

PENERAPAN ALGORITMA BAYES TEOREMA PADA APLIKASI PENGEJAN KATA

Marta Lenah Haryanti, Dwi Fegiannata
mharyanti@bundamulia.ac.id, dwifegiannata@yahoo.co.id
Teknik Informatika Universitas Bunda Mulia

Abstract

Mesin pencari kata menjadi salah satu fasilitas yang harus disediakan bagi setiap situs web. Pada mesin pencari, kesalahan dapat saja terjadi saat pengguna memasukkan kata kunci yang akan dicari. Baik karena kesalahan penulisan bahasa asing atau pengejaan kata kunci yang salah.

Penelitian ini menerapkan algoritma Bayes Teorema untuk memperbaiki kesalahan yang terjadi. Data yang akan diambil sebagai referensi pengkoreksian kata kunci akan disimpan ke dalam sebuah dokumen teks. Kata yang berada dalam dokumen tersebut akan diambil beberapa sebagai sampel, kemudian huruf dalam kata tersebut akan diacak untuk menguji tingkat pengkoreksian kata dari algoritma Bayes Teorema sesuai dengan yang diinginkan oleh pengguna.

Kata Kunci: mesin pencari, koreksi, ejaan, kata

PENDAHULUAN

Search Engine atau mesin pencarian sudah menjadi salah satu fasilitas yang harus disediakan bagi setiap situs web. Sebagai salah satu contoh adalah situs web *Social Media* yang terus berkembang yang dapat melakukan pencarian melalui jenis data seperti nama, pendidikan, tempat tinggal, dll. Contoh lainnya adalah situs *e-Commerce* agar konsumen dapat lebih mudah melakukan pencarian jenis data produk/jasa yang ditawarkan.

Pada mesin pencari, kesalahan dapat saja terjadi saat pengguna memasukkan kata kunci yang akan dicari. Baik karena kesalahan penulisan bahasa asing atau pengejaan kata kunci yang salah.

Dengan menerapkan algoritma Bayes Teorema, kesalahan tersebut dapat diminimalisir dengan cara, mesin pencari akan memberikan pengkoreksian kata yang benar melalui kata yang hampir mirip dengan kata kunci yang dimasukkan.

LANDASAN TEORI

Bayes Teorema adalah adalah rumus matematika sederhana yang digunakan untuk menghitung probabilitas bersyarat. Hal ini terlihat dalam pendekatan subjektif atau Bayesian untuk epistemologi, statistik, dan logika induktif. Subjektif, yang mempertahankan bahwa keyakinan rasional diatur oleh hukum-hukum probabilitas,

bersandar berat pada probabilitas bersyarat dalam bukti teori dan model belajar empiris. Bayes Teorema adalah menyederhanakan perhitungan probabilitas bersyarat dan karena hal itu akan menjelaskan fitur yang signifikan dari posisi subjektif [dikutip dari <http://plato.stanford.edu/entries/bayes-theorem/> diakses pada tanggal 28 Juni 2014].

$$\operatorname{argmax}_c P(w|c) P(c) / P(w)$$

Gambar 1. Bayes Teorema

[dikutip dari <http://norvig.com/spell-correct.html>]

Berikut merupakan penjelasan dari Bayes Teorema tersebut:

1. $P(c)$, probabilitas bahwa koreksi yang diusulkan c berdiri sendiri. yang disebut sebagai *model language*.
2. $P(w | c)$, probabilitas bahwa w adalah teks yang akan diketik oleh pengguna. Sedangkan c adalah model error.
3. Argmax_c adalah mekanisme kontrol yang diproses untuk menghitung semua nilai layak dari c dan kemudian memilih salah satu yang memberikan nilai probabilitas gabungan terbaik.

SIMULASI

Simulasi uji coba algoritma Bayes Teorema dengan menggunakan data yang diambil melalui situs <http://millercenter.org/president/lincoln/esays/biography/2> [diakses tanggal 18 juni 2014]. Berikut merupakan hasil uji cobanya :

TABEL 1 Uji coba kata have

Hasil yang diharapkan		<i>have</i>
Masukan kata kunci	Hasil yang didapat	Waktu(ms)
hav	had	21
hae	he	35
hve	he	22
habe	abe	23
ave	abe	21
has	his	29
havi	have	22
hase	have	37

Dalam dokumen teks, kata **have** terdapat sebanyak 5 buah, kata **had** sebanyak 17 buah, kata **he** sebanyak 31 buah, kata **abe** sebanyak 13 buah, dan kata **his** sebanyak 58 buah.

TABEL 2 Uji coba kata having

Hasil yang diharapkan		<i>having</i>
Masukan kata kunci	Hasil yang didapat	Waktu(ms)
havin	having	21
havng	having	23
havig	having	27
hving	having	23
aving	having	29
havi	have	22
hvng	having	60
avig	whig	59

Dalam dokumen teks, kata **having** terdapat sebanyak 2 buah, kata **have** sebanyak 5 buah dan kata **whig** sebanyak 12 buah.

TABEL 3 Uji coba kata political

Hasil yang diharapkan		<i>political</i>
Masukan kata kunci	Hasil yang didapat	Waktu(ms)
politica	political	117
politca	political	97
plitcal	political	90
polical	political	93
litical	political	95
pliical	political	98
palitica	political	110
poletocal	political	126

Dalam dokumen teks, kata **political** terdapat sebanyak 11 buah.

TABEL 4 Uji coba kata lincoln

Hasil yang diharapkan		<i>lincoln</i>
Masukan kata kunci	Hasil yang didapat	Waktu(ms)
licoln	lincoln	22
lcoln	lincoln	79
lacoln	lincoln	82
lncln	lincoln	93
loncaln	lincoln	97
localn	local	36
lincl	lincoln	68
licnol	lincoln	85

Dalam dokumen teks, kata **lincoln** terdapat sebanyak 14 buah dan kata **local** sebanyak 8.

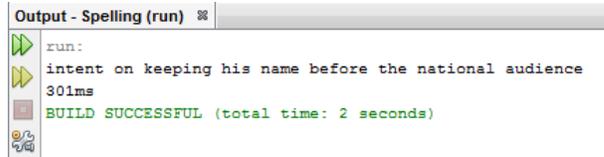
Uji Coba Kalimat

Kata yang dimasukkan =

intent on keping his name befre the national audence

Maka hasil yang didapatkan adalah

intent on keeping his name before the national audience



```
Output - Spelling (run)
run:
intent on keeping his name before the national audience
301ms
BUILD SUCCESSFUL (total time: 2 seconds)
```

Gambar 2. Hasil eksekusi kalimat pertama

Kata yang dimasukkan =

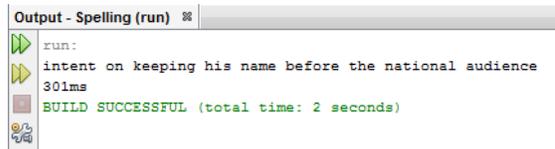
to all men in ther freedom to ear
the just rewards

Maka hasil yang didapatkan adalah

to all men in the freedom to war
the just rewards

Sedangkan hasil yang diharapkan adalah

to all men in their freedom to earn
the just rewards



```
Output - Spelling (run)
run:
intent on keeping his name before the national audience
301ms
BUILD SUCCESSFUL (total time: 2 seconds)
```

Gambar 4.7 Hasil eksekusi kalimat kedua

Dalam basis data, kata **their** terdapat sebanyak 7 buah, sedangkan kata **the** sebanyak 195 buah dan kata **earn** sebanyak 3 buah, sedangkan kata **war** terdapat sebanyak 13 buah.

SIMPULAN

Setelah melakukan uji coba kata dan kalimat, maka dapat diambil kesimpulan sebagai berikut :

- Pengkoreksian kata yang muncul adalah kata yang memiliki jumlah terbanyak.
- Apabila jumlah kata yang dicari memiliki jumlah yang sama, kata yang muncul adalah yang memiliki alphabet dalam urutan *ascending*.
- Pengkoreksian dalam kalimat tidak dapat digunakan karena pengecekan dilakukan melalui kata per kata, bukan kemiripan kalimat.

Saran

Adanya kekurangan dalam penelitian ini mungkin saja terjadi karena penerapan aplikasi dengan menggunakan bahasa pemrograman, fungsi, prosedur yang berbeda. Disarankan untuk membandingkan kembali dengan menggunakan beberapa macam bahasa pemrograman dan data yang berbeda.

DAFTAR PUSTAKA

- [1] <http://millercenter.org/president/lincoln/essays/biography/2> [diakses tanggal 18 juni 2014].
- [2] <http://norvig.com/spell-correct.html> [diakses tanggal 15 juni 2014].
- [3] http://www.ted.com/speakers/peter_norvig.html [diakses tanggal 15 juni 2014].
- [4] Burd, Barry, 2012, **Begining Programming with Java For Dummies**, 3rd Ed, Wiley, Indianapolis.