SWABUMI, Vol.5 Maret 2017, pp. 84-89
ISSN : 2355-990X
E-ISSN : 2549-5178

IMPLEMENTASI ALGORITMA RABIN KARP DAN STEMMING NAJIEF ANDRIANI UNTUK DETEKSI PLAGIARISME DOKUMEN

Satia Suhada [1], Saeful Bahri [2]
STMIK Nusa Mandiri Jakarta[1,2]
Jl. Damai No. 8, Warung Jati Barat
[1]satia.shq@nusamandiri.ac.id
[2]saeful.sel@nusamandiri.ac.id

Abstrak

Plagiarisme merupakan pengambilan karya orang lain kemudian menjadikannya nampak seperti karya sendirii. Praktek plagiat sering terjadi didunia akademis seprti pada dokumen proyek akhir. Untuk meminimalkan tindakan plagiarisme, diperlukan suatu sistem untuk menilai atau mengukur seberapa banyak kemiripan dalam sebuah document sepeti tugas akhir dan makalah ilmiah beberapa algoritma telah digunakan dalam proses deteksi plagiarism namun pada penelitian ini akan menerapkan . Algoritma Rabin Karp dalam mendeteksi plagiarisme karena algoritma ini terbukti efektif untuk membandingkan patern-patern yang ada pada sebuah essai dengan menggunakan fungsi hashing yang dapat menemukan bentuk-bentuk / pola dalam teks, untuk lebih meningkatkan keakuratan dalam proses penemuan pola pada sebuah teks kemudian digunakan algoritma Steeming Najief Andriani yang dapat menemukan kata-kata yang setara yang memiliki persamaan kata dasar yang sama, selanjutnya kedua algoritma yang digunakan akan diimplemntasikan kepada sebuah aplikasi web untuk pengujian document tugas akhir.

***Keywords**: Plagiarisme, Algoritma Rabin karp, Steeming Najief andriani*

1. Pendahuluan

Plagiarisme atau plagiat adalah penjiplakan atau pengambilan karangan, pendapat orang lain dan menjadikannya seolah-olah karangan sendiri (Kbbi:2014). merupakan sebuah aktifitas yang menjadi perhatian dibanyak sektor terutama sektor akademis (Salmuasih:2013), Kategori plagiarisme berdasrkan cara menggunakanya adalah

1. *Copy* dan *paste plagiaisme* tanpa merubah kata sedikitpun
2. *Disguised plagiarisme* proktek menutupi bagaian yang disalin
3. tecnical disguis teknik meringkas untuk menyembunyikan konten plagiat dari deteksi otomatis
4. *Undue paraphrasing*, sengaja menuliskan ulang pemikiran asing dengan pemilihan kata dan gaya plagiator dengan menyembunyikan sumber asli.
5. Translated plagiarism, mengkonversi konten dari satu bahasa ke bahasa lain.
6. Idea plagiarism, menggunakan ide asing tanpa menyatakan sumber.
7. *Self* plagiarism, penggunaan sebagian atau keseluruhan tulisan pribadi yang tidak dibenarkan secara ilmiah(Salmuasih:2013).

Untuk meminimalisir tindak plagiarisme pada penulisan tugas akhir perlu adanya sebuah mekanisme yang mampu mengukur tingkat kemiripan suatu dokumen tugas akhir, konsep yang diterapkan dalam proses pengecekan kemiripan suatu dokumen salah satu cara mendeteksi plagiarime seperti copy paste plagiarime dan disguised plagiarism . menggunakan metode finger printing menggunakan fungsi hash menggunakan algoritma rabin karp sebagai string matching memiliki kelebihan dalam proses pencocokan string dilakukan secara sederhana dengan mencari term dari masing-masing nilai hash pad setiap text yang diuji (Haytam.et,al,:2013). Untuk pengaplikasian text minning untuk tahap preposesing menggunakan algoritma najif andriani yang memiliki kelebihan dan presisi lebih besar dibanding algoritma steming seperti steming porter (Wacana, et.al,:2014).

Pada penelitian ini algoritma rabinkarp akan diterapkan untuk mendeteksi tingkat kemiripan suatu dokumen tugas akhir sedangkan algoritma stemming nazief

andriani diterapkan untuk mencari root kata atau kata dasar dari tiap-tiap kalimat pada tugas akhir yang akan diuji tingkat kemiripanya

2. Metode Penelitian

Pada penelitian ini dilkuakn beberapa tahapan penelitian seperti studi litelature menegenai algoritma rabin karp dan steming najief andriani berikut adalah beberapa point dari studi litelatur yang dimaksud

1. Ekstraksi Dokumen

Ekstaksi merupakan tahapan utama dalam proses analisa teks tahapan ekstraksi teks dibagi kedalam beberapa bagian diantaranya

a. Case folding dan Tokenizing
   Case folding merupakan sebuah langkah yang merubah huruf bentuk huruf yang semula UPPER CASE ke dalam bentuk lowercase sedangkan proses tokenizing adalah proses pemisahan kalimat kedalam bentuk kata (Made, et.al:2013)
b. Stemming
   Stemming merupakan suatu proses yang terdapat dalam sistem IR yang mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya (*root word*) dengan menggunakan aturan-aturan tertentu. Sebagai contoh, kata bersama, kebersamaan, menyamai, akan distem ke *root word* nya yaitu “sama” (Agusta, et.al:2009).

2. Algoritma rabin karp

Rabin karp adalah sebuah algoritma yang diciptakan pada tahun 1987 oleh Michael O rabin dan ricard M Karp yang menggunakan fungsi hashing untuk menemukan bentuk dalam sebuah teks, algoritma Rabin Karp merepresentasikan setiap karakter ke dalam bentuk desimal digit (digit radix-d) ∑ = {0, 1, 2, 3, …, d}, dimana d = |∑|. Sehingga didapat masukan string k berturut- turut sebagai perwakilan panjang k desimal. Karakter string 31415 sesuai dengan jumlah desimal 31,415. Kemudian pola p di-hash menjadi nilai desimal dan string direpresentasikan dengan penjumlahan digit-digit angka menggunakan aturan Horner's:

{ A, B, C, ..., Z } → { 0, 1, 2, ..., 26 }
BSI → 1 + 18 + 8 = 28
NURI → 13 + 20 + 17 +8 = 56

Untuk pola yang panjang dan teks yang besar, algoritma ini menggunakan operasi mod, setelah dikenai operasi mod q, nilainya akan menjadi lebih kecil dari q.

Rumus matematis dari algoritma rabin karp

$t_{s+1} = (d\ (ts – T\ [s + 1]\ h) + T\ [s + m + 1] \bmod q$

dimana

t_s = nilai desimal dengan panjang m dari substring T [s + 1 .. s + m], untuk s = 0, 1, ..., n - m

t_{s+1} = nilai desimal selanjutnya yang dihitung dari ts

d = radix desimal (bilangan basis 10)

$h = d^{m-1}$

n = panjang teks

m = panjang pola

q = nilai modulo

Psudo code dari rumus matematis algoritma rabin karp

```
RABIN-KARP (T, P, d, q)
n = T.length
m = P.length
h = dm-1 mod q
p = 0
t0 = 0
  for i = 1 to m
      p = (dp + P[i]) mod q
     t0 = (dt0 + T[i]) mod q
 for s = 0 to n – m
     if p == ts
     if P[1 .. m] == T[s + 1 .. s + m]
      print “cetak” s
if s < n – m
       ts+1 = (d(ts - T[s + 1] h) + T[s + m + 1]) mod q.
```

3. Stemming nazief andriani

Stemming yang akan digunakan dalam sistem pendeteksi plagiarisme ialah steming najif andriani yang akan digunakan untuk preprossing teks sebelum kemudian di cari kemiripan kata algoritma najif andriani memiliki kelebihan dari segi prosentasi keakuratan (presisi) lebih besar dibanding dengan algoritma stemming porter (Agusta, et.al:2009).Tahapan pada algoritma najief dan andriani :

a. Cari kata yang akan distem dalam kamus. Jika ditemukan maka diasumsikan bahwa kata tesebut adalah root word. Maka algoritma berhenti.
b. Inflection Suffixes (“-lah”, “-kah”, “-ku”, “-mu”, atau “-nya”) dibuang. Jika berupa particles (“-lah”, “-kah”, “-tah” atau “-pun”) maka langkah ini diulangi lagi

untuk menghapus Possesive Pronouns (“-ku”, “-mu”, atau “-nya”), jika ada.

c. Hapus Derivation Suffixes (“-i”, “-an” atau “-kan”). Jika kata ditemukan di kamus, maka algoritma berhenti. Jika tidak maka ke langkah 3a a. Jika “-an” telah dihapus dan huruf terakhir dari kata tersebut adalah “-k”, maka “-k” juga ikut dihapus. Jika kata tersebut ditemukan dalam kamus maka algoritma berhenti. Jika tidak ditemukan maka lakukan langkah 3b. b. Akhiran yang dihapus (“-i”, “-an” atau “-kan”) dikembalikan, lanjut ke langkah 4.

d. Hapus Derivation Prefix. Jika pada langkah 3 ada sufiks yang dihapus maka pergi ke langkah 4a, jika tidak pergi ke langkah 4b. a. Periksa tabel kombinasi awalan-akhiran yang tidak diijinkan. Jika ditemukan maka algoritma berhenti, jika tidak pergi ke langkah 4b.

e. b. For i = 1 to 3, tentukan tipe awalan kemudian hapus awalan. Jika root word belum juga ditemukan lakukan langkah 5, jika sudah maka algoritma berhenti. Catatan: jika awalan kedua sama dengan awalan pertama algoritma berhenti.

f. Melakukan Recoding.

g. Jika semua langkah telah selesai tetapi tidak juga berhasil maka kata awal diasumsikan sebagai root word. Proses selesai.

Analisa Penelitian

a. Planning
Merencanakan pembuatan aplikasi pendeteksi plagiarisme dengan algoritma rabin karp dan steming najief Andriani.

b. Desain
Merancang algoritma rabin karp dan steming najief andriani untuk di implemntasikan ke sistem pendeteksi plagiarisme berbasis php

c. Implementasi
Uji coba sistem pendeteksi plagiarisme dengan menggunakan sample dari dokument tugas akhir mahasiswa.

1. Metode Pengumpulan data

a. Studi Kepustakaan
Membantu dalam menyusun penulisan dengan literatur-literatur yang diambil dari buku, jurnal, serta media lain yang berkaitan dengan permasalahan yang dibahas, pembahasan masalah dan pemecahan masalah.

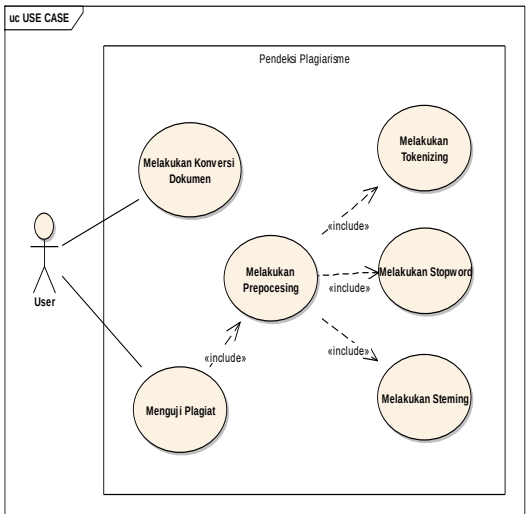
b. Observasi
Mengumpulkan data sample acak tugas akhir mahasiswa untuk kemudian di ujikan tingkat plagiarisme dari dokumen tersebut.

3. Hasil dan Pembahasan

Pada penelitian ini digunakan sebuah algoritma yang akan digunakan untuk membandingkan pola dari dari string dokumen tugas akhir. Sebelum di bandingkan string tersebut telah mengalami proses pencarian root kata menggunakan algoritma stemming najief andriani yang akan diterapkan kedalam sebuah sistem menggunakan bahasa pemrograman PHP .

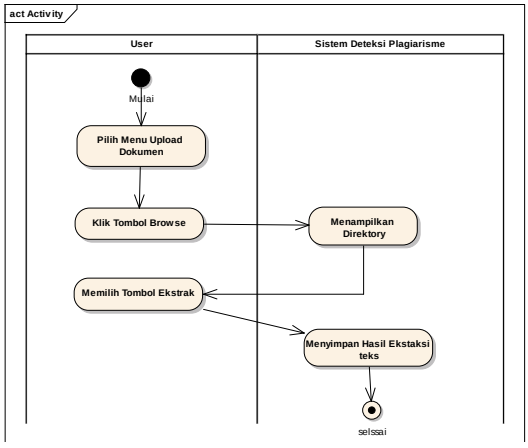
1. Rancangan Sistem deteksi plagiarisme

a. Use case



Gambar. 1. Use Case Sistem deteksi Plagiarisme

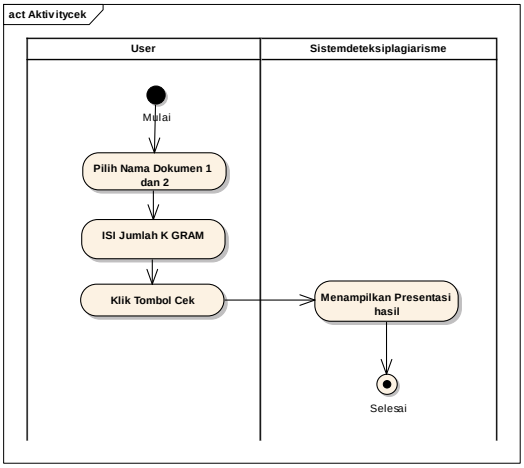
b. Aktivity ektraksi teks



Gambar. 2. Activity Extraksi Teks

ISSN : 2355-990X
E-ISSN : 2549-5178

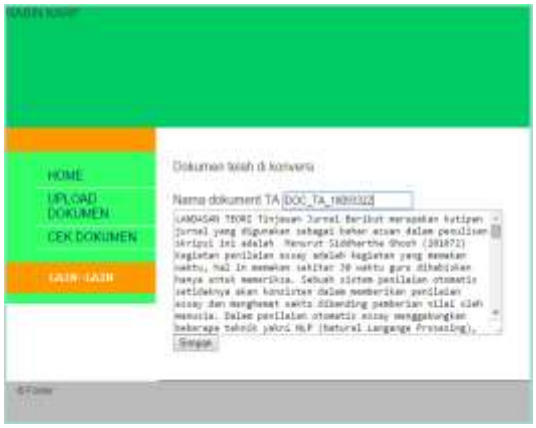
c. Aktivity pengecekan plagiarism



Gambar. 3. Activity Pengecekan plagiarisme

2. Tampilan Antar muka

Gambar. 4. Halaman Index



Gambar. 7. Halaman Konversi Doc Ke text

Gambar. 6. Hasil Cek Plagiarsme

3. Hasil Analisis

Analis dilakukan dengan melakukan pemeriksaan terhadap beberapa document atau naskah tugas akhir dari beberapa tahun kebelakang menggunakan software yang telah dirancang berikut hasil dari analisis document tugas akhir yang tersaji dalam Tabel 1.

Tabel 1. Hasil Analisis

No	Nama Dokumen	Dokumen Uji	Pembanding	Kemiripan	Kemiripan STEMMING
1	x1	Dokumen x1	Dokumen x2	33%	35%
2	x2	Dokumen x2	Dokumen x3	40%	42%
3	x3	Dokumen x3	Dokumen x4	20%	25%
4	x4	Dokumen x4	Dokumen x5	50%	55%
5	x5	Dokumen x5	Dokumen x6	60%	63%

6	x6	Dokumen x6	Dokumen x7	32%	32%
7	x7	Dokumen x7	Dokumen x8	67%	69%
8	x8	Dokumen x8	Dokumen x9	48%	50%
9	x9	Dokumen x9	Dokumen x10	50%	51%
10	x10	Dokumen x10	Dokumen x11	23%	28%
11	x11	Dokumen x11	Dokumen x12	45%	55%
12	x12	Dokumen x12	Dokumen x13	56%	50%
13	x13	Dokumen x13	Dokumen x14	55%	43%
14	x14	Dokumen x14	Dokumen x15	12%	15%
15	x15	Dokumen x15	Dokumen x16	23%	28%
16	x16	Dokumen x16	Dokumen x17	32%	33%
17	x17	Dokumen x17	Dokumen x18	45%	42%
18	x18	Dokumen x18	Dokumen x19	65%	68%
19	x19	Dokumen x19	Dokumen x1	44%	45%

Dari hasil pengujian menggunakan aplikasi yang dirancang pengujian menggunakan stemming dan tanpa stemming terdapat perubahan yang cukup signifikan, hal ini terjadi karena document yang diuji menggunakan stemming, berarti document tersebut telah melalui suatu tahapan yang disebut prepocesing pada pada tahapan ini teks akan diolah dengan membuang beberapa kalimat atau kata yang tidak memiliki maksna seperti awalan, imbuhan dan kata sambung, sehingga didapat sebuah kata dasar. Jadi pattern yang diabandingkan oleh algoritma rabin karp menjadi lebih teratur dan hasilnya pun cukup baik dibanding dengan rabinkarp tanpa algoritma nazief andriani.

Hasil pengujian diatas merupakan hasil pembuktian dari software yang dirancanga dan masih belum bisa dikatakan valid 100% karena software yang dirancang belum melalui tahap pengujian kualitas software, jadi perlu ada penelitian lanjutan tentang penerapan algoritma rabinkarp dan steming nazief andriani ini.

4. Kesimpulan

Dari hasil pengujian dan analisa pada penelitian ini, beberapa ditemukan beberapa kesimpulan sebagai salah satu acuan pengembangan dari penelitian ini diantaranya :

1. Algoritma rabin karp bisa digunakan untuk proses pengecekan pattern pada sebuah teks yang cukup besar akan tetapi algoritma ini sangat lambat dalam proses eksekusinya sehingga perlu adanya penelitian lanjutan yang mengoptimalkan waktu eksekusi dari algoritma rabinkarp.
2. Algoritma stemming najief andriani sangat cocok di terapkan dalam proses pencarian root kata atau kata dasr dalam bahasa indonesia berdasarkan kamus kata dasar
3. Ketika nilai k-gram pada algoritma rabin karp yang diberikan kecil yang terjadi adalah hasil persentase yang di hasilkan akan jauh dari harapan hal ini terjadi karena algoritma rabin karp butuh optimalisasi.
4. Aplikasi perlu pengujian kualitas agar mendapat nilai validasi yang baik, seperti pengujian white box, yang bertujuan untuk mengukur tingat efektifitas algoritma
5. Hasil pengujian diatas diambil hanya untuk mengetahui perbedaan antara penggunaan stemming pada proses pemerikasaan dokumen dan tanpa menggunakan stemming dan pengujian diatas bukan merupan hasil uji yang valid dari segi algoritma yang digunakan.
6. Penelitian selanjutnya diharapakan dapat mengembangkan teknik pengecekan document menggukan metode lain seperti NLP dan metode teks minning yang lainya.

Referensi

A. S. Salmuasih, “Perancangan sistem deteksi plagiat pada dokumen teks dengan konsep similarity menggunakan algoritma rabin karp naskah publikasi,” pp. 1–20, 2013.

Departemen Pendidikan dan kebudayaan, “PUSAT BAHASA DEPDIKBUD REPUBLIK INDONESIA,” 2008. [Online]. Available: http://www.badanbahasa.kemendikbud.go.id/kbbi/index.php. [Accessed: 31-Dec-2014].

Haytham I.M Alzeini, Shihab A Hamed, and Mohamed H Habaebi, "Optimizing OLAP Heterogeneous Computing Based on Rabin-Karp Algorithm," in *Proc. of the IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)*, Kuala Lumpur, November 2013, pp. 26-27.

L. Agusta, U. Kristen, and S. Wacana, “PERBANDINGAN ALGORITMA STEMMING PORTER DENGAN ALGORITMA NAZIEF & ADRIANI UNTUK STEMMING DOKUMEN TEKS BAHASA INDONESIA,” pp. 196–201, 2009.

Meusche Norman and Gip Bela, "Citation Pattern Matching Algorithms for Citation-based Plagiarism Detection: Greedy Citation Tiling, Citation Chunking and Longest Common Citation Sequence. Proceeding of the 11th ACM Symposium on Document Engineering," in *Proceeding of the 11th ACM Symposium on Document Engineering*, California, 2011.

N. Made, A. Lestari, I. K. Gede, D. Putra, A. A. Ketut, and A. Cahyawan, “Personality Types Classification for Indonesian Text in Partners Searching Website Using Naïve Bayes Methods,” vol. 10, no. 1, pp. 1–8, 2013.