



SOAL UJIAN TENGAH SEMESTER GASAL
UNIVERSITAS AMIKOM PURWOKERTO
TAHUN AKADEMIK 2021/2022

Mata Kuliah	: Pemrosesan Bahasa Alami
SKS/Semester	: 3 / Gasal
Fakultas	: Ilmu Komputer
Program Studi	: Informatika
Dosen	: Bambang Pilu Hartato, S.Kom., M.Eng.
Kelas	: IF 19 A, B, C, Sore
Hari/Tanggal	: Rabu / 17 November 2021
Jam	: 14.30 – 16.00 (IF 19 A, B, C) 16.30 – 18.00 (IF 19 Sore)
Sifat Ujian	: Penugasan

Petunjuk mengerjakan soal:

- Kerjakan soal-soal berikut dengan teliti dan seksama.
- Baca setiap petunjuk yang ada pada setiap soal.
- Ketik jawaban Anda ke dalam bentuk PDF (ukuran maksimal 2 MB). Jangan lupa cantumkan nama dan NIM Anda pada lembar jawab.
- Rename file lembar jawaban Anda dengan format **PBA_UTS2022_KELAS_NIM**.
- Unggah jawaban Anda pada sistem student center untuk mata kuliah Pemrosesan Bahasa Alami.

Soal:

- Pemrosesan Bahasa Alami merupakan salah satu cabang kecerdasan buatan yang berfokus bagaimana membuat mesin dan manusia dapat berkomunikasi dengan menggunakan bahasa manusia sealamiah mungkin. Namun dalam penerapannya terdapat cukup banyak masalah yang harus dihadapi. Sebutkan dan jelaskan setidaknya dua masalah utama yang membuat pemrosesan bahasa alami cukup sulit untuk dilakukan! **(Nilai: 10)**

- Jika diketahui bentuk Regular Expression adalah sebagai berikut:

$$[\backslash d]^+/[A-Z]^+/[A-S]\{3\}/[0-9]\{2\}/[IVXLMDCLXVI]^+/\backslash d\{4\}$$

Berikan contoh salah satu string yang memenuhi Regular Express tersebut. Berikan penjelasan mengapa string tersebut memenuhi Regular Expression yang telah ditentukan!

(Nilai: 20)

- Anda diberikan dua buah string yaitu:

String 1 : An act

String 2 : a cat

Hitunglah seberapa beda kedua string tersebut dengan teknik Levenshtein Distance / Minimum Edit Distance. Anda dapat menghitungnya dengan menggunakan matriks ataupun menghitungnya secara langsung. **(Nilai: 20)**

- Jika diketahui suatu string sebagai berikut:





Awakdewe tau duwe bayangan. Besok yen wes wayah omah – omahan, Aku moco koran sarungan kowe blonjo dasteran.

Buatlah bentuk 5-gram nya! (Nilai: 25)

5. Jika diketahui suatu corpus berisi teks sebagai berikut:

Ada seorang suami yang di dalam dompetnya terdapat foto istrinya. Saat teman-temannya melihat, ia dipuji sebagai suami yang sangat baik. Lalu, satu di antara temannya bertanya apa fungsinya membawa foto sang istri. Dia menjawab: "kalau aku punya permasalahan di kantor, aku selalu memandang foto itu, dan permasalahan yang dihadapi hilang begitu saja". "Wah alangkah berbahagianya kamu mempunyai istri seperti itu, bagaimana bisa begitu?" tanya teman-temannya. Sang suami menjawab kembali: "Ya, kalau saya melihat foto istri saya, semua permasalahan apa pun di kantor, menjadi tidak ada apa-apanya dibandingkan dengan permasalahan dengan dia!"

Diasumsikan Anda telah melakukan beberapa langkah pra-pengolahan teks terhadap corpus tersebut, seperti casefolding dan punctuation removal. Anda diminta untuk menghitung probabilitas munculnya string berikut.

“suami yang sangat baik”

Perlu diketahui, pada skenario ini Anda memutuskan untuk menghitungnya dengan model probabilistik bahasa dengan pendekatan bi-gram. Jelaskan langkah-langkah perhitungan yang Anda lakukan. (Nilai: 25)

-Tidak ada kesuksesan yang instan dalam hidup ini. Bahkan untuk membuat mie instan sekalipun kita harus merebus air terlebih dahulu-