



Applying Antconc in Regex Search on BBC News Text

Annisah Inriani Harahap^{*1} , Samuel Nugraha Cristy² , Sultan Salma Efendi³ ,
 Tengku Silvana Sinar⁴ 

¹Universitas Sumatera Utara, Medan, 20155, Indonesia

²Universitas Sumatera Utara, Medan, 20155, Indonesia

³Universitas Sumatera Utara, Medan, 20155, Indonesia

⁴Universitas Sumatera Utara, Medan, 20155, Indonesia

ARTICLE INFO

Article history:

Received 24 January 2023

Revised 30 April 2023

Accepted 2 December 2023

Available online 29 February
 2024

E-ISSN: 2964-1713

P-ISSN: 2775-5622

How to cite:

Annisah, I. H., Samuel, N. C., Sultan, S. E.(2024). Applying Antconc in Regex Search on BBC News Text. *Journal of Linguistics and Literary Research*, 5(1), 38-47.

ABSTRACT

The software AntConc has become a freeware and multi-platform tool used in corpus-based linguistic research. Corpus linguistics itself is identified as a methodology consisting of various related methods that can be used by scholars with different theoretical tendencies. This research aims to determine the formulas used in searching regular expressions in news articles. The news data was collected from a known media source called BBC News. The research results indicate that Regex in Antconc can be performed in various ways, including: (1) To find words starting with 'sp,' the formula $\wedge\text{sp}[a-z]$ can be used, (2) To find words starting with 'st' and ending with 'ed,' the formula $\wedge\text{st}[a-z]^*\text{ed}$ can be used, (3) To find words ending with '-ing,' the formula $[a-z]^*\text{ing}$ can be used, (4) To find a phrase containing two words, with the first word starting with the letter 's' and the second word starting with the letter 't,' the formula $\wedge\text{s}[a-z]^*\wedge\text{t}[a-z]$ can be used.

Keyword: Antconc, BBC News, Regex

ABSTRAK

Software AntConc menjadi alat freeware dan multiplatform yang digunakan dalam penelitian linguistik korpus berbasis data. Linguistik korpus sendiri diidentifikasi metodologi yang terdiri dari sejumlah metode terkait yang dapat digunakan oleh para sarjana dari banyaknya kecenderungan teoretis yang berbeda. penelitian ini bertujuan untuk mengetahui rumus apa saja yang digunakan dalam pencarian regular expression dalam berita. Data berita tersebut diambil pada sumber media yang dikenal sebagai BBC News. Hasil penelitian menunjukkan bahwa Regex pada Antconc dapat dilakukan dengan berbagai cara, antara lain: (1) Untuk menemukan kata yang diawali dengan 'sp' dapat dilakukan dengan rumus $\wedge\text{sp}[a-z]$, (2) untuk menemukan kata yang diawali dengan "st" dan berakhiran "ed" dapat dilakukan dengan rumus $\wedge\text{st}[a-z]^*\text{ed}$, (3) Untuk menemukan kata yang berakhiran "-ing" dapat dilakukan dengan rumus $[a-z]^*\text{ing}$, (4) Untuk menemukan Sebuah Frasa yang mengandung dua kata dengan kata pertama diawali dengan huruf "s" lalu dengan kata kedua diawali dengan huruf "t" dapat dilakukan dengan rumus $\wedge\text{s}[a-z]^*\wedge\text{t}[a-z]$.

Keyword: Antconc, BBC News, Regex



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International.

<https://doi.org/10.32734/lingtersa.v5i1.12364>

1. Pendahuluan

Dalam dunia pendidikan saat ini, evolusi teknologi membuktikan bahwa teknologi merupakan salah satu media pembelajaran yang paling efektif. Menurut Breyer (2008 dalam khairas, 2019: 189), media pembelajaran yang menggunakan teknologi dapat meningkatkan motivasi siswa dan menurunkan tingkat kebosanannya dalam belajar. Siswa dapat menggunakan perangkat lunak untuk proses pembelajaran. Sehingga, dapat membantu mereka untuk mendapatkan pemahaman lebih lanjut tentang topik yang sedang dipelajari.

Salah satu media pembelajaran yang menggunakan teknologi adalah *Software AntConc*. Perangkat lunak *Antconc* adalah alat yang dapat digunakan bagi siswa untuk memahami pembelajaran suatu bahasa. *Software AntConc* menjadi alat freeware dan multiplatform yang digunakan dalam penelitian linguistik korpus berbasis data (Anthony, 2012). Linguistik korpus didefinisikan, dalam dua dekade terakhir, sebagai pendekatan berbasis data yang inovatif dan canggih untuk menganalisis bahasa, Meskipun dianggap oleh banyak orang sebagai metode analisis linguistik (lihat Biber, 1998; Cheng, 2012). Korpus hampir tidak berguna tanpa semacam alat perangkat lunak komputer untuk memprosesnya dan menampilkan hasil dengan cara yang dapat dimengerti. Dua yang paling populer diantaranya adalah *MonoConc Pro*, dan *WordSmith Tools*, meskipun banyak concordancer dan program analisis korpus lainnya juga telah dikembangkan (Anthony, 2004:3).

dengan demikian, Linguistik korpus merupakan metodologi yang terdiri dari sejumlah metode terkait yang dapat digunakan oleh para sarjana dari banyaknya kecenderungan teoretis yang berbeda. Di sisi lain, tidak dapat disangkal bahwa korpus linguistik juga sering dikaitkan dengan pandangan bahasa tertentu. Inti dari pandangan ini adalah bahwa aturan bahasa didasarkan pada penggunaan dan bahwa perubahan terjadi ketika penutur menggunakan bahasa untuk berkomunikasi satu sama lain (Hans Lindquist, 2009: 1-2). Linguistik korpus mempertimbangkan tidak hanya ucapan, tetapi juga bahasa, bisa dibilang, dalam sudut pandang baru, memberikan visi baru dan potensi yang lebih kuat, yang berasal dari jumlah sampel bahasa belaka, skala cakupan yang berkembang (Gvishiani, 2008).

Software AntConc merupakan seperangkat alat analisis korpus umum yang dapat diterapkan pada sistem Windows, Linux dan Macintosh di komputer. Alat ini telah digunakan oleh penerjemah, guru bahasa dan pelajar di seluruh dunia. Alat ini tidak hanya menganalisis teks bahasa Inggris tetapi juga bahasa lain seperti Jepang, Korea, dll, termasuk bahasa Indonesia. Salah satu fitur dalam alat ini adalah *Concordancer Tool* yang digunakan untuk melihat kata-kata yang digunakan dalam sebuah teks (khairas, 2019: 189).

Concordancer Tool adalah salah satu alat utama yang digunakan dalam sebagian besar perangkat lunak analisis korpus yang dapat digunakan untuk memfasilitasi pembelajaran kosakata bagi pelajar bahasa. Sun & Wang (2003) sebagaimana dikutip oleh Anthony (2004) menyatakan bahwa alat ini telah terbukti sebagai alat bantu yang efektif untuk pemerolehan bahasa kedua atau asing dan pembelajaran kosa kata, kolokasi, tata bahasa dan gaya penulisan. Di *AntConc*, *Concordancer Tool* dimasukkan ke layar utama sehingga pengguna dapat mengoperasikannya secara langsung. Konkordansi Ini akan menunjukkan kepada Kata Kunci dalam tampilan Konteks/ *Keyword in Context* (disingkat KWIC), menggunakan bilah pencarian di bawahnya (azmovich, 2019).

Sebagian besar kumpulan leksikal saat ini diberi tanda POS (*POS-tagged*). bahkan korpus ahli bahasa yang bekerja dengan '*unannotated plain text*' pasti menerapkan beberapa metode untuk memisahkan istilah yang mereka minati dari kata-kata di sekitarnya. Dalam hal itu, anotasi dan abstraksi digabungkan dalam pencarian leksikal (Johanson, 1991).

Software ini juga dapat mencari regular expression (regex) dalam suatu kumpulan teks, termasuk teks berita yang termuat dalam media berita BBC News. BBC News merupakan salah satu lembaga penyiaran (broadcasting) yang tertua dan terbesar di dunia. Media ini juga telah menyediakan web portal yang dapat diakses dengan mudah (pascarina, 2018). Doumit dan Minai (2012) menjadikan media tersebut sebagai data dari berita-berita selama bulan Maret 2011. Korpus tersebut menghasilkan label topik yang berhubungan dengan pemberontakan Libya, gempa bumi Jepang, dan beberapa cerita lainnya. Ukuran dan kompleksitas pola jaringan menunjukkan tingkat detail dan signifikansi yang dialokasikan setiap sumber berita pada berita tersebut. Dengan sumber data yang sama, penelitian ini berfokus pada penemuan rumus untuk mencari regular expression (regex) dalam kumpulan teks berita BBC News.

Penelitian ini berangkat dari keingintahuan peneliti yang mengkaji seputar linguistik korpus menggunakan *software Antconc*. Kajian tersebut dilandaskan dengan menerapkan pencarian regex pada berita BBC News. Sebagai langkah awal, kami mengacu pada 2 penelitian landasan: yaitu penelitian rubbani dkk (2014) yang menggunakan *software Antconc* sebagai media pembelajaran bahasa inggris dan penelitian dan anthony (2004) yang mengkaji pengantar *software Antconc*.

Rubbani dkk (2014) menggunakan *software Antconc 3.2.4* sebagai alat penelitian. Untuk itu, sebuah korpus besar telah dikompilasi dan dianalisis melalui perangkat lunak Antconc 3.2.4. Dengan bantuan ekspresi reguler di Antconc, fitur sintaksis yang berbeda dari simile telah dianalisis. Penelitian ini telah menciptakan banyak twist dengan memberikan perintah melalui regular expression (regex). Didukung dengan temuan anthony (2004) yang menjelaskan masing-masing fitur dalam software tersebut serta keterbatasannya.

2. Metodologi Penelitian

Dalam langkah-langkah komprehensif, penelitian ini menggunakan metode gabungan kualitatif dan kuantitatif. Holmes & Meyerhoff (2003) menyatakan bahwa metode gabungan ini memanfaatkan sejumlah pola yang teridentifikasi lewat analisis kuantitatif yang menjadi landasan penting untuk memahami teks secara kualitatif secara mendalam. Terkait dengan penelitian linguistik korpus, Kwary dan Arum (2011: 213) menyebutkan bahwa penelitian linguistik korpus perlu menggunakan desain metode gabungan karena data statistik yang diperoleh dari analisis korpus perlu ditafsirkan lebih lanjut dengan menggunakan pertimbangan-pertimbangan kualitatif.

sampel tulisan atau ujaran L2 tertentu dianalisis sehubungan dengan semua kolokasi yang terjadi dalam teks yang tersedia. Metode lain yaitu dengan menggunakan kumpulan tulisan non-native, dalam hal ini dimungkinkan untuk menganalisis hanya kolokasi spesifik yang telah dipilih sebelumnya seperti yang terjadi dalam rentang teks seperti yang dilakukan dalam penelitian oleh Granger (1998). Dengan memperoleh konkordansi untuk butir-butir yang diselidiki, pola kolokasi teks non-native dapat dibandingkan dengan teks yang diproduksi oleh penutur asli. Studi ini mencoba untuk mengumpulkan data melalui rata-rata corpora berdasarkan tulisan akademis non-native tingkat lanjut. Metode ini disebut sebagai linguistik korpus (faizah, 2018: 49)

Data dalam penelitian ini diperoleh dari sejumlah pola dan rumus dalam pencarian regex pada berita kesehatan (health related) BBC News dalam bahasa inggris. Secara spesifik, datanya terdiri leksikal yang diawali dengan ‘sp’, leksikal yang diawali dengan “st” dan berakhiran “ed”, leksikal yang berakhiran “-ing”, Frasa yang mengandung dua kata dengan kata pertama diawali dengan huruf “s” lalu dengan kata kedua diawali dengan huruf “t”. Perangkat lunak yang akan digunakan dalam penelitian ini adalah *software* korpus bernama AntConc Tools Versi 3.4. Perangkat lunak ini dapat mengidentifikasi pola kata dalam teks dengan menggunakan beberapa fitur utamanya.

3. Pembahasan

AntConc adalah salah satu aplikasi populer dalam analisis korpus linguistik yang dapat membantu pengguna dalam menganalisis dan memproses teks atau data bahasa. Aplikasi ini dapat digunakan untuk mengumpulkan data dari berbagai sumber, termasuk web, dokumen teks, dan file lainnya, serta memberikan berbagai alat dan fitur yang berguna untuk analisis dan visualisasi data.

Beberapa fitur yang disediakan oleh AntConc antara lain:

- a) Pencarian dan analisis konkordansi
Fitur ini memungkinkan pengguna untuk mencari kata-kata atau frasa tertentu dalam teks atau data bahasa, dan menampilkan hasil pencarian dalam format konkordansi. Dalam format ini, kata atau frasa yang dicari ditampilkan bersama dengan konteksnya dalam teks.
- b) Analisis frekuensi kata
AntConc juga dapat menghitung frekuensi kata dalam teks atau korpus, dan menampilkan hasil dalam bentuk grafik atau tabel.
- c) Analisis kolokasi
Fitur ini dapat membantu pengguna untuk menemukan kata-kata yang cenderung muncul bersama-sama dalam teks atau korpus tertentu. Hal ini dapat membantu pengguna saat mengidentifikasi pola dan makna yang terkait dengan kata-kata tertentu.
- d) Pencarian dan penggantian teks
Fitur ini memungkinkan pengguna untuk mencari dan mengganti teks dalam korpus, serta mengatur penggantian teks secara otomatis atau manual.
- e) Visualisasi data
AntConc juga menyediakan berbagai jenis grafik dan visualisasi data, seperti histogram dan plot garis, untuk membantu pengguna dalam menganalisis dan memvisualisasikan data bahasa.

Dalam keseluruhan, *AntConc* merupakan alat yang sangat berguna untuk analisis korpus linguistik, baik untuk tujuan akademis maupun profesional. Aplikasi ini mudah digunakan dan memberikan berbagai fitur yang berguna untuk analisis dan visualisasi data bahasa, sehingga sangat membantu dalam menghasilkan hasil penelitian yang berkualitas dan relevan.

Aplikasi *Antconc* ini akan digunakan dalam mencari dan mengidentifikasi regular expression (regex) berdasarkan data yang berupa teks berita dari BBC News dalam jumlah banyak tentunya. Dengan mencari regex dalam aplikasi AntConc, pengguna dapat mengidentifikasi pola-pola dan tren yang terkait dengan data bahasa yang sedang dianalisis. Ini dapat membantu pengguna dalam mengambil kesimpulan dan membuat generalisasi tentang data bahasa yang sedang dianalisis, serta membantu dalam mengidentifikasi makna yang terkait dengan kata-kata atau frasa tertentu.

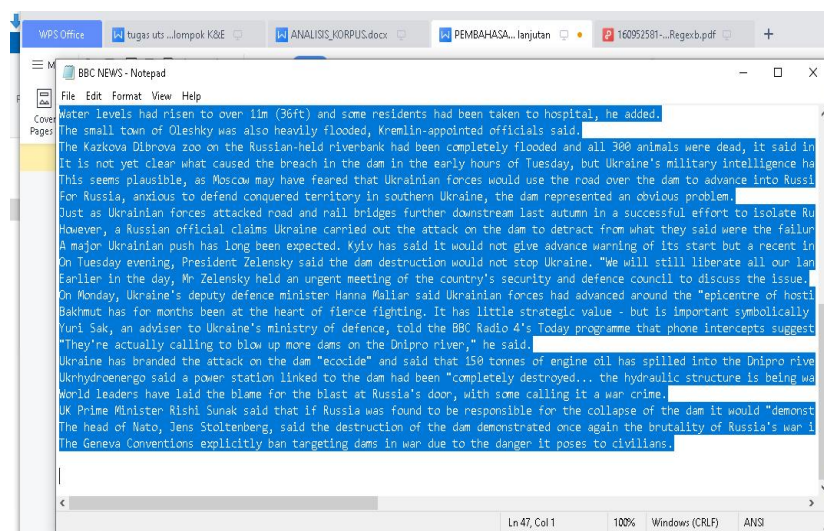
Contoh penggunaan regex dalam AntConc adalah sebagai berikut:

- i. Mencari kata-kata tertentu dalam teks atau korpus, seperti mencari semua kata yang dimulai dengan huruf "A".
- ii. Mencari frasa tertentu dalam teks atau korpus, seperti mencari semua frasa yang mengandung kata "pemerintah".
- iii. Mencari pola sintaksis tertentu dalam teks atau korpus, seperti mencari semua kalimat yang mengandung klausa terikat "jika".

Sehingga dengan mengidentifikasi regex, pengguna dapat melakukan analisis yang lebih tepat dan spesifik pada teks atau data bahasa yang sedang dianalisis, dan dapat membantu pengguna dalam mengambil kesimpulan yang lebih akurat dan relevan.

Adapun tahapan yang harus dilakukan dalam mencari dan mengidentifikasi regex pada aplikasi *Antconc* akan dipaparkan sebagai berikut. Tahap pertama pengumpulan data dengan menyalin artikel berita satu per satu ke dalam dokumen yang akan diubah nantinya ke dalam format txt. File berbentuk teks dokumen (txt) adalah format data yang dapat diolah oleh perangkat lunak komputer pengolah korpus. Data tersebut kemudian dibuat menjadi korpus linguistik. Aplikasi pengolah korpus linguistik yang digunakan yaitu AntConc, alat analisis korpus gratis untuk konkordansi dan analisis teks (Anthony, 2011).

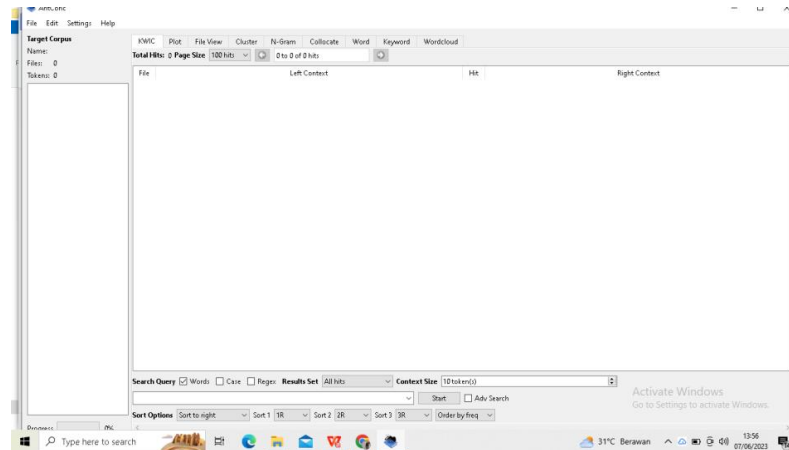
Langkah selanjutnya ialah membuka AntConc dan impor teks atau korpus yang ingin Anda analisis. Teks yang menjadi data disini adalah teks kumpulan berita BBC NEWS yang telah diubah ke dalam bentuk txt dan disimpan ke dalam file. Untuk menyimpan data yang telah diubah ke dalam bentuk txt adalah dengan mengklik menu 'file' lalu klik 'open file' selanjutnya pilih file yang akan dijadikan korpus dan langkah terakhir adalah pilih menu 'open'. Setelah itu menyimpan data tersebut peneliti dapat melakukan analisis berdasarkan corpus files yang telah dimasukkan.



Gambar 1. Lampiran berita dari media berita BBC News

Diatas merupakan lampiran berita yang peneliti ambil dari media berita BBC News dan telah diubah dan di simpan dalam bentuk txt.

Kemudian untuk mencari reguler expression dalam teks berita yang telah ditentukan, maka buka software Antcont dari leptop atau komputer anda masing masing. Perhatikan gambar berikut. Seperti inilah tampilan menu dalam software Antconc pada laman utama.

Gambar 2. Laman utama *software antconc*

Untuk melakukan semua pencarian dengan regular expression (regex), berikut langkah-langkahnya:

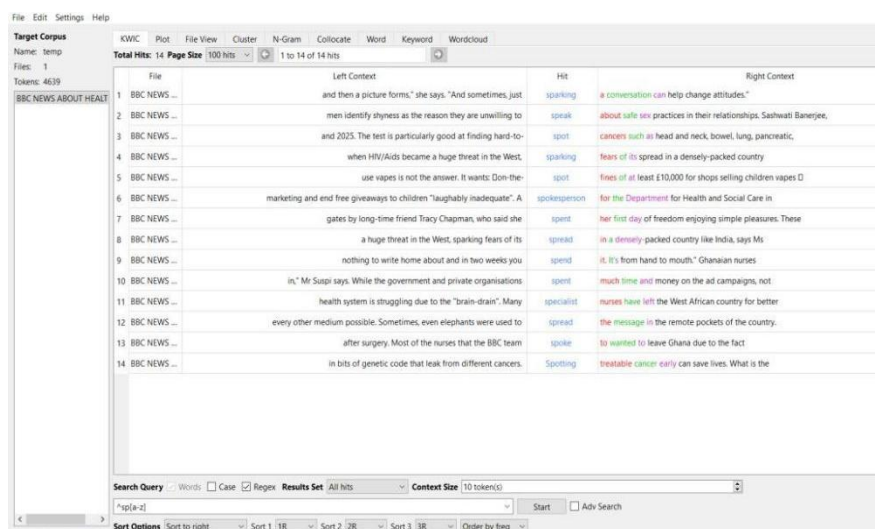
- 1) Klik pada menu "File".
- 2) Pilih opsi "Open Dir".
- 3) Pindah ke direktori "cartas&memos" atau "text_gen".
- 4) Pilih direktori yang diinginkan dan buka.
- 5) Di antara bagian "Files Processed" dan tombol "Start", centang kotak yang berlabel "Regex" (Ini memungkinkan pencarian menggunakan ekspresi reguler yang akan Anda ketikkan untuk meninjau tata bahasa.).
- 6) Klik pada tulisan biru hasil pencarian untuk melihat teks lengkap.
- 7) Untuk kembali ke daftar contoh, klik tab "Concordancer".

(Catatan: Direktori "cartas&memos" hanya berisi teks surat dan memorandum, sedangkan direktori "text_gen" berisi semua jenis teks dan berjumlah lebih dari 1juta kata)

Maka fokus penelitian ini adalah mencari regular expression (regex) pada data berita yang diambil dari media berita BBC NEWS dengan pola dan rumus sebagai berikut:

1. Kata yang diawali imbuhan “sp” Didalam bahasa Inggris yang ada di didalam BBC NEWS (Health Related) (versi 4.2.0)

Regex : ^sp[a-z]
 Jumlah : 14 hits
 Dari : 4639 token



Gambar 3. Regex pada leksikal yang diawali dengan imbuhan "sp"

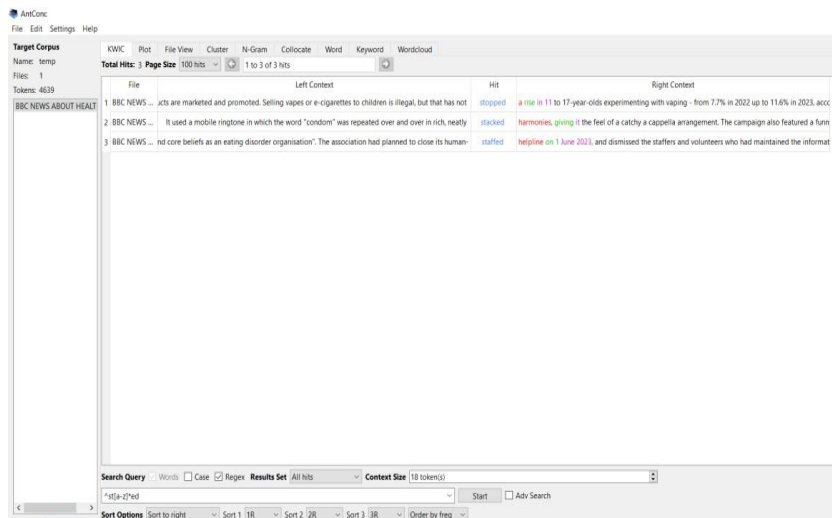
Pencarian regex yang pertama pada kata yang diawali dengan imbuhan "sp" dalam bahasa Inggris. Data yang dicari adalah kata-kata yang diawali dengan imbuhan "sp" dalam bahasa Inggris yang terdapat pada kumpulan berita BBC NEWS. Imbuhan "sp" diikuti oleh huruf kecil. Contoh kata-kata yang cocok dengan pola ini adalah "special" dan "spirit".

Dengan menggunakan regex `^sp[a-z]`, pola tersebut akan mencocokkan kata-kata yang dimulai dengan "sp" diikuti oleh huruf kecil apa pun. Simbol `^` menunjukkan bahwa pencarian harus dimulai dari awal kata, kemudian "sp" harus diikuti oleh satu huruf kecil [a-z]. Misalnya, jika kita menerapkan regex ini pada teks berita dari BBC NEWS (Health Related), kita dapat menemukan kata-kata seperti "specialist", "spontaneous", "sport", dan lainnya yang diawali dengan "sp" dan diikuti oleh huruf kecil apa pun.

Dengan menggunakan aplikasi AntConc, pengguna dapat dengan mudah mencari kata-kata yang memenuhi pola ini dalam teks berita dari BBC NEWS (Health Related) dan mengidentifikasi pola-pola dan tren terkait dengan kata-kata yang diawali dengan "sp" dalam bahasa Inggris.

2. Kata yang diawali "st" dan diakhiri dengan "ed" yang ada di dalam BBC NEWS (Health Related) (versi 4.2.0)

Regex : `^st[a-z]*ed`
 Jumlah : 3 hits
 Dari : 4639 Tokens



Gambar 4. Regex pada leksikal yang berimbuhan "sp" diawal dan "ed" diakhir

Pencarain regex kedua pada kumpulan berita BBC NEWS yaitu mencari kata yang diawali dengan "st" dan diakhiri dengan "ed": Untuk mencari pola ini, peneliti menggunakan regex `^st[a-z]*ed`. Pola ini akan mencocokkan kata-kata yang diawali dengan "st" dan diakhiri dengan "ed", di mana huruf di antara keduanya dapat berupa kombinasi dari huruf kecil. Contoh kata yang cocok dengan pola ini adalah "started" dan "studied".

Pola regex yang digunakan adalah `^st[a-z]*ed`, yang akan mencocokkan kata-kata yang memenuhi kriteria tersebut. Pola regex tersebut terdiri dari beberapa elemen. Simbol `^` menunjukkan bahwa pencarian harus dimulai dari awal kata. Kemudian, "st" mengindikasikan bahwa kata yang dicari harus diawali dengan kombinasi huruf "st". Setelah itu, `[a-z]*` memungkinkan adanya kombinasi huruf kecil di antara "st" dan "ed". Akhirnya, "ed" menunjukkan bahwa kata yang dicari harus diakhiri dengan "ed".

Sebagai contoh, kata-kata seperti "started" dan "studied" memenuhi pola ini karena keduanya diawali dengan "st" dan diakhiri dengan "ed". Dengan menggunakan aplikasi AntConc, pengguna dapat mencari kata-kata yang memenuhi pola ini dalam teks berita dari BBC NEWS (Health Related) dan mengidentifikasi pola serta tren yang terkait dengan kata-kata tersebut.

3. Kata yang diakhiri dengan "ing" yang ada di dalam BBC NEWS (Health Related) (versi 4.2.0)

Regex : `^sp[a-z]`
 Jumlah : 14 hits

Dari : 4639 tokens



Gambar 5. Regex pada leksikal yang berimbuhan "sp" diawal dan "ed" diakhir

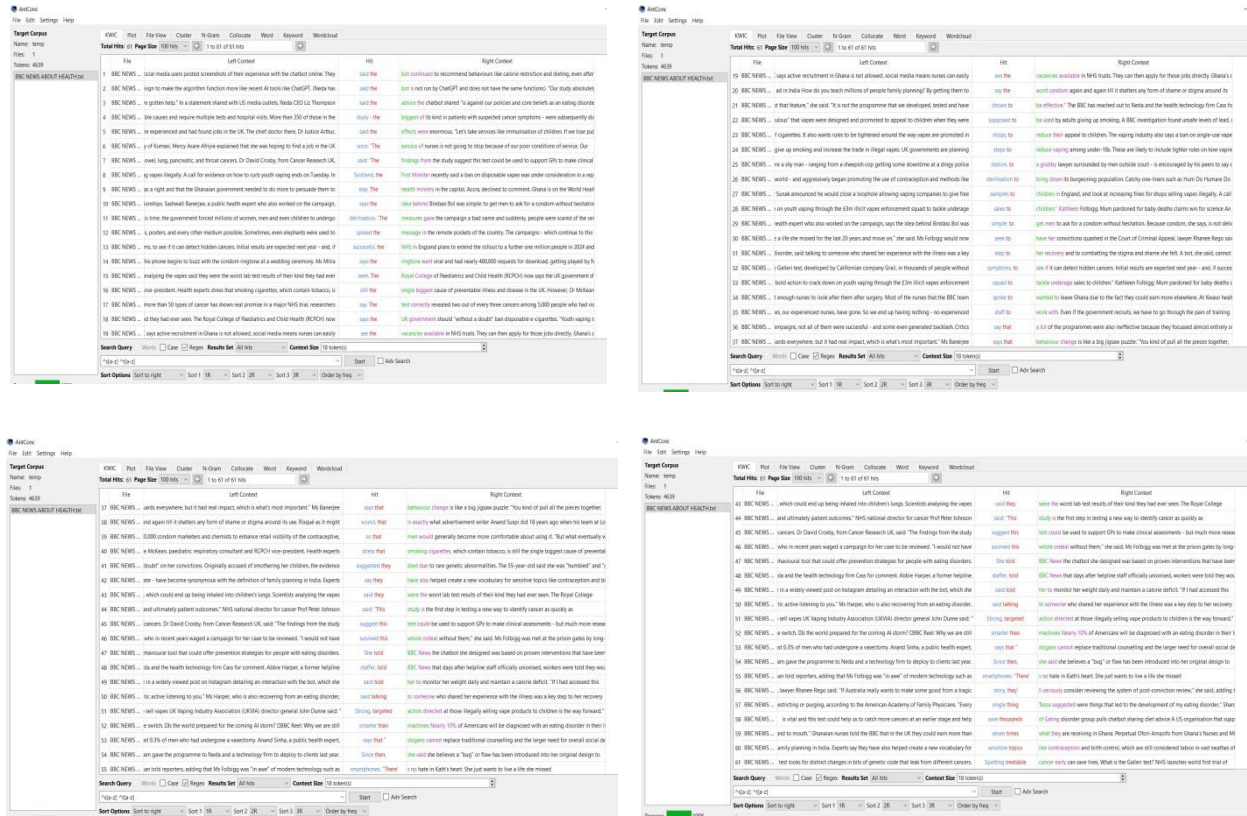
Pola regex berikutnya adalah [a-z]*ing, yang digunakan untuk mencari kata-kata yang diakhiri dengan "ing" dalam teks berita dari BBC NEWS (Health Related). Pola ini akan mencocokkan kata-kata yang berakhir dengan "ing". Pola regex [a-z]*ing mencocokkan kata-kata yang berakhir dengan "ing". Pola ini membantu dalam mengidentifikasi kegiatan atau proses yang sedang berlangsung dalam konteks kesehatan. Dalam teks berita BBC NEWS (Health Related), kata-kata yang memenuhi pola ini mungkin merujuk pada berbagai aktivitas atau proses yang sedang berlangsung yang berkaitan dengan topik kesehatan.

Contoh kata-kata yang dapat ditemukan menggunakan pola regex ini adalah "running", "building", "monitoring", "studying", dan sebagainya. Dengan mengidentifikasi kata-kata tersebut, pengguna dapat melihat tren atau pola aktivitas tertentu yang sedang dibahas dalam teks berita kesehatan. Hal ini dapat membantu memahami tren atau pola aktivitas tertentu yang sedang dibahas dalam teks berita kesehatan. Hal ini dapat membantu memahami tren atau pola aktivitas tertentu yang sedang dibahas dalam teks berita kesehatan.

Melalui penggunaan pola regex [a-z]*ing dan analisis lebih lanjut dengan AntConc, penelitian ini memberikan wawasan tentang kegiatan atau proses yang sedang berlangsung dalam teks berita kesehatan dari BBC NEWS (Health Related). Informasi ini dapat digunakan untuk mendapatkan pemahaman yang lebih baik tentang topik kesehatan yang dibahas dalam teks tersebut dan memperoleh gambaran yang lebih lengkap tentang tren dan pola kegiatan dalam konteks kesehatan.

4. Sebuah Frasa yang mengandung dua kata dengan kata pertama diawali dengan huruf “s” lalu dengan kata kedua diawali dengan huruf “t” yang berada di dalam BBC NEWS (Health Related) (versi 4.2.0)
 Regex: `^s[a-z]^t[a-z]`

Jumlah : 61 Hits
Dari: 4639 Tokens



Gambar 6. Regex pada Frasa yang mengandung dua kata berawalan “s” dan “t”

Pola regex $^s[a-z] ^t[a-z]$ digunakan dalam penelitian ini untuk mencari frasa yang terdiri dari dua kata, di mana kata pertama diawali dengan huruf "s" dan kata kedua diawali dengan huruf "t". Tujuan pencarian ini adalah untuk mengidentifikasi frasa-frasa yang memiliki pola tertentu dan relevan dengan topik kesehatan dalam teks berita dari BBC NEWS (Health Related).

Dengan menggunakan pola regex ini, penelitian berhasil menemukan frasa-frasa yang memenuhi pola tersebut dalam teks berita. Frasa-frasa ini mungkin berhubungan dengan topik kesehatan yang dibahas dalam teks berita, seperti nama obat atau penyakit, tindakan medis, atau istilah-istilah yang relevan dengan domain kesehatan. Contoh frasa yang dapat ditemukan menggunakan pola regex ini adalah "skin treatment", "surgical techniques", "safety tips", dan sebagainya. Pencarian frasa-frasa dengan pola tertentu ini membantu pengguna dalam mengidentifikasi informasi yang spesifik dan relevan dengan topik kesehatan dalam teks berita. Frasa-frasa tersebut dapat memberikan wawasan tambahan tentang aspek-aspek khusus dalam bidang kesehatan yang sedang dianalisis.

Dengan menggunakan pola regex $^s[a-z] ^t[a-z]$ dan analisis menggunakan AntConc, penelitian ini berhasil mengungkapkannya frasa-frasa yang memiliki pola tertentu dan terkait dengan topik kesehatan dalam teks berita dari BBC NEWS (Health Related). Informasi ini dapat membantu pengguna untuk memperoleh pemahaman yang lebih dalam tentang konteks kesehatan yang dibahas dalam teks tersebut, serta menemukan frasa-frasa yang memiliki relevansi dengan topik kesehatan yang sedang dianalisis.

4. Kesimpulan

Antconc merupakan aplikasi yang sangat berguna dalam menganalisis dan memproses teks atau bahasa. Antconc memungkinkan pengguna untuk mencari kata-kata atau frasa tertentu dalam teks atau data bahasa, dan menampilkan hasil pencarian dalam format konkordansi. AntConc juga dapat menghitung frekuensi kata dalam teks atau korpus, dan menampilkan hasil dalam bentuk grafik atau tabel. Untuk mengidentifikasi pola-pola dan makna yang terkait dengan kata-kata tertentu dapat menggunakan analisis kolokasi.

Aplikasi Antconc dapat digunakan untuk mengidentifikasi Regular Expression (Regex) berdasarkan data berupa teks. Hasil penelitian menunjukkan bahwa Regex pada Antconc dapat dilakukan dengan berbagai cara, antara lain: (1) Untuk menemukan kata yang diawali dengan ‘sp’ dapat dilakukan dengan rumus $^{\wedge}\text{sp}[\text{a-z}]$, (2) untuk menemukan kata yang diawali dengan “st” dan berakhiran “ed” dapat dilakukan dengan rumus $^{\wedge}\text{st}[\text{a-z}]*\text{ed}$, (3) Untuk menemukan kata yang berakhiran “-ing” dapat dilakukan dengan rumus $[\text{a-z}]*\text{ing}$, (4) Untuk menemukan Sebuah Frasa yang mengandung dua kata dengan kata pertama diawali dengan huruf “s” lalu dengan kata kedua diawali dengan huruf “t” dapat dilakukan dengan rumus $^{\wedge}\text{s}[\text{a-z}]^{\wedge}\text{t}[\text{a-z}]$.

References

- Anthony, L. (2004). AntConc: A learner and classroom friendly, multi-platform corpus analysis toolkit. *proceedings of IWLeL*, 7-13.
- Azmovich, T. V. (2019). Analysing Some Uzbek Texts Via Corpus Analysis Toolkit-“Antconc”. *Think India Journal*, 22(4), 4690-4700.
- Biber D, Conrad S, Reppen R (1998) *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Cheng W (2011) *Exploring Corpus Linguistics: Language in Action*. New York: Routledge
- Doumit, S., & Minai, A. (2012, April). Topic identification and analysis in large news corpora. In *Midwest Artificial Intelligence and Cognitive Science Conference* (p. 31).
- Faizah , Mohamad Nusri (2018) Features of lexical verbs in the discussion section of masters' dissertations / Faizah Mohamad Nusri. Masters thesis, University of Malaya..
- Gvishiani, N. B. (2008). *Praktikum po korpusnoy lingvistike [Workshop on corpus linguistics]*. Moskva: Vysshaya shkola. <https://www.twirpx.com/file/2341987/> [in Russian].
- Holmes, J & Meyerhoff, M. 2003. *The Handbook of Language & Gender*. USA: Blackwell Publishing Ltd
- Johanson S., Stenstrom A. 1991. *English Computer Corpora: Selected Papers and Research Guide*. Walter de Gruyter, pp.127-148
- Khairas, E. E. (2019). Using Antconc software as English learning media: The students’ perception. *EPIGRAM (e-journal)*, 16(2), 189-194.
- Kwary, D. A. dan Arum, K. W. A. 2011. 'Lincoln's vs. Obama's presidencies: A diachronic corpus based analysis of the adjectival collocates of [man] and [woman] in the American English'. *ReVEL* 9 (17): 211-225
- L, Hans. 2009. *Corpus Linguistics and the Description of English*. Edinburgh University Press p.1-8.
- Pascarina, H. (2018). Disphemisms and their translations in bbc online news texts (disphemisme dan terjemahannya pada teks berita bbc online). *Leksema: Jurnal Bahasa dan Sastra*, 3(1), 1-10.
- Rubbani, A., Akram, I., & Raza, A. 2014. Syntactic Features of Simile in English Learners’ Corpora.
- Sun, Y. and Wang, L. (2003). Concordancers in the EFL classroom: cognitive approaches and collocation difficulty. *Computer Assisted Language Learning* 16(1)