

## Gelombang 2 [50 poin]

Jawab dengan se jelas-jelasnya.

Tidak akan ada tanya-jawab terkait dengan soal. Silakan ambil dan tuliskan asumsi masing-masing (jika diperlukan) dalam menjawab soal.

**A. Benar/Salah** - jelaskan apakah pernyataan di bawah ini benar/salah, dan berikan alasan anda. [0.5 poin untuk jawaban benar/salah yang tepat, 3 poin untuk alasan yang tepat, total 3.5 poin masing-masingnya, total keseluruhan 17.5 poin]

1. Abstractive summarization lebih sulit daripada extractive summarization.  
Ya krn harus melakukan paraphrasr. Extract hanya lgsg ambil dr kalimat di dukmu
2. Abstractive summarization pasti menghasilkan ringkasan yang lebih bagus dibanding extractive summarization.  
Salah. Belum tentu
3. Ada dua sistem summarization A dan B yang dites dengan testbench yang sama. A memiliki nilai ROUGE-2 precision yang lebih baik ketimbang B. Maka ROUGE-1 precision A juga pasti lebih baik dibanding B.
4. Hasil ringkasan LexRank secara umum lebih *fluent* dibanding dengan TextRank.
5. Algoritma LexRank dan TextRank keduanya memodelkan kalimat-kalimat dalam dokumen ke dalam suatu *complete graph*.

### B. Esai

1. [10 poin] Anda dihadapkan dengan suatu korpus multilingual dan anda diminta untuk melakukan clustering terhadap korpus tersebut. Sebutkan dua langkah berbeda yang bisa anda ambil sehingga cluster-cluster yang dihasilkan dapat mengandung dokumen dengan bahasa yang berbeda-beda namun tetap akurat!
2. [22.5 poin] Salah satu tantangan utama dalam pembentukan *chatbot* yang ekspresif adalah untuk merespon pernyataan yang sifatnya *open-ended*. Misalnya, pesan dengan *intent* menyatakan kesukaan terhadap sesuatu, seperti:  
**User:** "Aku suka Harry Potter"  
Yang bebas dijawab seperti apa pun, bahkan jawaban *ngeles* seperti  
**Chatbot:** "hmm"  
Tentunya jawaban ini tidak menarik. Misalkan anda diminta untuk membuat suatu template jawaban untuk chatbot ini sehingga setiap kali ada pesan:  
**User:** "Aku suka [Entity\_X]"  
Maka chatbotnya harus menjawab:

**Chatbot:** “Oh ya? Aku suka sama [Entity\_Y yang berhubungan dengan X]. Dia itu yang [salah satu atribut/aktivitas yang dilakukan oleh Y] kan?”

Contoh: “Oh ya? Aku suka sama **Dumbledore**. Dia itu yang **mati di buku keenam kan?**”

- a. [13 poin] Asumsikan anda memiliki sistem IR, NER, dan POS tag yang akurat, dan korpus yang sangat lengkap untuk entity apa pun.
  - i. [3 poin] Jelaskan bagaimana anda bisa membuat chatbot ini memahami *intent* ini.
  - ii. [10 poin] Jelaskan bagaimana anda bisa membangun suatu sistem berbasis IR sehingga si chatbot bisa menjawab pertanyaan dengan *intent* tersebut sesuai dengan template yang disebutkan di atas.  
**Note:** bisa jadi ada banyak [Entity\_Y] dan [Aktivitas] yang benar (dalam contoh di atas, bisa jadi **Snape**, **Hermione**, dll yang berhubungan dengan Harry Potter). Sistem anda bebas mengambil yang manapun sesuai dengan kriteria yang anda tetapkan sendiri (jelaskan kriterianya).
- b. [3.5 poin] Katakanlah anda sudah memiliki suatu sistem yang menjalankan hal ini. Bagaimana cara anda mengevaluasi sebaik apakah sistem tersebut bekerja?
- c. [6 poin] Tentunya jika template yang dipakai selalu seperti itu dan/atau [Entity\_Y] dan [Aktivitas] yang diambil itu-itu saja, maka jawaban chatbotnya akan monoton dan membosankan. Sebutkan **tiga** cara berbeda untuk menanggulangi hal ini!