

TUGAS PEMROSESAN BAHASA ALAMI
STEMMING MULTIPROCESSING



Oleh:

Kelompok 2

Muhd. Muqsith Giga Saputra	09021181924023
Nurul Akhni	09021181924024
Citra Septivia Ardila	09021181924025
Alfath Aditya Putra	09021281924027
Jerrel Adriel A.Hutahaeen	09021281924031

PBA REGULER

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2022

Analisis Masalah

Stemming merupakan sebuah proses yang bertujuan untuk mereduksi jumlah variasi dalam representasi dari sebuah kata (Kowalski, 2011). Tujuan dari *stemming* sebenarnya adalah untuk meningkatkan performansi dan mengurangi penggunaan *resource* dari sistem dengan mengurangi jumlah *unique word* yang harus diakomodasikan oleh sistem. Jadi secara umum, algoritma *stemming* mengerjakan transformasi dari sebuah kata menjadi sebuah standar representasi morfologi (yang dikenal sebagai stem). Dalam melakukan proses *stemming* sebuah file terutama untuk file teks yang berukuran besar membutuhkan waktu yang relatif lama. Jika terdapat beberapa file teks yang perlu diproses, maka semakin banyak waktu yang dibutuhkan karena setiap file akan di-*stemming* secara berurutan.

Solusi

Berdasarkan analisis masalah diatas, diperlukannya aplikasi tambahan dalam melakukan *stemming* yaitu **Multiprocessing**. *Multiprocessing* mengacu pada kemampuan suatu sistem yang dapat menjalankan lebih dari satu proses secara bersamaan (paralel). Untuk mengetahui penerapan *multiprocessing* pada *stemming*, dilakukan pengujian (*testing*) dengan menggunakan 3 file berukuran 10 MB, 15 MB, dan 20 MB.

Analisis Hasil

Dari output program didapat hasil seperti berikut.

No	Proses	Waktu Eksekusi			
		10 MB	15 MB	20 MB	Keseluruhan
1.	Stemming tanpa Multiprocessing	260.9 seconds(s)	252.1 seconds(s)	249.0 seconds(s)	762.3 seconds(s)
2.	Stemming dengan Multiprocessing	452.1 seconds(s)	453.2 seconds(s)	454.4 seconds(s)	455.0 seconds(s)

Pada tabel diatas dapat dilihat bahwa adanya perbedaan waktu eksekusi yang signifikan pada proses *stemming*. Pada waktu eksekusi keseluruhan file, proses *stemming* tanpa multiprocessing membutuhkan waktu yang lebih lama dikarenakan setiap file tidak diproses secara bersamaan dan harus menunggu sampai seluruh file sebelumnya selesai dieksekusi. Sedangkan pada proses *stemming* dengan multiprocessing, waktu eksekusi keseluruhan file yang dibutuhkan hampir sama dengan waktu eksekusi setiap file. Hal tersebut dikarenakan multiprocessing membuat seluruh file dieksekusi secara bersamaan (paralel). Sehingga dapat disimpulkan bahwa proses Stemming dengan Multiprocessing cukup signifikan dapat menurunkan waktu proses *stemming* dan dapat mereduksi sampai 30% lebih dari waktu proses Stemming tanpa Multiprocessing. Hal tersebut juga bergantung pada ukuran file serta waktu eksekusi setiap file yang bisa berubah karena waktu tersebut tidak konstan.

Screenshot Program/Algoritma

1. Library

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from multiprocessing import Process
import time
```

2. Function

```
def stemming(file_i):
    factory = StemmerFactory()
    stemmer = factory.create_stemmer()
    teks_stem = stemmer.stem(file)

    # path tujuan penyimpanan file
    stem_file = f'C:/Users/ASUS/IdeaProjects/Python/src/NLP/File Stemming/file_stem_{i}.txt'
    with open(stem_file, 'w') as s:
        s.write(teks_stem)

    print(f'Result:success! file_stem{i}.txt saved')
```

3. Main

```
if __name__ == "__main__":
    # daftar file yang akan distemming
    files = ['10mb.txt', '15mb.txt', '20mb.txt']

    # input opsi proses yang akan dilakukan
    option_process = int(input('Pilih Proses Stemming \n1. Stemming tanpa Multiprocessing \n2. Stemming dengan Multiprocessing \nMasukkan Pilihan (angka) : '))

    process_memory = []
    i=10 # indentasi untuk penomoran file hasil

    start = time.perf_counter() # awal perhitungan waktu pemrosesan
    for file in range(len(files)):
        if option_process == 1: # Stemming tanpa Multiprocessing
            with open('C:/Users/ASUS/IdeaProjects/Python/src/NLP/File Proses Stemming/'+files[file], 'r') as p:
                openFile = p.read()
                stemming(openFile, i)

        elif option_process == 2: # Stemming dengan Multiprocessing
            with open('C:/Users/ASUS/IdeaProjects/Python/src/NLP/File Proses Stemming/'+files[file], 'r') as p:
                openFile = p.read()
                m = Process(target=stemming, args=(openFile, i))
                process_memory.append(m)

        i += 5
```

```

for p in proc:
    p.start()

for p in proc:
    p.join()

# akhir perhitungan waktu pemrosesan
finish = time.perf_counter()
print(f'Finished in {round(finish-start,4)} seconds(s)')

```

Penjelasan Algoritma

- Import Library

Hal pertama yang diperlukan pada program ini yaitu mengimport semua library yang diperlukan. Terdapat 3 library pada program yang telah dibuat yaitu Sastrawi, Multiprocessing dan time.

- Proses Stemming (stemming())

Fungsi stemming() berisi kode dan perintah yang melakukan perubahan untuk kata pada kalimat menjadi kata dasar dengan menggunakan class dari library Sastrawi.

- Multiprocessing

Pada aplikasi multiprocessing ini menggunakan kelas **Process** untuk mengalokasikan seluruh proses yang ingin dijalankan untuk disimpan pada memori dalam bentuk array (proc[]). Pada kelas **Process** terdapat parameter berupa **target** untuk mendefinisikan fungsi yang akan dipanggil dan **args** untuk menentukan argumen yang akan diteruskan (parameter dari fungsi). Setelah tersimpan ke memori, seluruh proses akan dimulai hampir pada waktu yang sama dengan memanggil **.start()** lalu **.join()** akan memastikan bahwa kode program lain setelahnya hanya berjalan setelah multiprocessing selesai.

- Main

Pada Main akan diminta untuk menginputkan angka dari pilihan untuk proses *stemming*.

- Jika menginputkan 1 (Stemming tanpa Multiprocessing) maka akan membaca file dan membawa data file tersebut langsung ke fungsi stemming(). Pada kondisi ini

proses *stemming* pada file akan dilakukan secara berurutan (*sequential*) yang dimana proses *stemming* pada setiap file akan dilakukan setelah file sebelumnya selesai di-*stemming*.

- Jika menginputkan 2 (Stemming dengan Multiprocessing). Pada kondisi ini akan menggunakan kelas dari library Multiprocessing yaitu **Process**. Fungsi *stemming* akan didefinisikan sebagai target dan file yang ingin diproses akan dibaca lalu diteruskan sebagai parameter (arguments). Setelah itu setiap proses akan disimpan ke dalam array `process_memory[]` untuk dilakukan loop dan memulai seluruh proses dalam waktu yang sama dengan memanggil `.start()`. Dengan mengaplikasikan multiprocessing, maka proses untuk setiap file akan dijalankan secara bersamaan (paralel) sehingga waktu yang dihasilkan merupakan waktu proses masing - masing file (bukan keseluruhan).

Output

- Stemming tanpa Multiprocessing

```
stemming x
C:\Users\ASUS\IdeaProjects\Python\Scripts\python.exe C:/Users/ASUS/IdeaProjects/Python/src/NLP/stemming.py
Pilih Proses Stemming
1. Stemming tanpa Multiprocessing
2. Stemming dengan Multiprocessing
Masukkan Pilihan (angka) : 1

Result:success! file_stem10mb.txt saved
file_stem10mb process finished in 260.9 seconds(s)

Result:success! file_stem15mb.txt saved
file_stem15mb process finished in 252.1 seconds(s)

Result:success! file_stem20mb.txt saved
file_stem20mb process finished in 249.0 seconds(s)

All Process Finished in 762.3 seconds(s)

Process finished with exit code 0
```

- Stemming dengan Multiprocessing

```
stemming x
C:\Users\ASUS\IdeaProjects\Python\Scripts\python.exe C:/Users/ASUS/IdeaProjects/Python/src/NLP/stemming.py
Pilih Proses Stemming
1. Stemming tanpa Multiprocessing
2. Stemming dengan Multiprocessing
Masukkan Pilihan (angka) : 2

Result:success! file_stem10mb.txt saved
file_stem10mb process finished in 452.1 seconds(s)

Result:success! file_stem15mb.txt saved
file_stem15mb process finished in 453.2 seconds(s)

Result:success! file_stem20mb.txt saved
file_stem20mb process finished in 454.4 seconds(s)

All Process Finished in 455.0 seconds(s)

Process finished with exit code 0
```

- Hasil Stemming

file_stem_10mb - Notepad

File Edit Format View Help

Jembatan yang bentang di atas sungai mus kota Palembang ini milik panjang 1 177 meter lebar 22 meter dan tinggi 63 meter dan jarak antara menara 1 lagi bangun jembatan saat sidang pleno yang langsung pada 29 oktober 1956 bangun jembatan ini bilang cukup nekat sebab pada saat itu anggaran presiden jembatan sebut beri nama jembatan bung karno karena dengan sungguh-sungguh juang warga Palembang untuk milik buah jembatan di atas sungai m orientasi dari air ke darat komunitas dagang buah dan sayur di pasar benteng rupa komunitas yang dulunya jual di pasar apung sungai mus pempek at empek yakin asal dari sebut apek atau pek yaitu sebut untuk paman atau lelaki tua tionghoa dasar cerita masyarakat sekitar tahun 1617 orang apek iphone milik antarmuka guna yang rancang pada layar multi sentuh perangkat ini hubung ke jaring seluler atau wi-fi dan dapat laku panggil jajah w bagai merek yang populer ponsel cerdas bentuk pipih dan cipta pasar yang besar untuk aplikasi ponsel cerdas atau sebut ekonomi aplikasi pada janu at tegar rokr e1 batas simpan hanya 100 lagu itunes untuk hindar saing dengan ipod nano milik apple cingular bebas apple untuk kembang perangkat ilis iphone 3g di lebih dari delapan puluh negara dan wilayah 33 apple umum iphone 3gs pada 8 juni 2009 serta rencana untuk rilis pada bulan juni nuari 2011 verizon umum dalam jumpa pers bahwa mereka telah sepakat dengan apple dan akan mulai jual iphone 4 cdma verizon ungkap prapemesanan ip baik dengan 53 dapat asal dari jual 37 juta iphone dengan harga jual rata hampir 660 per unit harga jual rata cukup konstan jika banding dengan m ebih ringan dengan berat 112 gram 4 oz pada tanggal 6 juli 2013 lapor bahwa apple negosiasi dengan operator seluler korea sk telecom untuk rilis cupertino dua perangkat milik layar yang lebih besar dari dahulu masing-masing ukur 4 7 inci 120 mm dan 5 5 inci 140 mm telah iphone 6 dan 6 plus lensa telefoto baru yang fungsi lebih baik dalam kondisi kurang cahaya dan tingkat fungsi kamera untuk ar pada 12 september 2018 apple cara resmi a model bekal lensa ultra lebar yang muncul bidang pandang lebih luas model pro milik fitur kamera tiga lensa dan lapis bahan kaca matte seri iph radisional kait dengan pengaruh kuliner tionghoa akan tetapi bagi sejarawan sebut bahwa pempek mungkin asal dari makan kuno yang sebut les makan an tuna putih saji pempek tani oleh kuah saus warna hitam cokelat yang sebut cuka atau cuko bahasa Palembang cuko buat dari air yang dididh kemudi aksan tekwan model celimpungan dan lenggang laksanakan dan celimpungan saji dalam kuah yang kandung santan sedang model dan tekwan saji dalam kuah is ni kemudian jadi tumpu tumbuh usaha ini pasca mati lee pada tahun 1987 samsung pisah jadi empat grup bisnis yakni samsung group shinsegae group c 2 samsung milik pengaruh kuat pada kembang ekonomi politik media dan budaya di korea selatan serta jadi dukung utama di balik ajaib di sungai han di bagai macam industri samsung pun ekspansi ke sektor asuransi sekuritas dan ritel pada tahun 1947 cho hong-jai diri hyosung group sama diri sam ia yang rancang kembang dan jual telepon cerdas aplikasi seluler dan elektronik konsumen diri serta ceo xiaomi adalah lei jun yang rupa orang kay ui bas android pada bulan agustus 2012 xiaomi umum telepon cerdas yang nama mi 2 ponsel ini dukung oleh snapdragon s4 pro apq8064 dari qualcomm 1 da tengah tahun 2014 banyak 26 1 juta telepon cerdas pada tahun 2014 xiaomi umum akan luas pasar ke luar tiongkok xiaomi mulai debut internasionala a bulan april 2014 xiaomi beli domain baru dengan nama mi com dengan harga 3 6 juta dolar as atau sekitar 44 5 miliar rupiah mi com jadi nama dom as jadi salah satu usaha teknologi swasta yang paling harga di dunia ferrari adalah buah produsen mobil super dan mobil balap asal italia perform dapat global 118 3 miliar dolar as tahun 2009 usaha pertama henry ford untuk milik usaha mobil dengan nama sendiri mulai pada henry ford company sar bagai salah satu usaha yang kontrol keluarga besar di dunia ford motor company tetap kontrol oleh keluarga ford lama lebih dari 100 tahun tel una slogan iklan have you driven a ford lately apakah anda telah kendara ford untuk buat merk mereka tampak lebih modern pada tahun 1990 dan 1994 ang bas platform unibody tidak lagi sasis body-on-frame untuk kembang teknologi gerak listrik hibrida untuk suv ford escape hybrid ford ambil lis a 2 3 miliar dolar as tanggal 29 oktober 2012 ford umum bahwa mereka jual bisnis komponen atur cuaca ke detroit thermal systems llc bisnis ini ru saham di mazda dari 33 4 jadi 13 4 tanggal 18 november 2010 ford kurang lagi milik saham hingga jadi 3 saja hal ini tuju untuk tingkat fleksibili