya bisa saja terjadi pada kalangan anak sekolah dasar, menengah, atau sekolah tinggi atau kalangan umumnya yang dapat terjadi dimasyarakat.

Mengutip dari penelitian terdahulu menurut Rinci KH dan Yunus JS (dalam Jurnal Prosiding MMT-ITS, Surabaya, 24 Januari 2015) mengatakan bahwa penggunaan *Stemming* ini dilakukan untuk mencari kata-kata dasar dalam bahasa Indonesia dengan pencarian kata berimbuhan, dengan *stemming* melakukan prosesnya dengan tiga tahap yaitu tahap pertama, *parsing* dokumen dilakukan untuk memecah dokumen menjadi kata-kata atau disebut token, sedangkan tahap ke dua peneliti melakukan *stopword* yaitu penghilangan kata-kata yang dianggap tidak penting dalam dokumen kemudian melalui tahap ke tiga, *confix stripping* ini merupakan konsep dari suatu algoritma *stemming* guna pemenggalan kata-kata yang berimbuhan awalan dan akhiran bergabung menjadi sebuah kata dasar, hasil pengujian dari *stemming* tersebut memiliki nilai 94,8%. Sedangkan menurut sumber penelitian lainnya pendapat dari A. Yudi Permana dalam jurnal Teknologi Pelita Bangsa (SIGMA-Vol.7 No.1 September 2017) telah melakukan perbandingan dalam penggunaan Algoritma yang terdapat dalam *Stemming* dibandingkan dengan Algoritma Bayes guna mendapatkan hasil akurasi terhadap soal UN dengan mengklasifikasikan topik-topik yang terdapat dalam soal ujian tersebut dengan membagi kelas pengklasifikasian kata yang sama atau yang sering muncul. Sedangkan menurut Lasmedi Afuan dalam (jurnal Telematika Vol.6 No.2 Agustus 2013) melakukan bahwa dari apa yang diperoleh dengan menggunakan teknik STKI pada pencarian dokumen atau informasi agar lebih mudah dan efisien. Oleh sebab itu dari permasalahan diatas penulis mencoba membuat suatu terobosan setelah membandingkan dari peneliti sebelumnya bahwa *stemming* hanya digunakan untuk menentukan pemfilteran kata dalam bahasa Indonesia untuk itu berdasarkan pembaruan yang diperoleh penulis mencoba suatu

terobosan baru dengan membuat sebuah software aplikatif dan belum pernah ada digunakan pada suatu aplikasi *chat* sebelumnya dengan metode .

## II STUDI PUSTAKA

Mengutip dari penelitian terdahulu menurut Rinci KH dan Yunus JS (dalam Jurnal Prosiding MMT-ITS, Surabaya, 24 Januari 2015) yang berjudul "*Stemmin Artikel Berbahasa Indonesia Dengan Pendekatan Confix-Stripping*" mengatakan bahwa penggunaan *Stemming* ini dilakukan untuk mencari kata-kata dasar dalam bahasa Indonesia dengan pencarian kata berimbuhan, dengan stemming melakukan prosesnya dengan tiga tahap yaitu tahap pertama, *parsing* dokumen dilakukan untuk memecah dokumen menjadi kata-kata atau disebut token, sedangkan tahap ke dua peneliti melakukan *stopword* yaitu penghilangan kata-kata yang dianggap tidak penting dalam dokumen kemudian melalui tahap ke tiga, *confix stripping* ini merupakan konsep dari suatu algoritma *stemming* guna pemenggalan kata-kata yang berimbuhan awalan dan akhiran bergabung menjadi sebuah kata dasar, hasil pengujian dari stemming tersebut memiliki nilai 94,8%. Sedangkan menurut sumber penelitian lainnya pendapat dari A. Yudi Permana dalam jurnal Teknologi Pelita Bangsa (SIGMA-Vol.7 No.1 September 2017) dengan penelitiannya yang berjudul "Impelementasi *Stemming Porter* KBBI Untk Klasifikasi Topik Soal UN Bahasa Indonesia Menggunakan Algoritma Naïve Bayes" telah melakukan perbandingan dalam penggunaan Algoritma yang terdapat dalam *Stemming* dibandingkan dengan Algoritma Bayes guna mendapatkan hasil akurasi terhadap soal UN dengan mengklasifikasikan topik-topik yang terdapat dalam soal ujian tersebut dengan membagi kelas pengklasifikasian kata yang sama atau yang sering muncul. Sedangkan menurut Lasmedi Afuan dalam (jurnal Telematika Vol.6 No.2 Agustus 2013) yang berjudul "*Stemming Dokumen Teks Bahasa Indonesia Menggunakan Algoritma Porter*" mengatakan bahwa dalam penelitiannya penulis bahwa dari apa yang diperoleh dengan menggunakan teknik STKI pada pencarian dokumen atau informasi agar lebih mudah dan efisien, sehingga aplikasi yang dibuat penulis masih harus dikembangkan agar dapat membaca berbagai dokumen. Dari pemaparan yang terdapat dalam masing-masing peneliti metode *stemming* setelah diamati belumlah menemukan kelebihan dari apa yang mereka buat dan kebanyakan metode tersebut digunakan untuk penentuan kosa kata saja untuk itu saya selaku penulis ingin mencoba metode tersebut ke dalam mesin aplikasi *chat* yang tujuannya adalah guna memfilter kata-kata tersebut. *Stemming is a heuristic process that "chops off" the ends of words in the hope of achieving the goal correctly most of the time* (Ceri,Bozzon, Brambilla, Valle, Fraternali dan Quarteroni, 2013:18). Ada banyak algoritma yang digunakan untuk melakukan proses stemming, khususnya *stemming* bahasa indonesia, diantaranya adalah Algoritma Nazief dan Adriani. Algoritma ini mengacu pada aturan morfologi bahasa Indonesia yang mengelompokkan imbuhan, yaitu imbuhan yang diperbolehkan atau imbuhan yang tidak diperbolehkan. Pengelompokan ini termasuk imbuhan di depan (awalan), imbuhan kata di belakang (akhiran), imbuhan kata di tengah (sisipan) dan kombinasi imbuhan pada awal dan akhir kata (konfiks). Algoritma ini menggunakan kamus kata keterangan yang digunakan untuk mengetahui bahwa proses stemming telah mendapatkan kata dasar. Aturan morfologi Bahasa Indonesia mengelompokkan imbuhan ke dalam beberapa kategori sebagai berikut:

1. *Inflection suffixes* yakni kelompok akhiran yang tidak merubah bentuk kata dasar. Sebagai contoh, kata "duduk" yang diberikan akhiran "-lah" akan menjadi "duduklah". Kelompok ini dapat dibagi menjadi dua:
  - i. *Particle* (P) atau partikel yakni termaksud di dalamnya "-lah", "kah", "tah" dan "pun".
  - ii. *Possessive pronoun* (PP) atau kata ganti kepunyaan, termaksud di dalamnya "-ku", "-mu" dan "-nya".
2. *Derivation suffixes* (DS) yakni kumpulan akhiran asli Bahasa Indonesia yang secara langsung ditambahkan pada kata dasar yaitu akhiran "-i", "-kan", dan "-an".
3. *Derivation prefixes* (DP) yakni kumpulan awalan yang dapat langsung diberikan pada kata dasar murni, atau pada kata dasar yang sudah mendapatkan penambahan sampai dengan 2 awalan. Termaksud di dalamnya adalah:
  - i. Awalan yang dapat bermorfologi ("me-", "be-", "pe-" dan "te").
  - ii. Awalan yang tidak bermorfologi ("di-", "ke-" dan "se-").

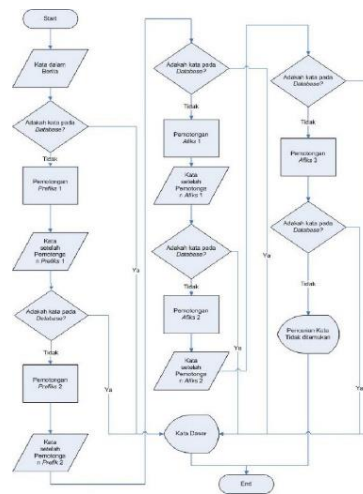
Berdasarkan pengklasifikasi imbuhan-imbuhan di atas, maka bentuk kata berimbuhan dalam Bahasa Indonesia dapat dimodelkan sebagai berikut:

**[DP+ [DP+ [DP+] ] ] Kata Dasar [ [+DS] [+PP]**

Gambar 1. Model Kata Berimbuhan dalam Bahasa Indonesia

### III METODE PENELITIAN

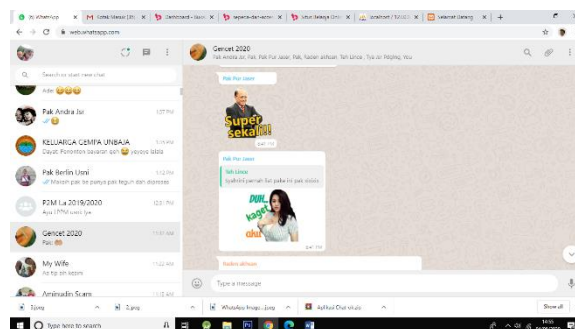
Objek penelitian yang penulis lakukan adalah terhadap pengguna dimasyarakat khususnya para remaja, dewasa dan mahasiswa di lingkungan sekitar kampus Universitas Satya Negara Indonesia atau dari luar kampus yang akan menjadi responden penelitian guna kelangsungan pembuatan aplikasi chat berbasis web dengan metode *stemming*. Sample yang akan diambil sebanyak 50 atau dengan 100, dengan 30 sample kata jorok atau yang sekiranya dapat mewakili. Metode yang dilakukan oleh saya selaku peneliti adalah dengan menggunakan Metode Stemmning yang dipadukan dengan algoritma lain dimana teknik ini dilakukan dalam setiap pengkodean dengan menyortir kata-kata yang tak pantas sehingga pada proses implementasi saat aplikasi digunakan akan terkontrol dengan baik ketika dijalankan dapat lebih mudah untuk mengetahui apakah aplikasi menemukan debug atau tidak. Berikut bentuk FOD (*Flow of Data*). Dalam FOD ini digambarkan bagaimana proses dalam aktifitas Aplikasi *Chat* dari awal sampai akhir dalam proses pengelolaan kata dengan metode *stemming*, sebagai berikut :



Gambar 2. Alur Flow Data Penggunaan Aplikasi *Chat* dengan *Stemming*

#### IV ANALISA HASIL DAN PEMBAHASAN

Dari analisa yang didapat penulis melihat apa yang ada dalam tampilan program Aplikasi *chat* seperti WA dan yang lainnya yang berjalan masih banyak kekurangan-kekurangan yang belum terpenuhi. Prosedur yang dilakukan dalam aplikasi chat tersebut sama saja kinerjanya dengan proses manual atau ibaratnya mulut manusi kalo berbicara gak pake rem atau bablas seenaknya, kemudian *user lain* juga dapat melihat dan menerima gambar katalog atau text kata-kata jorok yang dikirim secara seronok sehingga dalam aplikasi itu ketika akan melakukan pembalasan *chat* bersama teman atau lawan tidak akan terjadi filter kata yang diharapkan, dengan demikian usulan yang diharapkan dalam membuat Aplikasi *Chat* dapat menyempurnakan kekurangan-kekurangan yang ada pada Aplikasi lainnya seperti WA dan Aplikasi-aplikasi yang tidak bisa menyaring kata atau gambar yang jorok. Berikut bentuk dari rancangan program tampilan sistem informasi pada layar WA (WhatsApp) atau contoh lain yang sedang berjalan:

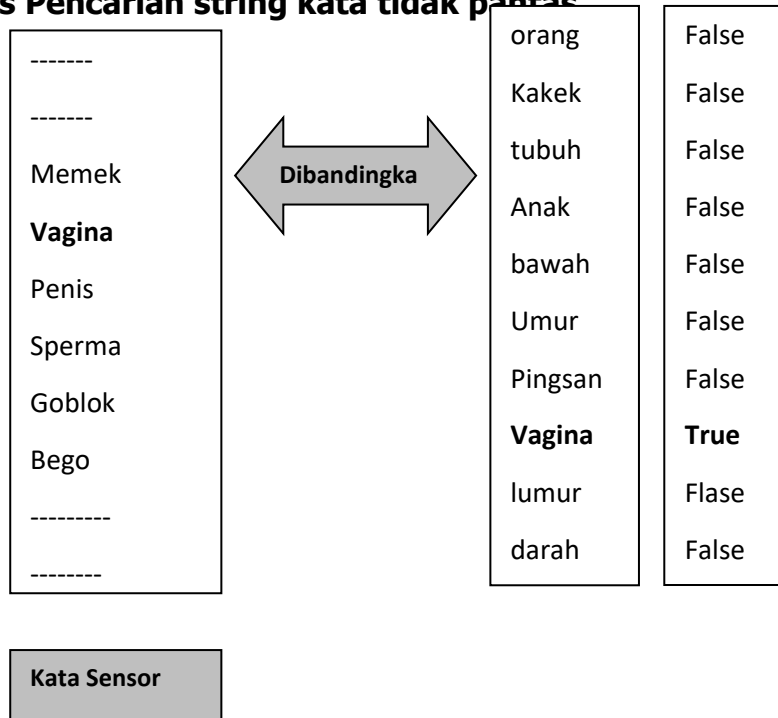


Gambar 3. Tampilan WA (Aplikasi *Chat*) dengan gambar kurang pantas

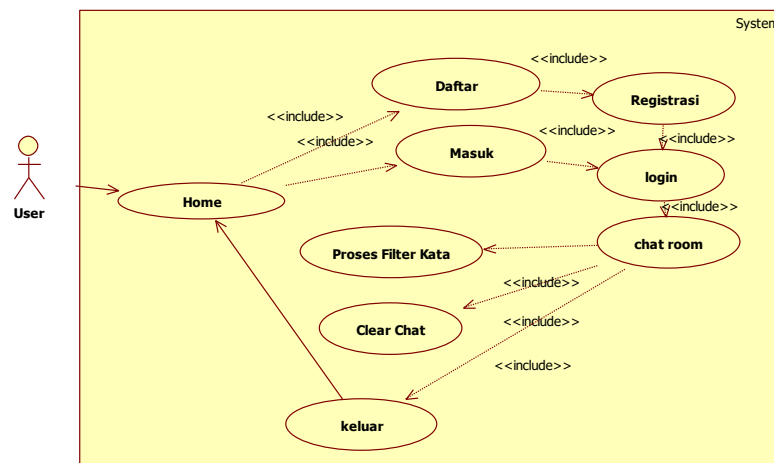
## Proses Stemming

Contoh kalimat : Seorang kakek menyetubuhi anak dibawah umur sampai pingsan dengan vagina yang berlumuran darah, jadi dalam penelitian ini penulis membuat aplikasi chat hanya dapat memfilter kosa kata atau ucapan yang diketik pada mesin *chat* berbasis *web* yang dibuat adapun proses gambarannya seperti dibawah ini:

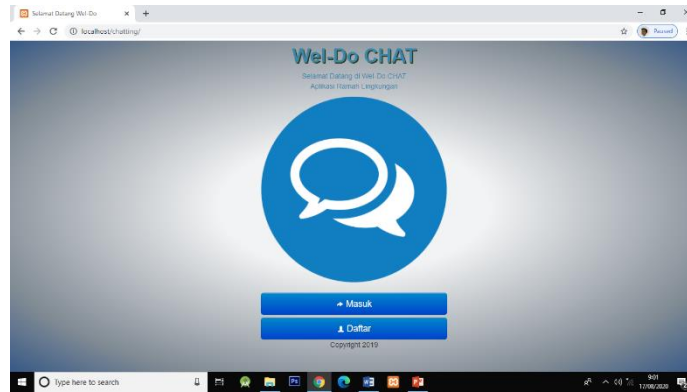
## Proses Pencarian string kata tidak pantas



Gambar 4. Pencarian string kata sensor



Gambar 5. Use case diagram Aplikasi Chat



**Gambar 6.** Tampilan Halaman *Chat*

### **Daya Tampung dan Kesuksesan Aplikasi *Chat***

Dilihat dari segi perspektif pengguna kemampuan dari Aplikasi *chat* ini, memiliki daya tampung bisa lebih dari 50 pengguna bahkan bisa 100 *user*. Dalam aplikasi *chat* ini penulis membuat dengan sederhana dan lebih tepatnya dapat digunakan dalam sebuah ruangan kerja atau dengan kata lain *chat room*, dengan demikian berdasarkan analisa dan hasil kuisisioner dengan kesuksesan yang didapat 83% sampai dengan 85%, aplikasi ini setelah diuji cobakan dari tingkat kemampuan dan keberhasilan dalam melayani *chat* sampai dengan menolak kata-kata yang tidak pantas dikeluarkan dari pengguna sistem sudah mampu menyensor kata dan ternyata alhasil sistem aplikasi dapat bekerja dengan baik, walaupun masih membutuhkan tahap pengembangan selanjutnya.

## **V KESIMPULAN**

Dari beberapa kesimpulan yang diambil berdasarkan uji kelayakan terhadap sebaran kuisisioner dan uji kelayakan sistem dengan white box-blackbox dari penelitian yang telah dilakukan oleh penulis, adalah ternyata dari setiap penerapan sistem aplikasi chat WhatsApp dan lain-lain, penulis mengidentifikasi masalah pada sistem yang berjalan ada beberapa banyak kekurangan setelah melihat kondisi dilapangan yang ada pada sistem tersebut baik dari segi model maupun Aplikasi chat yang digunakan oleh kebanyakan orang. Dalam menentukan kebutuhan sistem bukan suatu hal yang mudah akan tetapi kita harus tahu siapa yang akan berperan, apa yang akan dilakukan setelah sistem itu dibuat, kemudian kapankah sistem itu akan digunakan tergantung waktu, sehingga penulis melakukan perancangan serta implementasi dalam mengembangkan sistem *aplikasi chat berbasis web* untuk diujikan kepada pengguna dengan menggunakan metode *algoritma stemming* yang memiliki beberapa tahapan yang merupakan hal yang terpenting dalam membuat sebuah rancangan dari awal analisa sampai dengan hasil yang sesuai dengan apa yang di inginkan oleh pengguna. Hal ini berdasarkan beberapa kuisisioner atau permodelan rancangan *UML* yang dilakukan dalam tahapan analisa data secara kualitatif dan kuantitatif dari mulai mewawancarai, observasi sampai ke tahapan selanjutnya memperoleh hasil 83% sampai dengan 85% tanggapan dari user terhadap sistem aplikasi *chat* yang dibuat. Dan cara kerja sistem ini memiliki kelebihan dalam memfilter kata jorok

yang tidak dimiliki oleh aplikasi lain seperti WhatsApp dan aplikasi lainnya. Penulis belum begitu yakin sepenuhnya bahwa dalam melakukan penelitian dan membuat sebuah rancangan ini belum begitu baik dan sempurna dalam mengimplementasikan sistem *Aplikasi Chat*, oleh sebab itu perlu pengembangan ke arah yang lebih baik dan akan dikembangkan dalam penelitian selanjutnya dengan metode-metode yang mendukung dalam penelitian ini.

### 5.1 Saran

Adapun saran-saran dari penulis sesuai dengan pengetahuannya penulis adalah sebagai berikut :

1. Untuk Penelitian pada tahap selanjutnya dapat dikembangkan dengan kombinasi dengan Metode Information Retrieval yang lain sehingga filter kata sensor dapat lebih akurat dan cepat.
2. Pada penelitian ini hanya memfilter kata saja, diharapkan penelitian selanjutnya dapat memfilter tidak hanya kata saja, tetapi kalimat, gambar dan suara sehingga pengguna akan lebih nyaman lagi dengan penggunaan aplikasi chatting.

### DAFTAR PUSTAKA

- Adi Nugroho, 2009, " Rekayasa Perangkat Lunak Menggunakan UML dan JAVA", Yogyakarta : Andi Offset.
- A. Yudi Permana, Implementasi Stemming Porter KBBI Untuk Klasifikasi Topik Soal Ujian Nasional Bahasa Indonesia Menggunakan Algoritma Naïve Bayes, SIGMA-Jurnal Teknologi Pelita Bangsa, Volume 7 No.1 September 2017.
- B. Sakur, Stendy . 2010. "PHP 5 Pemrograman Berorientasi Objek". Yogyakarta : Andi Offset.
- Bunafit Nugroho, 2008, "Panduan Lengkap Menguasai Perintah SQL". Jakarta: Media Kita.
- Bunafit Nugroho, 2019, "Aplikasi Pemrograman WEB Dinamis dengan PHP dan MySQL", Gaya Media.
- Ceri, Stefano., Alessandro, Bozzon., Brambilla, Marco., D. Valle, Emanuele., Fraternali dan Quarteroni, Silvia 2013. " Web Information Retrieval", Italy: Springer
- D.Manning, Christopher., Raghavan, Prabhakar., Schütze, Hinrich . 2008. "Introduction To Information Retrieval", New York: Cambridge University Press.
- Dr. Widodo Budiharto, "Knowledge Information Retrieval". Deepublish; 2016. ISBN: 978-602-401-418-6
- Janner Simarmata, 2010. "Rekayasa Web Analisis dan Desain Sistem " Yogyakarta : Andi Offset.
- Lasmedi Afuan, Stemming Dokumen Teks Bahasa Indonesia Menggunakan Algoritma Porter, Jurnal Telematika Vol.6 No.2 Agustus 2013.

- Miftahul Firodh Fatroni, "Kecerdasan Buatan Dalam Program *Chatting* Untuk Merspon Emosi dari Teks Berbahasa Indonesia Menggunakan Teks Mining dan Naïve Bayes; Jurnal ITS-Surabaya.
- Musliadi K.H, 2013, " Tips Cara Cepat Bekerja dengan Database", Yogyakarta : Andi Offset.
- Ni Luh Ratniasih, Made Sudarma, Nyoman Gunantara, "Penerapan *Text Mining* dalam *Spam filtering* Untuk Aplikasi *Chatt*", Jurnal Teknologi Elektro; Volume 16, No.3 September – desember 2017.
- Pradana Nur Prasetya, Hero Wintolo, Nurcahyani Dewi Retnowati, "Perancangan Aplikasi *Chatt Translator* Berbasis Desktop Untuk Komunikasi Dua Bahasa dalam Jaringan Komputer", Jurnal Compiler; Volume 3 No.1 Mei 2014.
- Rosa A. S, M. Shalahuddin, 2014, "Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek", Informatika Bandung.
- Roni Setiawan, Edhy Sutanta, "Membangun Aplikasi *Chatting* Berbasis Multiuser" Jurnal DASI; Volume 10 No.1 Maret 2009.
- Rinci Kembang Hapsari dan Yunus Juli Santoso, Stemming Artikel Berbahasa Indonesia Dengan Pendekatan *Confix-Stripping*.*Prosiding Seminar Nasional Manajemen Teknologi ITS-Surabaya*; Januari 2015.