

SISTEM RINGKASAN TEKS BAGI ARTIKEL WEB (MYWEBSUMMARIZER)

OSMAN BIN ISMAIL

FADILLA 'ATYKA NOR RASHID

*Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia, 43600 UKM Bangi,
Selangor Darul Ehsan, Malaysia*

ABSTRAK

Masyarakat hari ini tidak dapat lari daripada menggunakan Internet untuk komunikasi, hiburan, dan mendapatkan maklumat dengan pantas. Dengan kemudahan ini, berjuta-juta maklumat dikongsi setiap hari tanpa henti. Namun, perkongsian maklumat yang berterusan menyebabkan masyarakat perlu menghadam pelbagai kandungan daripada pelbagai sumber, sering kali memerlukan rujukan silang dari satu halaman ke halaman yang lain bagi mendapatkan konteks yang lebih lengkap. Untuk mencari maklumat dari sesebuah laman web, pembaca kebiasaannya perlu mengambil masa yang lama untuk memproses teks artikel yang panjang. Kadangkala, mereka juga perlu membaca rujukan luar yang disertakan dalam artikel untuk mendapatkan konteks yang lebih relevan. Ini akan mengambil masa yang lama bagi seseorang itu menghadam maklumat tersebut. Oleh yang demikian, projek ini membangunkan sistem ringkasan teks bagi artikel web, di mana teks utama artikel serta kandungan daripada pautan rujukan akan diringkaskan bersama untuk menambah konteks yang lebih bermakna dalam sistem ringkasan teks. Bagi menggabungkan artikel utama dan kandungan pautan tersebut, persamaan kosinus akan dikira. Sekiranya nilai persamaan melebihi satu paras yang ditetapkan, kedua-dua kandungan tersebut akan digabungkan. Gabungan kandungan ini diproses melalui dua model Transformer pra-latih; model PEGASUS digunakan untuk mengekstrak ayat-ayat penting manakala T5 berperanan menstruktur semula ayat bagi menghasilkan ringkasan yang lebih abstraktif. Hasil ujian menunjukkan purata skor Flesch adalah sekitar 49.78, purata skor SMOG adalah 11.61, dan penilaian kualiti ringkasan menunjukkan purata penilaian bintang sebanyak 3 daripada 5. Sistem ini memudahkan pengguna mendapatkan maklumat yang relevan dengan lebih pantas, walaupun masih terdapat ruang untuk penambahbaikan dari segi kelajuan pemprosesan dan konsistensi ringkasan.

Kata kunci: Ringkasan Teks, Artikel Web, Transformer, PEGASUS, T5

PENGENALAN

Sebelum bermulanya era digital, di mana kita pada hari ini boleh mengakses pelbagai maklumat dengan sekeli mata, para penyelidik telah menyelesaikan pelbagai masalah berkaitan pemprosesan teks dan bahasa sejak awal 1950-an lagi. Usaha ini didorong oleh peningkatan jumlah teks seperti akhbar, jurnal, dan kertas kajian pada waktu itu. Selain itu,

perkembangan komputer turut memainkan peranan penting dalam memudahkan pengiraan model matematik yang rumit. Sejak penciptaan komputer, para penyelidik telah memanfaatkan kelajuan pengiraannya untuk menyelesaikan masalah ini. Kini, kebolehan dan kemajuan teknologi semasa dalam aspek pemprosesan teks dan bahasa adalah amat memberangsangkan dan sangat ke hadapan. Salah satu fokus utama para penyelidik dalam bidang ini adalah membangunkan model ringkasan teks yang berkesan. Ringkasan teks amatlah membantu pengguna dan pembaca dalam mengekstrak maklumat penting dan relevan, terutamanya dengan peningkatan berterusan jumlah teks yang dimuat naik ke Internet setiap hari. Dari semasa ke semasa, teknik bagi model ringkasan teks terus berkembang dan kian makin berkesan.

Salah satu cabang dalam ringkasan teks ialah ringkasan artikel web. Ringkasan teks untuk artikel web sudah pun wujud dan digunakan secara meluas oleh pengguna. Setiap sistem yang tersedia menggunakan teknik dan model yang berbeza untuk mengekstrak artikel dan meringkaskan teks tersebut. Namun, satu aspek yang sering diabaikan oleh kebanyakan sistem adalah konteks tambahan atau rujukan yang disertakan dalam artikel, seperti pautan yang memberi pembaca peluang untuk mendalami topik dengan lebih lanjut. Artikel web kini semakin banyak menambah pautan luar (external link) sebagai rujukan tambahan bagi pengguna yang ingin memahami konteks dengan lebih mendalam. Namun, penulis artikel sering kali menggunakan pautan ini secara berlebihan untuk meningkatkan jumlah klik pada artikel mereka, yang seterusnya meningkatkan trafik laman web dan memberikan manfaat dari segi pengiklanan atau tujuan lain. Walaupun begitu, rujukan ini boleh menjadi sangat penting apabila ia benar-benar relevan dan memberi nilai tambahan kepada artikel utama.

Dalam model transformer, penambahan konteks yang relevan ini boleh menghasilkan ringkasan yang lebih komprehensif, kerana ia dapat menghubungkan maklumat yang ada dengan lebih tepat. Jika pautan-pautan (external link) ini tidak diambil kira, model akan kehilangan maklumat tambahan yang boleh memberikan gambaran yang lebih lengkap. Bagi pengguna pula, cabaran utama adalah masa yang diperlukan untuk membaca keseluruhan artikel dan pautan-pautan yang disertakan. Pengguna sering perlu melompat dari satu halaman ke halaman lain sebelum memperoleh gambaran penuh mengenai isi artikel. Dengan adanya sistem ringkasan yang mengintegrasikan konteks daripada pautan, pengguna dapat memahami maklumat dengan lebih mendalam dan pantas tanpa perlu membaca setiap pautan secara berasingan.

Sebagai penyelesaian, sistem ringkasan teks yang dicadangkan akan menggabungkan artikel utama dengan artikel rujukan berdasarkan skor persamaan kosinus untuk memastikan ringkasan mempunyai konteks tambahan yang relevan. Pendekatan ini lebih efisien pada peringkat awal sebelum menggunakan model Transformer yang lebih komprehensif, yang akan dipilih melalui kajian literatur. Transformer pra-latih yang lebih sesuai digunakan dalam memastikan kualiti ringkasan yang dihasilkan adalah tepat dan bermakna. Bagi hasil akhir, sistem web akan dibangunkan bagi membolehkan pengguna memasukkan pautan artikel dan menetapkan had maksimum perkataan, dengan teknik ekstraksi seperti BeautifulSoup digunakan untuk mendapatkan kandungan. Pembersihan data juga dilakukan bagi

memastikan ringkasan berkualiti tinggi, manakala maklum balas pengguna serta penilaian kebolehbacaan seperti SMOG dan Flesch-Kincaid akan digunakan untuk penambahbaikan berterusan.

METODOLOGI KAJIAN

Metodologi yang akan digunakan bagi melaksanakan projek ini adalah berdasarkan model Spiral. Model Spiral dipilih kerana ia menggabungkan elemen pembangunan berulang (iterative development) dengan penilaian risiko yang berterusan, membolehkan perubahan dan penyesuaian sepanjang pembangunan sistem. Metodologi ini amat sesuai untuk projek ini kerana keperluan dan strategi yang terbaik bagi menghasilkan model ringkasan teks yang berkesan, dan mungkin berbeza seiring dengan perkembangan projek dan hasil penilaian. Melalui model Spiral ini juga, kos dan risiko yang berkaitan dapat diberi perhatian yang serius tanpa melibatkan risiko yang tinggi. Setiap fasa dalam model Spiral akan menghasilkan prototaip yang boleh diuji dan dianalisis untuk mengenal pasti kekurangan yang perlu ditambah baik. Dengan itu, pembangunan projek dapat diubah secara fleksibel dan berterusan.

Fasa pelan

Dalam fasa ini, kajian literatur yang mendalam akan dijalankan untuk menilai dan memilih model Transformer pra-latih yang paling sesuai dengan keperluan projek. Kajian ini akan memberi tumpuan kepada model seperti BERT, PEGASUS, T5, dan lain-lain berdasarkan prestasi mereka dalam tugas ringkasan teks. Reka bentuk awal sistem akan dibangunkan bagi memberikan gambaran keseluruhan mengenai aliran kerja, termasuk bagaimana artikel utama dan pautan rujukan diekstrak, dianalisis, dan diringkaskan. Ini akan merangkumi pemetaan aliran data, struktur pangkalan data, serta senibina sistem yang menghubungkan bahagian belakang (pemprosesan model Transformer) dengan bahagian depan (antara muka pengguna). Kod pseudo akan disediakan untuk menggambarkan logik algoritma sistem secara terperinci, termasuk langkah-langkah pemprosesan data, pengiraan persamaan kosinus, serta penggunaan model Transformer bagi menghasilkan ringkasan. Selain itu, antara muka pengguna dengan fideliti rendah akan direka bentuk bagi menggambarkan aliran interaksi pengguna dengan sistem. Ini termasuk skrin input pautan artikel, pilihan had panjang ringkasan, serta paparan hasil ringkasan yang dihasilkan oleh model.

Fasa analisis risiko

Fasa penilaian risiko akan dilakukan bagi mengenal pasti keperluan pemprosesan data yang besar serta kos sumber yang diperlukan untuk menjalankan model Transformer. Oleh kerana pemprosesan kandungan artikel di Internet kebiasaannya besar, had maksimum perkataan yang diterima akan diubah sesuai mengikut kemampuan sumber komputer yang ada. Masa yang diambil untuk menghasil ringkasan teks juga akan dianalisis, bagi menyesuaikan ia agar hasil akhir dipaparkan tidak mengambil masa terlalu lama.

Fasa pembangunan

Pada fasa ini, sistem akan dibangunkan secara menyeluruh merangkumi bahagian belakang (backend), bahagian depan (frontend), dan integrasi antara model-model sistem. Dari aspek

backend, sistem akan merangkumi proses ekstraksi artikel menggunakan BeautifulSoup untuk mengekstrak kandungan artikel utama serta pautan rujukan, memastikan hanya teks yang relevan dikekalkan. Proses pengiraan persamaan kosinus akan dijalankan menggunakan TF-IDF untuk mengira relevansi antara artikel utama dan pautan rujukan sebelum dihantar ke model Transformer bagi menghasilkan ringkasan. Dari aspek frontend, antara muka pengguna akan direka bentuk menggunakan HTML, CSS, dan JavaScript untuk membolehkan pengguna memasukkan pautan artikel, menetapkan had panjang ringkasan, dan melihat hasil ringkasan dengan kemas. Sistem maklum balas pengguna juga akan disediakan bagi membolehkan pengguna memberikan penilaian bintang dan maklum balas terhadap keberkesanan dan kepuasan mereka terhadap ringkasan teks yang dihasilkan. Keseluruhan integrasi sistem akan dilakukan menggunakan Flask sebagai API bagi menghantar data antara backend dan frontend dengan lancar.

Fasa penilaian

Dalam fasa penilaian, kebolehgunaan, dan keberkesanan sistem ringkasan teks akan diuji secara menyeluruh. Ujian akan dijalankan untuk mengenal pasti sebarang masalah atau kekurangan yang ada dalam sistem. Dari aspek kebolehgunaan, maklum balas pengguna akan di kumpul bagi menilai sejauh mana sistem memudahkan pengguna dalam mendapatkan ringkasan yang bermakna. Bagi menilai keberkesanan dan kualiti ringkasan teks, tahap kebolehbacaan ringkasan akan diuji menggunakan kaedah SMOG dan Flesch-Kincaid bagi memastikan ringkasan yang dihasilkan dapat difahami oleh pengguna. Penilaian bintang juga akan diambil bagi mendapatkan tahap kepuasan pengguna terhadap hasil akhir ringkasan teks. Berdasarkan hasil ujian dan maklum balas, pengubahsuaian akan dilakukan pada parameter model Transformer serta fungsi sistem lain bagi meningkatkan kualiti ringkasan. Proses penyesuaian dan penambahbaikan akan dilakukan secara iteratif untuk memastikan sistem memenuhi keperluan pengguna dengan lebih baik.

KEPUTUSAN DAN PERBINCANGAN

Satu sistem ringkasan teks berasaskan web, dinamakan MyWeb Summarizer, telah berjaya dibangunkan. Pengguna boleh memasukkan pautan ke dalam kotak "Website Link" serta menetapkan had maksimum perkataan bagi ringkasan teks yang dihasilkan melalui kotak "Maximum Words". Had maksimum perkataan mestilah melebihi 30 patah perkataan bagi memastikan ringkasan yang dihasilkan tidak terlalu pendek atau terpotong.

MyWebSummarizer

Quickly summarizes both the main article and any related links within it, giving you a clear overview of the entire content.

How it Works?

Website Link:

Maximum Words:

Summarize

The summary will appear here.

Rajah 1 Antara Muka Akhir Pengguna (Halaman Utama)

Setelah kandungan berjaya diekstrak dan diringkaskan di bahagian backend, hasil akhir ringkasan teks akan dipaparkan. Pengguna kemudian boleh memilih untuk menilai ringkasan teks atau membuat ringkasan baharu bagi pautan lain.

MyWebSummarizer

Quickly summarizes both the main article and any related links within it, giving you a clear overview of the entire content.

How it Works?

Website Link:

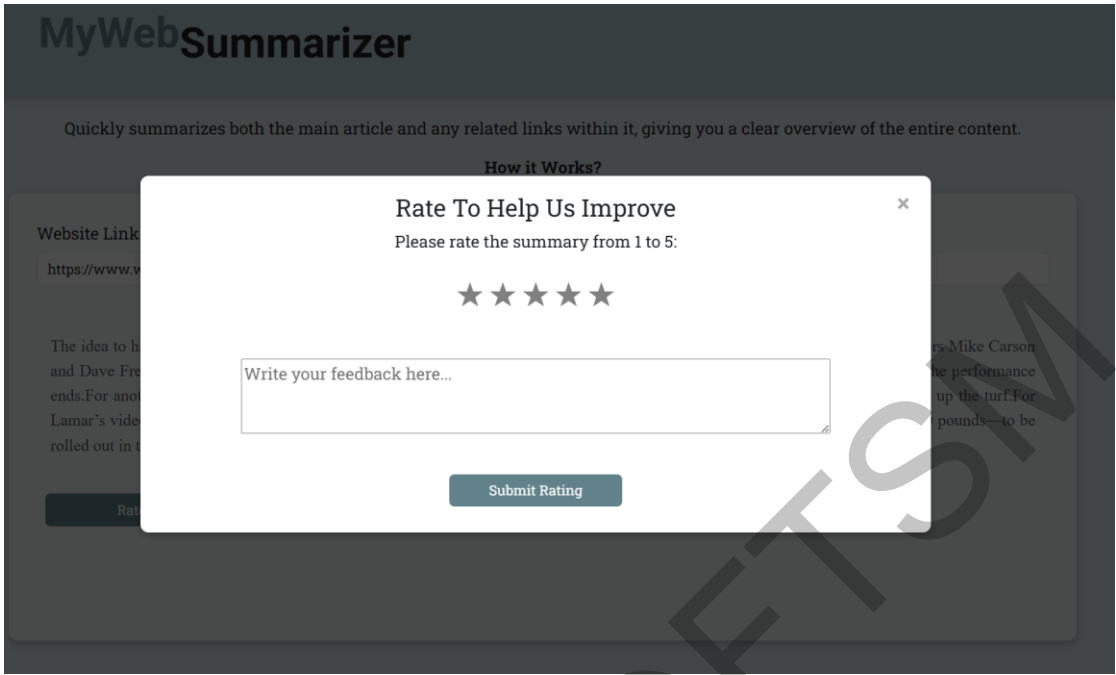
Maximum Words:

The idea to have the halftime show performance space evoke video games came from Kendrick Lamar and his frequent collaborators Mike Carson and Dave Free. All of the staging, lights, and performers have to get on the field in 7.5 minutes and then off of it 6 minutes after the performance ends. For another, you have to do all of this on a field where the biggest NFL game of the year is being played without screwing up the turf. For Lamar's video game themed show that meant more than 50 carts of staging and equipment each weighing between 1,500 and 3,500 pounds—to be rolled out in the center of the stadium.

Rate **Summarize Again**

Rajah 2 Antara Muka Akhir Pengguna (Contoh Input)

Pengguna juga boleh memberikan penilaian bintang bagi ringkasan teks yang telah dibaca serta memberikan maklum balas tambahan, yang akan digunakan untuk meningkatkan kualiti sistem pada masa hadapan.



Rajah 3 Antara Muka Akhir Pengguna (Contoh Maklum Balas)

Perbincangan

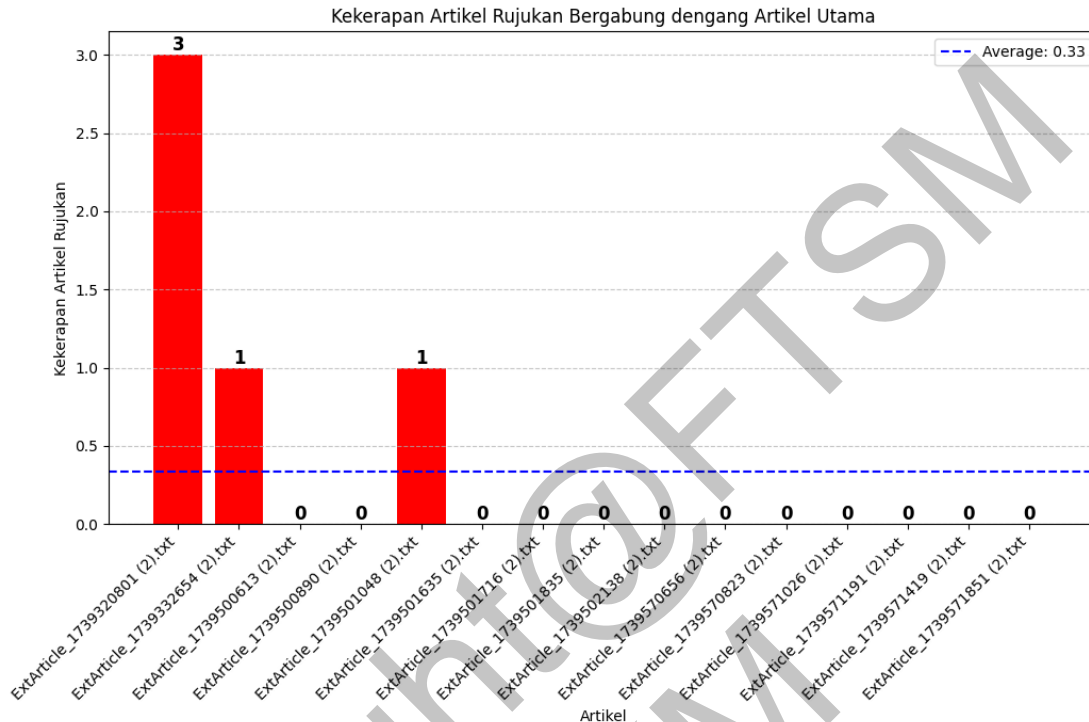
Sejumlah 15 keputusan hasil akhir ringkasan teks telah diambil berdasarkan maklum balas yang diterima melalui laman web. Hasil yang didapati terdapat beberapa laman web yang telah dicubai. Daripada jumlah 15 halaman yang diuji, Britannica.com mencatatkan bilangan tertinggi dengan 4 halaman, manakala laman lain seperti Wired.com dan Medium.com masing-masing menyumbang 2 halaman. Data ini menunjukkan bahawa model telah diuji ke atas pelbagai jenis laman web yang menawarkan kandungan yang berbeza. Berikut merupakan bilangan halaman laman web yang telah digunakan di dalam sistem.

Jadual 1 Senarai Laman Web yang dimasukkan Sebagai Pautan

Laman Web	Bilangan
Brittanica.com	4
Wired.com	2
Medium.com	2
Screenrant.com	1
Nhs.uk	1
People.cn	1
Thestar.com	1
TheVerge.com	1
Mirror.co	1
Wikipedia.com	1
Jumlah	15

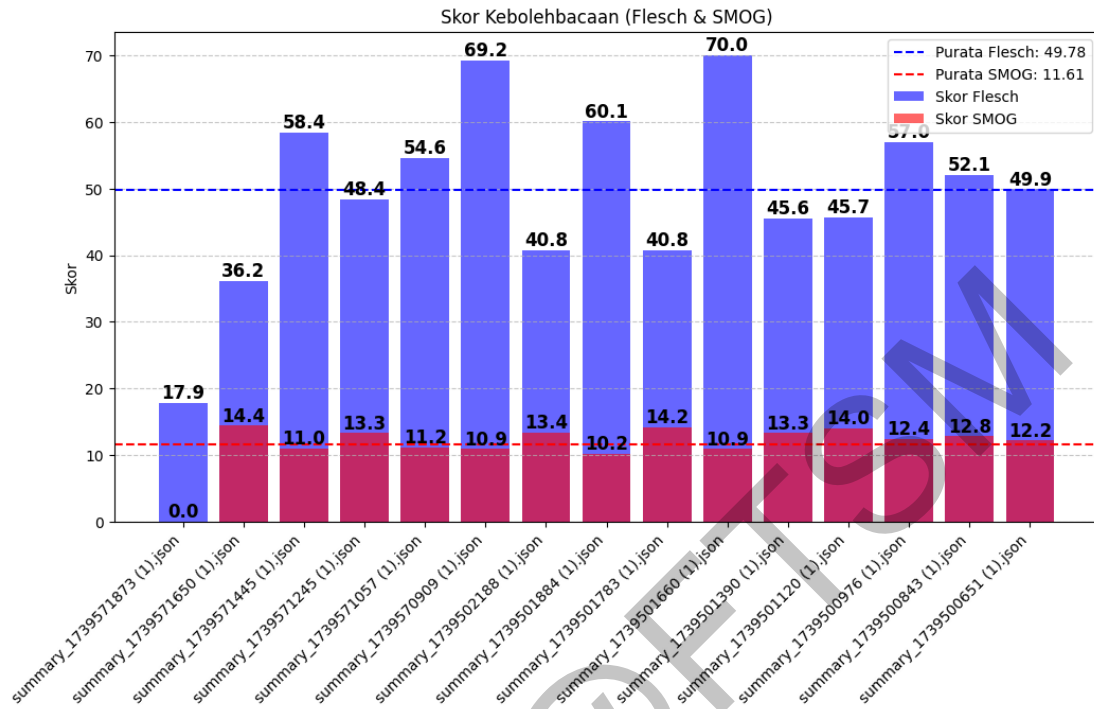
Sebelum melihat kepada hasil ringkasan model menggunakan model Transformer, satu aspek yang juga akan dilihat adalah kekerapan penggabungan artikel rujukan ke dalam artikel utama. Data menunjukkan bahawa hanya beberapa artikel mempunyai rujukan yang digabungkan; contohnya, Artikel 1 mengandungi 3 artikel rujukan manakala Artikel 2 dan

Artikel 5 mengandung 1, manakala selebihnya tidak melibatkan penggabungan rujukan. Purata kekerapan penggabungan adalah 0.33, menunjukkan bahawa kebanyakan ringkasan tidak memerlukan atau tidak menggabungkan artikel rujukan sebagai sebahagian daripada kandungan utama.



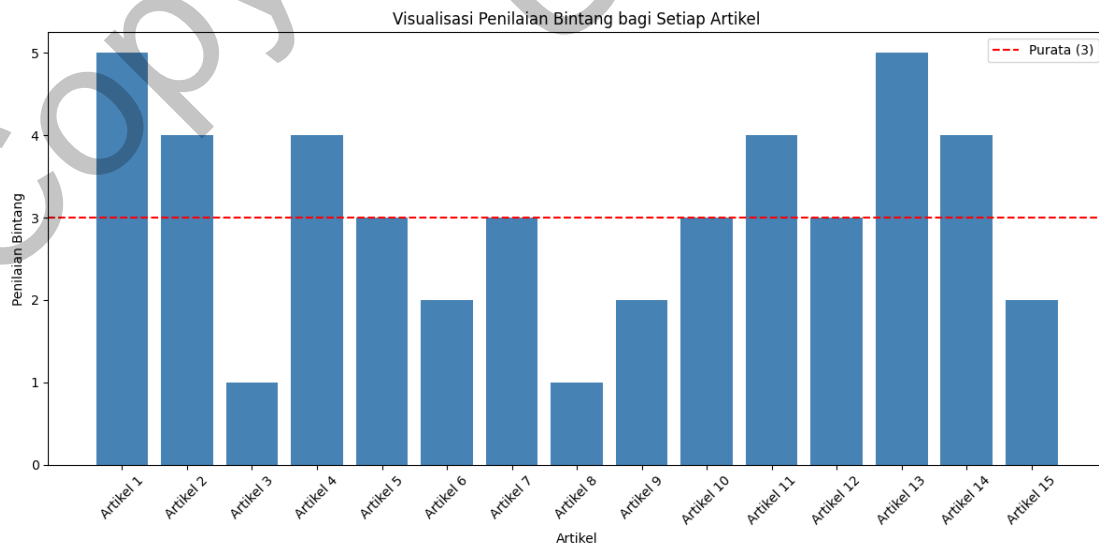
Rajah 4 Carta Palang Kekerapan Artikel Rujukan yang bergabung

Penilaian kebolehbacaan juga dilakukan menggunakan skor Flesch dan skor SMOG bagi setiap artikel ringkasan. Purata skor Flesch adalah sebanyak 49.78, manakala purata skor SMOG adalah sekitar 11.61. Skor Flesch yang rendah bagi sesetengah artikel (contohnya, Artikel 1 dengan skor 17.85) menunjukkan bahawa terdapat ringkasan yang agak mencabar untuk dibaca, sementara artikel lain mencatatkan skor yang lebih tinggi (contohnya, Artikel 10 dengan skor 70.02). Berdasarkan penilaian menggunakan metrik SMOG, skor ringkasan teks yang dihasilkan berkisar antara 10.2 hingga 14.4, dengan purata sekitar 11.61. Secara umum, nilai ini menunjukkan bahawa teks ringkasan memerlukan tahap pendidikan setaraf pelajar sekolah menengah atas untuk dapat difahami dengan baik. Skor yang lebih tinggi, seperti 14.4, mencadangkan bahawa teks tersebut agak kompleks, manakala skor yang lebih rendah menandakan kebolehbacaan yang lebih mudah. Dalam kajian ini, terdapat beberapa ringkasan yang menunjukkan skor SMOG mendekati nilai tertinggi, yang memberi isyarat bahawa meskipun ringkasan telah berjaya mengurangkan jumlah teks, kompleksiti struktur ayat masih tinggi. Ini mengakibatkan kesukaran kepada pengguna untuk segera menangkap maksud utama teks tanpa perlu meluangkan masa yang lebih lama membaca keseluruhan teks artikel.



Rajah 5 Carta Palang Skor Flesch dan SMOG

Bagi penilaian kualiti ringkasan, sistem menggunakan penilaian bintang daripada 1 hingga 5 yang boleh diberikan selepas mendapat hasil ringkasan. Purata markah keseluruhannya adalah 3, mencerminkan bahawa terdapat variasi dalam keberkesanan ringkasan yang dihasilkan oleh model. Artikel seperti Artikel 1 dan Artikel 13 yang memperoleh skor 5 menunjukkan bahawa ringkasan tersebut sangat berkesan dan memenuhi jangkaan pengguna. Sebaliknya, Artikel 3 dan Artikel 8 yang hanya mendapat skor 1 menandakan bahawa ringkasan tersebut tidak berjaya menyampaikan maklumat dengan baik atau mudah difahami oleh pengguna.



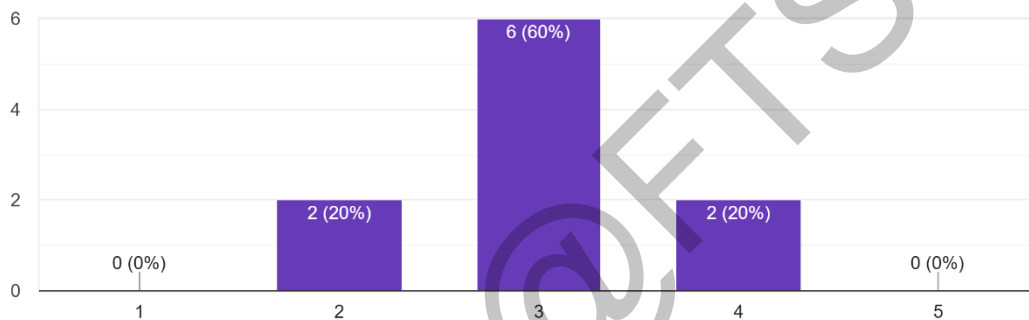
Rajah 6 Carta Palang Penilaian Bintang bagi Setiap Artikel

Ujian Kebolegunaan

Satu tinjauan telah dilakukan bagi memastikan sistem ringkasan teks MyWebSummarizer memenuhi keperluan dan jangkaan pengguna. Beberapa soalan telah diajukan, dengan penekanan utama diberikan kepada masa yang diambil untuk menghasilkan ringkasan teks. Berdasarkan maklum balas yang diterima, 60% daripada responden menilai kelajuan sistem sebagai neutral, menunjukkan bahawa tempoh masa untuk menghasilkan ringkasan tidak terlalu lama tetapi juga tidak terlalu cepat. Manakala, 2 responden merasakan proses ini agak lama dan 2 responden lagi merasakan ia cepat.

Berapa lama masa yang diambil untuk menghasilkan ringkasan teks?

10 responses



Rajah 7 Carta Palang bagi Masa yang diambil untuk menghasilkan Ringkasan Teks

Cadangan Penambahbaikan

Hasil dari keputusan ujian keberkesanan model dan kebolegunaan sistem, beberapa cadangan penambahbaikan boleh dilakukan pada masa hadapan. Antara cadangan yang boleh dilaksanakan adalah meningkatkan keupayaan sumber pemprosesan, seperti menggunakan GPU berkuasa tinggi, bagi mengurangkan masa penjanaan ringkasan teks. Bagi meningkatkan kualiti ringkasan teks pula, dari segi teknik pengekstrakan, pembaikan diperlukan untuk memastikan data yang diekstrak lebih bersih dan bebas daripada unsur tidak relevan. Ini boleh dilakukan dengan integrasi teknik topic modelling dapat membantu memastikan bahawa ringkasan yang dihasilkan lebih betul-betul relevan dengan kandungan asal artikel. Kajian yang lebih mendalam mengenai kualiti integrasi antara artikel utama dan rujukan juga perlu dilakukan bagi menentukan sejauh mana ia benar-benar membantu dalam memberikan konteks tambahan kepada artikel. Model Transformer juga perlu dilatih semula khusus untuk konteks gabungan artikel seperti ini bagi memastikan penjanaan ringkasan teks yang lebih efektif dan berkualiti tinggi.

KESIMPULAN

Sistem Ringkasan Teks bagi Web Artikel telah berjaya dihasilkan dan mencapai objektif yang diinginkan. Sistem ini membantu pengguna mendapatkan maklumat dengan lebih jelas dalam bentuk yang lebih padat. Dengan penggunaan model Transformer PEGASUS dan T5, sistem ini dapat menghasilkan ringkasan jenis abstraktif. Hasil kajian menunjukkan bahawa sistem

ini masih mampu meringkaskan teks dengan berkesan dan memberikan kemudahan kepada pengguna dalam mendapatkan maklumat dengan lebih padat. Walau bagaimanapun, keberkesanan ringkasan yang dihasilkan menunjukkan variasi yang berbeza-beza, menyebabkan pengalaman pengguna tidak sepenuhnya konsisten. Penilaian kualiti ringkasan berdasarkan bintang mencatatkan purata sebanyak 3 daripada 5, yang menunjukkan bahawa sebahagian besar pengguna masih berpuas hati, walaupun terdapat ringkasan yang kurang tepat atau kurang jelas. Dari segi kebolehbacaan, purata skor SMOG adalah 49.78 dan purata skor Flesch adalah 11.62, menandakan bahawa tahap kesukaran pembacaan yang tidak seragam, sesetengah ringkasan mudah difahami, manakala yang lain adalah lebih kompleks.

Kekuatan Sistem

Sistem ringkasan teks, MyWebSummarizer, telah menunjukkan bahawa rata-rata pengguna berpuas hati dengan antara muka laman web serta pengalaman penggunaannya yang mesra. Teks yang dihasilkan umumnya boleh difahami dan diringkaskan dengan baik, menjadikannya lebih mudah untuk pengguna mendapatkan maklumat dengan cepat. Penilaian bintang yang diberikan oleh pengguna mencerminkan maklum balas positif daripada majoriti responden, menunjukkan bahawa sistem berjaya memenuhi keperluan utama mereka dalam mengekstrak maklumat penting daripada artikel.

Kelemahan Sistem

Proses penjanaan ringkasan mengambil masa yang agak lama, mengakibatkan pengalaman pengguna terjejas, terutamanya apabila mereka memerlukan maklumat dengan segera. Punca yang menjadikan kelewatan ini adalah kerana sistem kekurangan sumber pemprosesan yang mencukupi. Selain itu, oleh kerana belum ada sistem lain yang melakukan pendekatan ini untuk menggabungkan artikel rujukan dan artikel utama, pengguna mungkin kurang memahami manfaatnya, menyebabkan kesukaran dalam menilai keberkesanan sistem. Tambahan pula, teknik pengekstrakan teks masih belum mencapai tahap yang ideal kerana sistem kadangkala menyertakan teks yang tidak relevan, seperti komen pengguna atau iklan.. Kualiti ringkasan teks yang dihasilkan pula berbeza-beza setiap kali dihasilkan. Ini menunjukkan bahawa sistem masih kurang konsisten dalam menghasilkan ringkasan yang tepat dan berkualiti bagi pengguna.

PENGHARGAAN

Penulis kajian ini ingin mengucapkan setinggi-tinggi penghargaan ditujukan kepada penyelia kajian, Dr. Fadilla 'Atyka Nor Rashid, atas segala nasihat, bimbingan, dan dorongan yang telah diberikan sepanjang pembangunan projek ini. Tanpa berkat kesabaran dan kefahaman beliau terhadap kelemahan dan kekurangan penulis, serta membantu mengatasi masalah penulis, projek ini tidak berjaya.

Penulis kajian ini juga ingin mengucapkan terima kasih kepada semua pihak yang membantu secara langsung mahupun tidak langsung dalam menyempurnakan projek ini. egala sokongan dan kerjasama yang diberikan amat dihargai, kerana tanpa bantuan mereka, projek ini tidak dapat direalisasikan dengan baik. Semoga segala usaha mereka mendapat keberkatan Tuhan.

RUJUKAN

- Devlin, J. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. <https://arxiv.org/abs/1810.04805v2>.
- George, S. 2022. AUTOMATED SUMMARIZATION OF RESTAURANT REVIEWS USING HYBRID APPROACHES. | ICTACT Journal on Soft Computing | EBSCOHost. openurl.ebsco.com.
- Guo, M., Ainslie, J., Uthus, D., Ontanon, S., Ni, J., Sung, Y.-H. & Yang, Y. 2021. LongT5: efficient Text-To-Text transformer for long sequences. *arXiv (Cornell University)*.
- Lewis, M.A., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V. & Zettlemoyer, L. 2019. BART: Denoising Sequence-to-Sequence Pre-training for natural language generation, Translation, and Comprehension. *arXiv (Cornell University)*.
- Lin, C.-Y. 2004. ROUGE: a package for automatic evaluation of summaries. <https://aclanthology.org/W04-1013>.
- Luhn, H.P. 1958. The automatic creation of literature abstracts. *IBM Journal of Research and Development* 2(2): 159–165.
- McKeown, K., Barzilay, R., Evans, D.K., Hatzivassiloglou, V., Klavans, J.L., Nenkova, A., Sable, C., Schiffman, B. & Sigelman, S. 2002. Tracking and summarizing news on a daily basis with Columbia's Newsblaster. .
- McKeown, K., Passonneau, R.J., Elson, D.K., Nenkova, A. & Hirschberg, J. 2005. Do summaries help? .
- Mc Laughlin, G. H. (1969). SMOG Grading-a New Readability Formula. *Journal of Reading*, 12(8), 639–646.
- Nenkova, A. 2011. Automatic summarization. *Foundations and Trends in Information Retrieval* 5(2): 103–233.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W. & Liu, P.J. 2019. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *arXiv (Cornell University)*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. & Polosukhin, I. 2017. Attention is all you need. *arXiv (Cornell University)*.
- Zhang, J., Zhao, Y., Saleh, M. & Liu, P.J. 2019. PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. *arXiv (Cornell University)*.

Osman bin Ismail (A190073)

Dr. Fadilla 'Atyka Nor Rashid

Fakulti Teknologi & Sains Maklumat

Universiti Kebangsaan Malaysia