

LEMBAR LATIHAN: PENGAMBILAN KATA DASAR (STEMMING DAN LEMATISASI)

Mata Kuliah: Pemrosesan Bahasa Alami (Natural Language Processing)

Dosen Pengampu: _____

Nama Mahasiswa: _____ NIM: _____

Tujuan Pembelajaran

Mahasiswa mampu:

1. Membedakan antara stemming dan lematisasi.
2. Melakukan proses pengambilan kata dasar menggunakan kedua metode tersebut.
3. Menganalisis hasil dan memilih metode yang sesuai untuk kasus tertentu.

Bagian A — Teori Singkat (Isian)

1. Jelaskan secara singkat apa yang dimaksud dengan stemming.
2. Jelaskan perbedaan utama antara stemming dan lematisasi.
3. Sebutkan satu kelebihan dan satu kekurangan lematisasi.
4. Mengapa hasil stemming terkadang bukan kata yang valid secara linguistik?
5. Kapan lebih baik menggunakan stemming dibanding lematisasi?

Bagian B — Praktik Pemrograman

Gunakan Python dan pustaka berikut:

- ✓ Sastrawi untuk Bahasa Indonesia
- ✓ nltk untuk Bahasa Inggris

1. Stemming Bahasa Indonesia

Tuliskan kode dan hasil dari teks berikut:

kalimat = "Pemerintah sedang mempersiapkan peraturan baru tentang pendidikan nasional."

Tugas:

- a. Lakukan stemming menggunakan Sastrawi.
- b. Tuliskan hasil kata dasar setiap kata.

2. Lematisasi Bahasa Inggris

Gunakan teks berikut:

sentence = "The children were running and studying hard for their exams."

Tugas:

- a. Lakukan lemmatization dengan WordNetLemmatizer (nltk).
- b. Tampilkan pasangan kata asli dan hasil lema-nya dalam format tabel.

Bagian C — Analisis Hasil

1. Apakah hasil stemming menghasilkan kata dasar yang bermakna dengan benar?
2. Apakah lemmatization lebih akurat dari stemming pada contoh di atas?
3. Menurut Anda, bagaimana pengaruh penggunaan stemming atau lematisasi terhadap hasil analisis teks (misal dalam klasifikasi berita atau pencarian dokumen)?

Bagian D — Tugas Tambahan (Opsiional)

Carilah 3 kalimat berbahasa Indonesia dari berita online, lalu lakukan:

1. Stemming dengan Sastrawi
2. Lematisasi (jika tersedia model)

Tuliskan perbandingan hasilnya dan simpulkan metode mana yang lebih sesuai untuk Bahasa Indonesia.

Petunjuk Pengumpulan

1. Jawaban diketik rapi dalam format Word atau PDF.
2. Sertakan kode program dan hasil output.
3. File diberi nama: LembarLatihan_NLP_NamaMahasiswa.docx
4. Kumpulkan melalui moca.unimma.ac.id pertemuan 6 paling lambat 3 hari setelah perkuliahan.

Kriteria Penilaian

Aspek yang Dinilai	Bobot
Pemahaman konsep (Bagian A)	25%
Kebenaran kode dan hasil (Bagian B)	40%
Analisis hasil dan kesimpulan (Bagian C)	25%
Kerapian dan sistematika laporan	10%