



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)**

Jawaban:

Dua masalah utama pemrosesan bahasa alami adalah:

- Sifat mesin yang hanya dapat mengolah data yang sifatnya terstruktur sedangkan bahasa alami manusia masih jauh dari sifat terstruktur. Tentu kedua hal ini merupakan hal yang sangat bertentangan sehingga perlu dipikirkan cara bagaimana mesin dapat mengakomodir sifat bahasa manusia tersebut.
- Bahasa alami manusia masih memiliki sifat ambiguitas. Dengan adanya ambiguitas makna dari bahasa yang digunakan manusia maka mesin terkadang salah menafsirkan bahasa manusia. Mesin harus benar-benar mengetahui konteks dari pembicaraan agar dapat meminimalisir kekurangan tersebut.

Penilaian:

- Jika jawaban menyebutkan 2 beserta alasan lengkap dengan pembahasannya maka mendapat poin: 10 poin





- Jika jawaban menyebutkan 2 namun tanpa menyebutkan alasan lengkap maka mendapat 8 poin
- Jika jawaban menyebutkan 1 beserta alasan lengkap dengan pembahasannya maka mendapat poin: 5 poin
- Jika jawaban menyebutkan 1 namun tanpa menyebutkan alasan lengkap maka mendapat 3 poin
- Jika asal menjawab mendapat 1 poin
- Jika tidak menjawab sama sekali 0 poin

Soal:

2. Jika diketahui bentuk Regular Expression adalah sebagai berikut:

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

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

(Nilai: 20)

Jawaban:

Salah satu contoh string yang memenuhi regex

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

adalah

$$2/XY/BPS/46/VI/5758$$

String tersebut memenuhi kriteria regex karena:

- Segmen pertama setidaknya terdiri dari sebuah digit angka lalu diikuti dengan garis miring (*slash*).
- Segmen ke-2 setidaknya terdiri dari sebuah karakter alfabet kapital mulai dari huruf A-Z diikuti dengan garis miring (*slash*).
- Segmen ke-3 terdiri tepat 3 karakter alfabet kapital mulai dari huruf A-S diikuti dengan garis miring (*slash*).
- Segmen ke-4 terdiri tepat 2 digit angka mulai dari angka 0-9 diikuti dengan garis miring (*slash*).





- e. Segmen ke-5 terdiri dari setidaknya 1 karakter romawi yang merupakan kombinasi dari I, V, X, L, M, D, dan C diikuti dengan garis miring (*slash*).
- f. Segmen ke-6 terdiri tepat 4 digit angka.

Penilaian:

- Jika jawaban string yang diberikan seluruhnya benar disertai alasan yang benar maka mendapat 20 poin
- Jika jawaban string yang diberikan seluruhnya benar namun tidak disertai alasan yang benar maka mendapat 16 poin
- Jika jawaban string yang diberikan setidaknya 3 segmen benar disertai alasan yang benar maka mendapat 10 poin
- Jika jawaban string yang diberikan setidaknya 3 segmen benar namun tidak disertai alasan yang benar maka mendapat 6 poin
- Jika asal menjawab mendapat 1 poin
- Jika tidak menjawab sama sekali maka mendapat 0 poin

Soal:

3. 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)**

Jawaban:

Jarak levenshtein antara An act dan a cat adalah

A	n		a	c	t
a			c	a	t
1	1	0	1	1	0

Jarak levenshtein: $1 + 1 + 0 + 1 + 1 + 0 = 4$

Penilaian:





- Jika jawaban nilai jarak levenshtein benar dan disertai cara menghitungnya (menghitung langsung / menggunakan matrik) yang benar juga maka mendapat 20 poin
- Jika jawaban nilai jarak levenshtein benar namun tidak disertai cara menghitungnya (menghitung langsung / menggunakan matrik) yang benar juga maka mendapat 16 poin
- Jika jawaban nilai jarak levenshtein salah namun sudah berusaha menghitungnya dengan benar maka mendapat 10 poin
- Jika jawaban nilai jarak levenshtein salah dan juga cara menghitung salah maka mendapat 6 poin
- Jika asal menjawab maka mendapat 1 poin
- Jika tidak menjawab sama sekali maka mendapat 0 poin

Soal:

4. 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**)

Jawaban:

Token:

[Awakdewe] [tau] [duwe] [bayangan] [.] [Besok] [yen] [wes] [wayah] [omah] [-] [omahan] [.] [Aku] [moco] [koran] [sarungan] [kowe] [blonjo] [dasteran] [.] (21 token)

#1: Awakdewe tau duwe bayangan .

#2: tau duwe bayangan . Besok

#3: duwe bayangan . Besok yen

#4: bayangan . Besok yen wes

#5: . Besok yen wes wayah

#6: Besok yen wes wayah omah

#7: yen wes wayah omah –

#8: wes wayah omah – omahan

#9: wayah omah – omahan ,

#10: omah – omahan , Aku





- #11: – omahan , Aku moco
- #12: omahan , Aku moco koran
- #13: , Aku moco koran sarungan
- #14: Aku moco koran sarungan kowe
- #15: moco koran sarungan kowe blonjo
- #16: koran sarungan kowe blonjo dasteran
- #17: sarungan kowe blonjo dasteran .0

Penilaian:

- Jika jawaban benar semua termasuk case huruf / besar kecilnya huruf dan tanda baca juga diikutkan maka mendapat poin 25.
- Jika jawaban benar (tanda baca dianggap token) tapi ada kesalahan case huruf maka mendapat 20 poin.
- Jika berhasil melakukan 5-gram (formasinya benar) namun mengabaikan case huruf / tanda baca maka mendapat 12 poin.
- Jika salah formasi n-gram maka mendapat 5 poin
- Jika asal menjawab maka mendapat 1 poin
- Jika tidak menjawab sama sekali maka mendapat 0 poin.

Soal:

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”

SUCCESS, SPIRIT & CREATIVE





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)

Jawab:

Corpus yang sudah melalui tahap pra-pengolahan teks:

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

$$p(\text{suami yang sangat baik}) = p(\text{yang|suami}) * p(\text{sangat|yang}) * p(\text{baik|sangat})$$

$$p(\text{yang|suami}) = c(\text{suami yang}) / c(\text{suami}) = 2 / 3 = 0.667$$

$$p(\text{sangat|yang}) = c(\text{yang sangat}) / c(\text{yang}) = 1 / 3 = 0.333$$

$$p(\text{baik|sangat}) = c(\text{sangat baik}) / c(\text{sangat}) = 1 / 1 = 1.000$$

$$p(\text{suami yang sangat baik}) = 0.667 * 0.333 * 1 = 0.222$$

Penilaian:

- Jika jawaban benar disertai dengan cara menghitung secara sistematis maka mendapat 25 poin
- Jika jawaban benar namun tanpa cara yang sistematis maka mendapat poin 12
- Jika cara sistematis namun jawaban masih keliru maka mendapat poin 10
- Jika asal menjawab maka mendapat 1 poin
- Jika tidak menjawab sama sekali maka mendapat 0 poin.

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

