

DEVELOPMENT OF A WEB-BASED TONSEA LANGUAGE CORPUS USING THE EVOLUTIONARY PROTOTYPING METHOD

Sherly M. Congressco^{*1}, Vivi P. Rantung², Quido C. Kainde³

^{1,2,3}Informatics, Engineering Faculty, Universitas Negeri Manado, Indonesia
Email: ¹19210031@unima.ac.id, ²vivirantung@unima.ac.id, ³quidokainde@unima.ac.id

(Article received: May 7, 2024; Revision: May 31, 2024; published: August 21, 2024)

Abstract

Without realising it, when humans increasingly follow technological advances and are only concerned with the demands of the times where technology can coexist with humans, humans sometimes forget to preserve their culture, one of which is local language. One of them is Tonsea language. Tonsea language is a regional language originating from North Sulawesi. The purpose of this study is to research material for Tonsea language corpus linguistics in the preparation of dictionaries and add Tonsea language resources to preserve the Tonsea language by developing a website to analyse the Tonsea language corpus. Corpus analysis can also be used to research or study variations in the use of the Tonsea language because the corpus can help linguists and lexicographers in the preparation of dictionaries in working on dictionary microstructures which include lemmas/sublemmas, word classes, definitions and writing. As a result, there are six key concepts of corpus analysis techniques, namely tokens, word frequency, concordance, collocation, ngrams, and word lists. In the Token feature, the Token can be used by pekamus to create a dictionary and linguists can also analyse the Tonsea language on Ankorsea and for the Concordance, Collocation and Ngram features users can search for keywords to find out the meaning of the use of a language. This research uses the Evolutionary Prototyping method.

Keywords: *Corpus, Tonsea Language.*

PENGEMBANGAN KORPUS BAHASA TONSEA BERBASIS WEB MENGGUNAKAN METODE EVOLUTIONARY PROTOTYPING

Abstrak

Tanpa disadari ketika manusia semakin mengikuti kemajuan teknologi dan hanya mementingkan tuntutan zaman yang dimana teknologi dapat berdampingan dengan manusia tetapi manusia terkadang lupa untuk tetap melestarikan kebudayaannya salah satunya adalah bahasa daerah. Salah satunya adalah bahasa Tonsea. Bahasa Tonsea merupakan bahasa daerah yang berasal dari Sulawesi Utara. Tujuan penelitian ini adalah untuk bahan penelitian Lingustik Korpus Bahasa Tonsea dalam penyusunan kamus dan menambah sumber daya bahasa Tonsea untuk melestarikan bahasa Tonsea dengan mengembangkan website untuk menganalisis korpus bahasa Tonsea. Analisis Korpus juga dapat digunakan untuk meneliti atau mengkaji variasi dari penggunaan bahasa tonsea karena dengan adanya korpus dapat membantu para linguistik dan leksikograf dalam penyusunan kamus dalam mengerjakan mikrostruktur kamus yang meliputi lema/sublema, kelas kata, definisi dan penulisan. Hasilnya, terdapat enam Konsep kunci teknik analisis korpus yaitu token, frekuensi kata, konkordansi, kolokasi, ngram, dan daftar kata. Pada fitur Token, dari Token tersebut dapat dimanfaatkan oleh pekamus untuk menciptakan kamus dan para linguistik juga dapat menganalisis bahasa Tonsea pada Ankorsea dan untuk fitur Konkordansi, Kolokasi dan Ngram pengguna dapat mencari kata kunci untuk mengetahui makna dari penggunaan suatu bahasa. Penelitian ini menggunakan metode Evolutionary Prototyping.

Kata kunci: *Bahasa Tonsea, Korpus.*

1. PENDAHULUAN

Seiring berjalannya perkembangan teknologi dan informasi pada saat ini selalu berinovasi dengan kreatif secara terus-menerus, sehingga dengan adanya akses internet dan kemajuan teknologi sangat memudahkan manusia dalam berkomunikasi dan

membantu manusia dalam pekerjaan yang biasanya dilakukan. Namun, tanpa disadari ketika manusia semakin mengikuti kemajuan teknologi dan hanya mementingkan tuntutan zaman yang dimana teknologi dapat berdampingan dengan manusia tetapi manusia terkadang lupa untuk tetap melestarikan

kebudayaannya salah satunya adalah bahasa daerah [1].

Bahasa daerah merupakan kebudayaan identitas bangsa Indonesia yang cenderung kurang disadari oleh masyarakat, khususnya generasi muda yang ada di daerah pengguna bahasa tersebut. Salah satunya adalah bahasa Tonsea [1]. Bahasa Tonsea merupakan bahasa daerah yang berasal dari Sulawesi Utara [2].

Pada tahun 2021, Profesor Gufran A. Ibrahim melakukan survei mengenai kebiasaan penggunaan bahasa daerah di lingkungan rumah pada keluarga penutur jati beberapa bahasa daerah di Sulawesi Utara dengan memberikan 6 pertanyaan pokok [3]. Responden pada survei ini adalah ibu yang berusia 50-an tahun. Mereka merupakan penutur asli bahasa Tonsea, Tondano, Tontemboan, Ratahan, dan Sangihe. Hasil dari survei tersebut menunjukkan bahwa para ibu sudah sangat jarang menggunakan bahasa ibu atau dialek bahasa daerah dan dalam berkomunikasi setiap hari kepada anak-anak, mereka lebih memilih menggunakan bahasa melayu-manado [3]. Hal ini mengakibatkan bahasa daerah lambat laun mulai ditinggalkan dan juga kurang minatnya generasi muda sebagai generasi penerus terhadap penggunaan bahasa Tonsea dan pengabaian bahasa ibu yang justru dilakukan oleh ibu, yang seharusnya menjadi orang pertama memperkenalkan bahasa daerah supaya tidak terjadi putusnya warisan bahasa daerah antargenerasi.[3].

Tujuan penelitian ini adalah untuk bahan penelitian Linguistik Korpus Bahasa Tonsea dalam penyusunan kamus dan menambah sumber daya bahasa Tonsea untuk melestarikan bahasa Tonsea. Oleh karena itu, upaya untuk melestarikan bahasa Tonsea pada generasi muda saat ini dengan memanfaatkan penggunaan teknologi adalah penyusunan kamus digital. Dalam pembuatan kamus, bukan hanya asal disusun saja, melainkan mempunyai tujuan dari setiap penggunaan kata [4]. Sebelum kamus diciptakan, data bahasa alamiah dalam jumlah besar dan di olah secara digital untuk tujuan penyelidikan linguistik dan mengetahui penggunaan suatu bahasa yang disebut dengan korpus atau *machine-readable text corpora* [5].

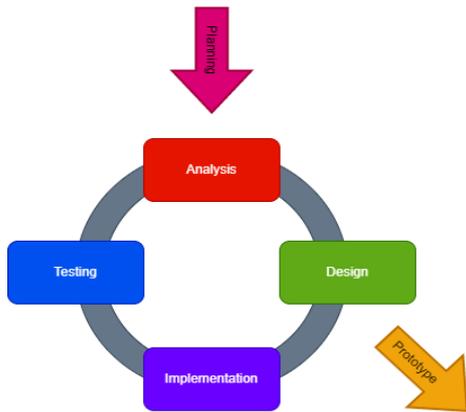
Di Indonesia, sudah tersedia 3 korpus yaitu Korpus Indonesia (KOIN), Korpus Nusantara (KORTARA), Korpus Universitas Indonesia yang memiliki 4 pilihan bahasa yaitu korpus bahasa Indonesia, korpus bahasa Jawa, korpus bahasa Minang dan korpus bahasa Sunda [6]. Akan tetapi, para peneliti korpus terdahulu belum ada yang mengembangkan analisis korpus bahasa Tonsea dan pengembangan korpus dalam bidang linguistik, pengolahan bahasa alami, dan tentang informasi terkait masih sangat minim. Analisis korpus bahasa secara digital ini sangat diperlukan agar mudah diakses oleh seluruh peneliti bahasa yang nantinya dapat dimanfaatkan sebagai sarana dan media untuk penelitian linguistik, linguistik forensik dan penerjemahan.

Korpus digital ini berupa platform yang dapat memudahkan para peneliti bahasa dalam menganalisis korpus bahasa Tonsea sesuai dengan enam konsep kunci teknik analisis korpus yaitu yang pertama, *Token* merupakan jumlah keseluruhan kata yang muncul dalam korpus yang bersifat unik [6], yang kedua, *Frekuensi kata* merupakan daftar kemunculan kata dari yang paling sering digunakan dan juga sampai yang jarang [7], yang ketiga, *Konkordansi* merupakan daftar kemunculan kata untuk mengetahui konteks makna dengan cara membandingkan satu baris konteks kalimat dengan baris lainnya dalam jumlah yang banyak [8], yang keempat, *Kolokasi* merupakan kata yang paling sering muncul bersama-sama atau berdampingan diantara kata lainnya [9], yang kelima, *N-gram* merupakan kata yang muncul berurutan [10], dan yang terakhir *Daftar kata* merupakan keseluruhan kata yang terdapat dalam suatu korpus disertai dengan frekuensi [6]. Manfaat penggunaan korpus untuk menganalisis bahasa yaitu dapat memudahkan para ahli linguistik untuk memperoleh data sumber daya bahasa (SDB) daerah yaitu bahasa Tonsea pada sekumpulan teks digital yang nantinya akan digunakan untuk penyusunan kamus, leksikografi, pengajaran bahasa, analisis sentimen, dan penerjemah mesin translasi.

Setelah peneliti menganalisis masalah diatas, maka peneliti menemukan solusi dengan memanfaatkan media sosial, artikel-artikel, dan juga mewawancarai penutur asli sebagai narasumber untuk membuat Pengembangan Korpus Bahasa Tonsea Berbasis Web menggunakan Metode *Evolutionary Prototyping*. Penggunaan metode *Evolutionary Prototyping* diperlukan karena pada saat mengimplementasikan metode kepada pengguna, pengguna dapat mengetahui lebih awal kelebihan dan kekurangan dari fitur dan fungsi pada website dari tahapan awal yaitu analisis kebutuhan pengguna serta desain awal yaitu pembuatan UML dan adanya kerjasama yang baik antar pengguna untuk mengetahui apa saja yang dibutuhkan oleh pengguna dengan demikian menghasilkan sistem sesuai dengan jadwal waktu penyelesaian yang telah ditentukan [11].

2. METODE PENELITIAN

Metode *prototype* memiliki 3 pendekatan yaitu *Throw Away*, *Incremental* dan *Evolutionary Prototyping* [12]. Dari ketiga pendekatan tersebut peneliti menggunakan pendekatan *Evolutionary Prototyping*. *Evolutionary Prototyping* merupakan tahapan dimana sistem dibangun dengan cara yang terstruktur dan terus menerus diperbaikinya, dengan cara menentukan kebutuhan user sampai menjadi sistem akhir [13].



Gambar 1 Tahapan Metode penelitian *Evolutionary Prototyping* [13]

Berikut ini adalah tahapan pengembangan sistem Metode Prototype dengan pendekatan *Evolutionary Prototyping*:

a. Analisis

Pada tahapan analisis ini, menggunakan teknik pengumpulan data untuk mengumpulkan data bahasa tonsea yaitu dengan mewawancarai penutur asli bahasa tonsea dengan cara peneliti melakukan sesi tanya jawab langsung kepada penutur asli yaitu orang tua, masyarakat dan sanak saudara yang berasal dari Desa Laikit, Minahasa Utara. Dan juga memanfaatkan studi literatur mengenai pengembangan korpus. Berdasarkan hasil analisis tersebut, memberikan solusi bahwa untuk penggunaan teknik analisis pada korpus menggunakan Library NLTK, *NLTK* merupakan library yang digunakan untuk mengolah dan memproses data bahasa manusia. *Library NLTK* memiliki berbagai macam text processing mulai dari klasifikasi, tokenisasi, konkordansi, ngram, kolokasi dan lain sebagainya [14].

b. Desain

Pada tahapan ini merupakan perancangan sistem berfokus pada konsep berorientasi objek dengan menggunakan *Unifed Modeling Language (UML)* seperti *Use Case Diagram*, *Use Case Scenario*, *Activity Diagram*, *Sequence Diagram* dan *Class Diagram*.

c. Implementasi

Pada tahapan ini dilakukan proses pengkodean atau pembuatan sebuah sistem.

d. Pengujian

Pada tahapan ini dilakukan pengujian pada sistem yaitu menggunakan *Blackbox Testing*.

3. HASIL DAN PEMBAHASAN

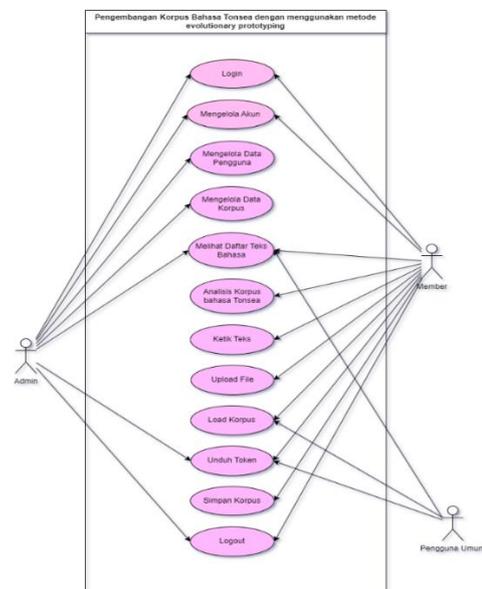
a. Analisis

a) Analisis Kebutuhan Pengguna

1. Terdapat 3 pemakai sistem yaitu pengguna umum, member dan admin.
2. Pengguna umum dapat mengunduh Token teks bahasa Tonsea, melihat daftar kata bahasa tonsea, memuat ulang teks bahasa

3. Jika pengguna umum memiliki kepentingan untuk menganalisis korpus bahasa Tonsea, pengguna umum harus menghubungi admin terlebih dahulu supaya dapat dibuatkan akun untuk masuk ke dalam website Ankorsea.
4. Member yang memiliki kepentingan untuk menggunakan korpus bahasa tonsea dapat menganalisis korpus, menyimpan daftar kata hasil analisis dan melihat daftar kata bahasa tonsea.
5. Admin bertugas untuk memanajemen pendaftaran pengguna, mengelola data pengguna dan dapat mengelola data teks bahasa Tonsea.
6. Member harus mendaftarkan diri kepada admin terlebih dahulu.
7. Pada website ini, Member dapat mengakses Ankorsea untuk menganalisis korpus dengan cara diketik, diunggah, dan dapat mereset kata yang diketik.
8. Hasil analisis korpus akan mengacu pada 6 fitur yaitu Daftar kata, Token, Frekuensi kata, Konkordansi, Kolokasi, dan Ngram. Member juga dapat mengunduh Token kata bahasa Tonsea.
9. Member juga dapat mencari kata kunci pada fitur Konkordansi, Kolokasi, dan Ngram.

b. Use Case Diagram



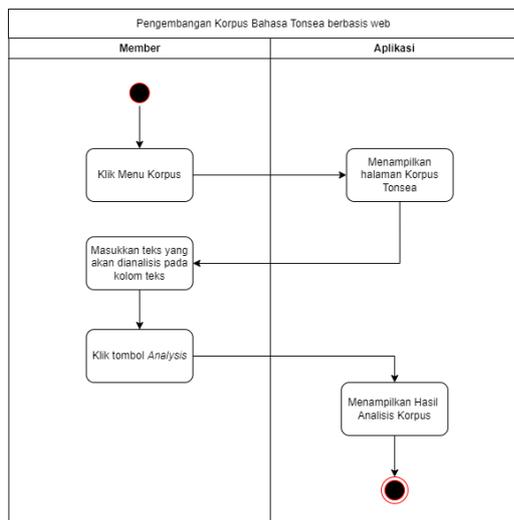
Gambar 2. *Use Case Diagram*

Terdapat 3 aktor yaitu Pengguna Umum, *Member* dan *Admin*. Yang pertama, Pengguna Umum hanya memiliki akses untuk melihat daftar korpus yang sudah pernah dianalisis, memuat ulang daftar hasil analisis, dan dapat mengunduh *Token*. Jika pengguna umum berkeinginan untuk menganalisis teks bahasa Tonsea, pengguna umum harus menghubungi admin terlebih dahulu supaya memiliki

hak untuk mengakses website Ankorsea. Kedua, Member merupakan pengguna yang memiliki akses penuh untuk menganalisis teks bahasa tonsea pada website Ankorsea. Member dapat melakukan *Login* dan mendapatkan alamat email yang telah didaftarkan serta *password* dari *Admin*. Pada analisis Korpus bahasa Tonsea terdapat 2 cara penggunaan yaitu memasukkan teks bahasa tonsea dan mengunggah file .txt. Member juga dapat menyimpan hasil teks bahasa tonsea yang telah dianalisis dan dapat memuat ulang hasil analisis korpus yang telah disimpan serta mengunduh *Token* berupa file .txt. Ketiga, *Admin* memiliki tugas untuk mengelola data pengguna yaitu menambah data pengguna baru, menghapus data pengguna dan memperbaharui data pengguna. *Admin* juga dapat mengelola data korpus yaitu melihat daftar kata bahasa Tonsea dan menghapus data bahasa Tonsea jika terdapat bahasa yang tidak sopan.

c. Activity Diagram

1. Analisis Korpus dengan memasukkan Teks

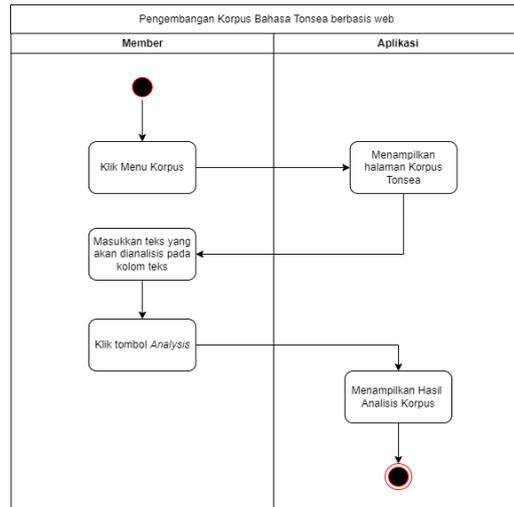


Gambar 3. Activity Diagram Analisis Korpus dengan memasukkan teks

Pada gambar 3 merupakan activity diagram dari analisis korpus dengan memasukkan teks. Pada bagian ini, pengguna yang memiliki fitur ini adalah member. Untuk menganalisis data bahasa tonsea. Ankorsea terdapat 2 cara yaitu mengunggah file dan memasukkan teks bahasa tonsea dengan cara diketik ataupun ditempel.

2. Analisis Korpus dengan mengunggah file

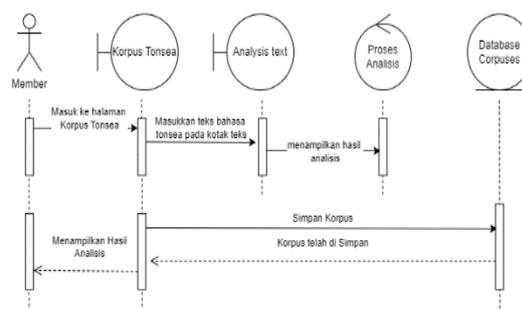
Pada gambar 4 merupakan activity diagram dari analisis korpus dengan mengunggah file. pengguna yang dapat mengakses fitur ini adalah member. Untuk menganalisis data bahasa tonsea, Ankorsea terdapat 2 cara yaitu mengunggah file dan memasukkan teks bahasa tonsea dengan cara diketik ataupun ditempel. Pada aktivitas ini, hanya member yang dapat mengunggah file teks bahasa dengan menggunakan tipe file .txt.



Gambar 4. Activity Diagram Analisis Korpus dengan mengunggah file

d. Sequence Diagram

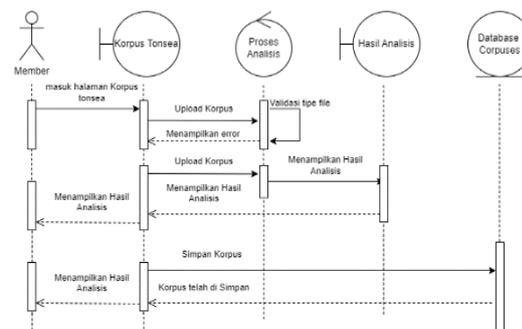
1. Analisis Korpus dengan memasukkan teks



Gambar 5. Sequence diagram analisis korpus dengan memasukkan teks

Pada gambar 5 merupakan proses analisis korpus mempunyai 2 cara yaitu mengetikkan teks bahasa kedalam kolom teks dan dapat mengunggah file. Dimulai dari pengguna membuka masuk ke halaman analisis korpus kemudian mengetik atau menempel teks bahasa yang akan dianalisis kemudian menekan tombol analisis lalu sistem akan menampilkan korpus yang telah dianalisis.

2. Analisis korpus dengan mengunggah file

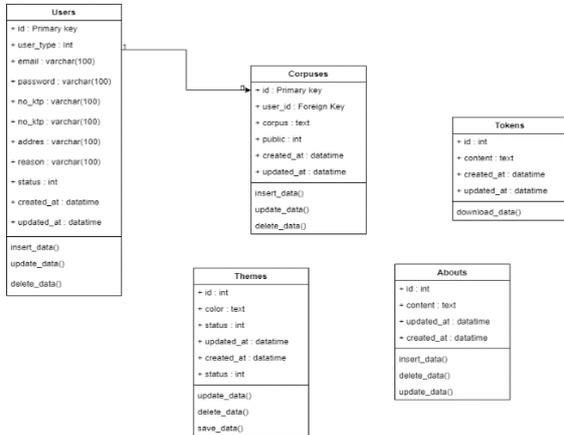


Gambar 6 Sequence diagram analisis korpus dengan mengunggah file

Pada gambar 6, dijelaskan proses analisis korpus mempunyai 2 cara yaitu mengetikkan teks

bahasa kedalam kolom teks dan dapat mengunggah file. Dimulai dari pengguna masuk ke halaman analisis korpus lalu mengunggah file yang akan dianalisis dalam bentuk file .txt setelah itu sistem akan menampilkan korpus yang telah dianalisis.

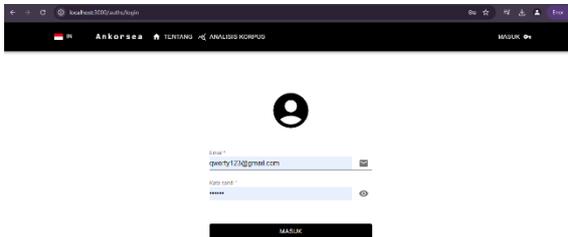
e. Class Diagram



Gambar 7. Class Diagram

Pada gambar 7, merupakan *class diagram* yang terdiri dari 5 kelas yaitu *Corpuses*, *Users*, *Tokens*, *Themes*, dan *Abovts*.

f. Implementasi



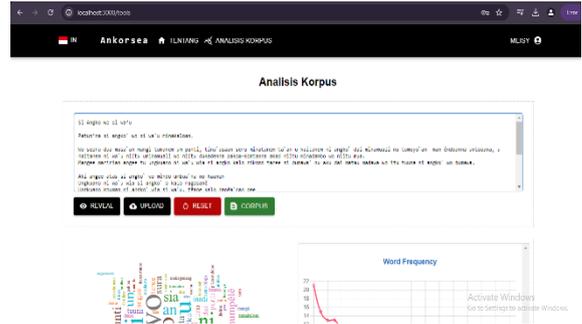
Gambar 8. Halaman Login

Pada gambar 8 merupakan halaman login. Member dan admin memiliki akses untuk login dengan cara memasukkan alamat email dan password. Jika pengguna umum ingin menganalisis korpus harus login terlebih dahulu dan menghubungi admin.



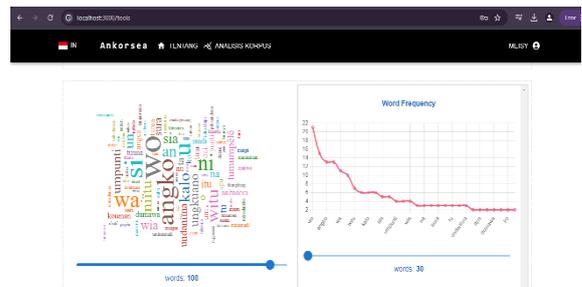
Gambar 9. Halaman Tentang website Ankorsea

Pada gambar 9 merupakan Halaman Tentang didalamnya terdapat informasi mengenai Website Ankorsea.



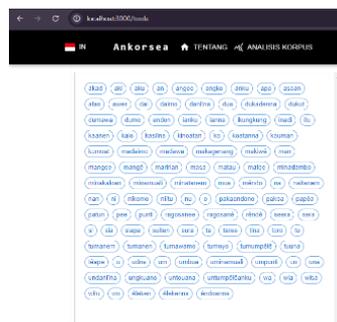
Gambar 10. Halaman Analisis Korpus

Pada gambar 10 merupakan Halaman Analisis Korpus, pada halaman ini pengguna dapat menganalisis korpus dengan 2 cara yaitu mengetik teks yang akan dianalisis dan mengunggah file. Pengguna juga dapat menyimpan korpus yang telah dianalisis dan dapat melihat daftar hasil analisis yang telah disimpan. Pada bagian ini juga terdapat 6 teknik analisis korpus yaitu Frekuensi kata, daftar kata, token, konkordansi, kolokasi dan ngram.



Gambar 11. Analisis Korpus berdasarkan daftar kata dan frekuensi kata

Pada gambar 11 merupakan hasil dari analisis korpus berdasarkan teknik analisis korpus yaitu daftar kata dan frekuensi kata. Daftar kata merupakan keseluruhan kata yang terdapat dalam suatu korpus disertai dengan frekuensi [6]. Pada bagian daftar kata terdapat jumlah keseluruhan kata yaitu 100 kata. Frekuensi kata merupakan daftar kemunculan kata dari yang paling sering digunakan dan juga sampai yang jarang [7]. Pada gambar 17 terdapat frekuensi daftar kata yang paling sering digunakan yaitu *wo* berjumlah 21 kata dan yang paling jarang digunakan yaitu *ko*, berjumlah 2 kata.



Gambar 12. Analisis Korpus berdasarkan token

Pada gambar 12 merupakan hasil dari analisis korpus berdasarkan teknik analisis korpus yaitu token

dan konkordansi. Token merupakan jumlah keseluruhan kata yang muncul dalam korpus yang bersifat unik [6]. Pada token terdapat 106 kata yang unik.

Mencari kata kunci *nyaku* artinya saya pada hasil analisis dari data korpus dengan menggunakan fitur korpus yaitu Konkordansi, Kolokasi dan Ngram.

Konkordansi merupakan daftar kemunculan kata untuk mengetahui konteks makna dengan cara membandingkan satu baris konteks kalimat dengan baris lainnya dalam jumlah yang banyak [8]. Kolokasi adalah kata yang paling sering muncul bersama-sama atau berdampingan diantara kata lainnya [9]. Ngram adalah kata yang muncul berurutan [10].

The screenshot shows a web interface for Konkordansi analysis. At the top, there is a 'Concordances' header. Below it, there are two input fields: 'Width' set to 80 and 'keyword' set to 'nyaku'. Below these fields is a table with columns 'ID', 'Left', 'Term', and 'Right'. The table contains two rows of data:

ID	Left	Term	Right
1	ou nepa ndee een ko wo manto wisa ko	nyaku	tou laiki
2	n uree mo maanto wia batam puuna tii	nyaku	koki kar

Gambar 13. hasil analisis konkordansi berdasarkan kata kunci nyaku

The screenshot shows a web interface for Konkordansi analysis. At the top, there is a 'Concordances' header. Below it, there are two input fields: 'Width' set to 80 and 'keyword' set to 'nyaku'. Below these fields is a table with columns 'Term' and 'Right'. The table contains two rows of data:

Term	Right
manto wisa ko nyaku	tou laikit taan uree mo maanto wia b
batam puuna tii nyaku	koki kampee kinuan ni tantaku sa ko

Gambar 14. hasil analisis konkordansi berdasarkan kata kunci nyaku

Pada gambar 13 dan 14 merupakan kata kunci *nyaku* dari teknik analisis konkordansi. Konkordansi dapat membandingkan kata yang berdampingan dari kata kunci *nyaku* yaitu terdapat kata-kata sebelum *nyaku* dan kata-kata sesudah kata *nyaku*. Pada gambar 13 merupakan kata-kata sebelum *nyaku* pada bagian kiri yang pertama, yaitu *tou nepa ndee'een ko wo manto wisa ko* lalu *nyaku*. Pada gambar 14 merupakan Kata-kata sesudah *nyaku* pada bagian kanan yaitu *tou laikit taan uree mo maanto wia Batam*. Yang kedua, kata-kata sebelum *nyaku* yaitu *tou Laikit taan uree mo maanto wia Batam Puuna tii* lalu *nyaku*. kata-kata sesudah *nyaku* adalah *koki kampee kinuan ni tantaku 'sa ko sela mo'*.

The screenshot shows a web interface for Collocates analysis. At the top, there is a 'Collocates' header. Below it, there is a 'keyword' input field containing 'nyaku'. Below this field is a table with columns 'ID', 'Term', 'Collocate', and 'Count'. The table contains two rows of data:

ID	Term	Collocate	Count
1	nyaku	koki	1
2	nyaku	tou	1

Gambar 15. hasil analisis kolokasi berdasarkan kata kunci nyaku

Pada gambar 15, merupakan kata kunci *nyaku* dari teknik analisis kolokasi. Kolokasi yang paling sering muncul dan berdampingan secara bersamaan. Berdasarkan hasil analisis diatas, term atau subjek merupakan kata *nyaku* dan hasil kata yang bersamaan pada kata *nyaku* adalah *koki*. Jumlah kolokasi dan term atau subjek pada kata *Nyaku* dan *koki* yaitu 1. Lalu yang kedua, term atau subjek dari kata *nyaku* adalah *tou* yang merupakan hasil kata yang bersamaan dengan kata *nyaku*. Jumlah kolokasi dan term atau subjek pada kata *Nyaku* dan *tou* yaitu 1.

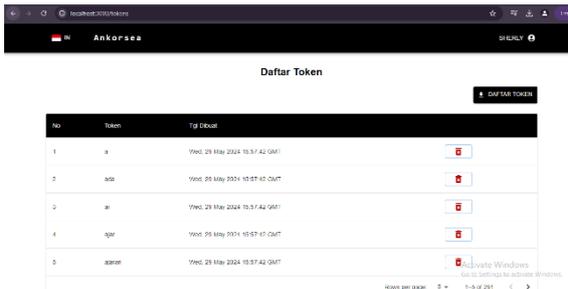
The screenshot shows a web interface for N-Grams analysis. At the top, there is an 'N-Grams' header. Below it, there are two input fields: 'Size' set to 2 and 'keyword' set to 'nyaku'. Below these fields is a table with columns 'ID', 'Type', and 'Frequency'. The table contains two rows of data:

ID	Type	Frequency
1	nyaku,koki	1
2	nyaku,tou	1

Gambar 16. hasil analisis kolokasi berdasarkan kata kunci nyaku

Pada gambar 16, merupakan kata kunci *nyaku* dari teknik analisis Ngram. Ngram merupakan kata yang paling sering muncul secara berurutan. Berdasarkan hasil analisis diatas, setelah kata kunci *nyaku* adalah *koki*. Hasil tersebut dapat dibuktikan dari hasil analisis konkordansi sebelumnya yang merupakan membandingkan kata yang berdampingan seperti berikut ini *nyaku* kata-kata sesudah *nyaku* adalah *koki kampee kinuan ni tantaku 'sa ko sela mo'*. Membuktikan bahwa kata *nyaku* berurutan dengan kata *koki*.

Hasil analisis yang kedua, setelah kata kunci *nyaku* adalah *tou*. Hasil tersebut dapat dibuktikan dari hasil analisis konkordansi sebelumnya yang merupakan membandingkan kata yang berdampingan seperti berikut ini *nyaku tou laikit taan uree mo maanto wia Batam*. Membuktikan bahwa kata *nyaku* berurutan dengan kata *tou*.



Gambar 17. daftar semua Token

Pada gambar 17, merupakan halaman daftar token. Admin, member dan juga pengguna umum memiliki akses untuk mengunduh semua daftar token dari teks bahasa.

g. Pengujian

Pada tahapan ini, peneliti melakukan pengujian fungsionalitas pada aplikasi dengan menggunakan metode *Black Box Testing*.

Berikut pengujian Website Korpus Bahasa Tonsea dengan menggunakan metode *Black Box Testing*:

Tabel 1 *Blackbox Testing*

No	Kasus Uji	Hasil	Aktor	Keterangan
1.	About	Menampilkan Halaman tentang Ankorsea	Member dan Pengguna Umum	Sesuai
2.	Korpus Tonsea	Menampilkan Halaman Analisis Korpus	Member dan Admin	Sesuai
3.	Login	Menampilkan Halaman Login	Member dan Admin	Sesuai
4.	Analisis	Menampilkan 6 fitur pada korpus yaitu daftar kata, frekuensi kata, token, konkordansi, kolokasi dan Ngram.	Member dan Admin	Sesuai
5.	Unggah File	Menampilkan 6 fitur pada korpus yaitu daftar kata, frekuensi kata, token, konkordansi, kolokasi dan Ngram.	Member dan Admin	Sesuai
6.	Reset	Menghapus hasil analisis korpus yang keluar	Member dan Admin	Sesuai
7.	Hasil Analisis Korpus	Menampilkan 6 fitur pada korpus yaitu daftar kata, frekuensi kata, token, konkordansi, kolokasi dan Ngram.	Member dan Admin	Sesuai

8.	Korpus	Menampilkan daftar teks bahasa tonsea yang telah disimpan	Member dan Admin	Sesuai
9.	Profil	Menampilkan data diri pengguna	Member dan Admin	Sesuai
10.	Simpan Korpus	Menyimpan data korpus kedalam database	Member dan Admin	Sesuai
11.	Load Korpus	Memuat kembali korpus yang sudah pernah dianalisis	Member dan Admin	Sesuai
12.	Logout	Menampilkan halaman login awal	Member dan Admin	Sesuai

4. DISKUSI

Penelitian ini sebelumnya sudah pernah dilakukan, di Indonesia terdapat beberapa website mengenai korpus yaitu Korpus Indonesia (KOIN), Korpus Nusantara (KORTARA), Korpus Universitas Indonesia yang memiliki 4 pilihan bahasa yaitu korpus bahasa indonesia, korpus bahasa jawa, korpus bahasa minang dan korpus bahasa sunda [6]. Kemudian referensi dari penelitian ini adalah pengaplikasian AntConc pada korpus bahasa Melayu Ambon, hasil dari penelitian ini adalah untuk mengetahui bagaimana penerapan penggunaan aplikasi AntConc dalam korpus bahasa Melayu Ambon, dan kosakata-kosakata apa saja yang muncul pada korpus yang telah dibuat [15]. Perbedaannya adalah pada penelitian ini menggunakan bahasa Tonsea sedangkan penelitian terdahulu menggunakan bahasa Melayu-Ambon. Persamaannya berupa hasil analisis korpus berdasarkan teknik analisis korpus.

5. KESIMPULAN

Berdasarkan hasil dari penelitian yang dilakukan pada Pengembangan Korpus Bahasa Tonsea berbasis website dengan menggunakan Metode *Evolutionary Prototyping*, maka dari itu penulis dapat memberikan kesimpulan bahwa telah tersedia website pengembangan korpus bahasa tonsea. Dengan adanya Ankorsea atau singkatan dari Analisis Korpus Tonsea, bahasa tonsea dapat terus menerus dikembangkan dan para linguistik korpus yang ingin mencari sumber data secara digital mengenai Bahasa Tonsea untuk kepentingan penyusunan kamus akan lebih mudah dan sistematis untuk di akses karena Ankorsea hadir untuk menjadi jawaban dari para linguistik korpus, Pada Ankorsea para linguistik dapat mengunduh teks bahasa yang sudah dianalisis pada fitur Token, dari Token tersebut dapat dimanfaatkan oleh pekamus untuk menciptakan kamus dan para linguistik juga dapat menganalisis bahasa Tonsea pada Ankorsea yang memiliki 6 fitur teknik analisis korpus yaitu Daftar kata, Frekuensi kata, Token, Konkordansi, Kolokasi dan Ngram.

Pada fitur Konkordansi, Kolokasi dan Ngram para linguistik dapat mencari kata kunci dari fitur-fitur tersebut agar supaya dapat mengetahui makna dari penggunaan suatu bahasa.

DAFTAR PUSTAKA

- [1] S. D. Ratumanan *et al.*, “Upaya Pembedayaan Penggunaan Bahasa Daerah Melalui Budaya Literasi Digital,” *J. Elem. Educ.*, vol. 05, no. 01, pp. 69–76, 2022.
- [2] A. J. Senduk, “Profil Pengajaran Bahasa Tonsea pada Peserta Didik di Sekolah Dasar di Kecamatan Kauditan di Kabupaten Minahasa Utara: Suatu Survey,” *LPPM Bid. EkoSosBudKum*, vol. 3, no. 1, pp. 21–46, 2016.
- [3] A. G. Ibrahim, “Membunuh Bahasa Sendiri,” *kompas.id*. [Online]. Available: https://www.kompas.id/baca/artikel-opini/2022/02/22/membunuh-bahasa-sendiri?status=sukses_login&status_login=login&loc=hard_paywall
- [4] W. A. Dewandono, “Leksikologi Dan Leksikografi Dalam Pembuatan Dan Pemaknaan Kamus,” *Paramasastra*, vol. 7, no. 1, p. 16, 2020, doi: 10.26740/paramasastra.v7n1.p16.
- [5] Priyono, “Prospek Penggunaan Korpus untuk Studi Kebahasaan dan Proses Pembelajaran Bahasa Kedua,” *Jurnal Ilmu Pendidikan Universitas Negeri Malang*, vol. 6, no. 2, pp. 75–88, 1999. [Online]. Available: <http://journal.um.ac.id/index.php/jip/article/view/703>
- [6] R. Almos, P. Pramono, S. Seswita, R. A. Asma, and N. O. Putri, “Linguistik Korpus: Sarana dan Media Pembelajaran pada Mata Kuliah Leksikologi dan Leksikografi di Perguruan Tinggi,” *Lect. J. Pendidik.*, vol. 14, no. 1, pp. 45–59, 2023, doi: 10.31849/lectura.v14i1.11705.
- [7] N. W. Sri Arini, I. B. Putu Widja, and I. K. R. Yasa Negara, “Analisis Frekuensi Kata untuk Mengekstrak Kata Kunci dari Artikel Ilmiah Berbahasa Indonesia,” *Eksplora Inform.*, vol. 8, no. 2, pp. 80–84, 2019, doi: 10.30864/eksplora.v8i2.162.
- [8] Bloomfield Maurice, *A Vedic Concordance*. Harvard university, Cambridge, 1906.
- [9] D. R. M. S. I. ERIYANTO, *Analisis Wacana Kritis Berbasis Korpus*. Remaja Rosdakarya, 2022. [Online]. Available: <https://books.google.co.id/books?id=jE0N0AEACAAJ>
- [10] A. Z. Broder, S. C. Glassman, M. S. Manasse, and G. Zweig, “Syntactic clustering of the Web,” *Comput. Networks ISDN Syst.*, vol. 29, no. 8, pp. 1157–1166, 1997, doi: [https://doi.org/10.1016/S0169-7552\(97\)00031-7](https://doi.org/10.1016/S0169-7552(97)00031-7).
- [11] O. Irnawati, I. Darwati, O. Irnawati, and I. Darwati, “Evolutionary Prototype Dalam Perancangan Sistem,” *JUSIM (Jurnal Sist. Inf. Musirawas)*, vol. 6, no. 1, pp. 1–8, 2021.
- [12] A. Z. dan D. Yusri, “Landasan teori Evolutionary Prototyping,” *J. Ilmu Pendidik.*, vol. 7, no. 2, pp. 809–820, 2020.
- [13] J. Crinnion, *Evolutionary systems development : a practical guide to the use of prototyping within a structured systems methodology*. New York SE - 369 Seiten ; 26 cm : Illustrationen.: Plenum Press, 1991. doi: LK - <https://worldcat.org/title/636467974>.
- [14] E. J. Rifano, A. C. Fauzan, A. Makhi, E. Nadya, Z. Nasikin, and F. N. Putra, “Text Summarization Menggunakan Library Natural Language Toolkit (NLTK) Berbasis Pemrograman Python,” *Ilk. J. Comput. Sci. Appl. Informatics*, vol. 2, no. 1, pp. 8–17, 2020, doi: 10.28926/ilkomnika.v2i1.32.
- [15] N. H. Hasan, “PENGAPLIKASIAN ANTCONC PADA KORPUS BAHASA MELAYU AMBON (The Application of AntConc on Ambon Malay Language Corpus),” *Kandai*, vol. 17, no. 2, p. 177, 2021, doi: 10.26499/jk.v17i2.2605.
- [16] J. Eska, M. F. Larasati, P. Studi, and S. Informasi, “Penggunaan K-Means Clustering Untuk Mengelompokkan Kemampuan Bahasa Inggris Siswa Lembaga Kursus Jason English Course,” *J. Tek. Inform.*, vol. 3, no. 3, 2022.