

ஆய்வுக்குறிப்பு எண் - தொ. மா. 119

தொல்காப்பிய மெய்ம்மயக்கப் பைத்தான் நிரலாக்க மேம்பாடு

(Tholkaappiyam Melmayakkam Python Programming Development)

சத்தியராஜ் தங்கச்சாமி, தமிழ் உதவிப்பேராசிரியர் & விக்கிமீடியர்,
ஸ்ரீ கிருஷ்ணா ஆதித்யா கலை மற்றும் அறிவியல் கல்லூரி, கோயம்புத்தூர் - 641042.

Sathiyaraj Thangasamy, Assistant Professor in Tamil & Wikimediar,
Sri Krishna Adithya College of Arts and Science, Coimbatore -641042.
sathiyarajkkt@gmail.com, ORCID: <https://orcid.org/0000-0002-6754-9095>

பரமேஷ்வர் அருணாசலம், மென்பொறியாளர், காஞ்சிபுரம்
Parameshwar Arunachalam, Software Engineer, Kanchipuram parameshwar273@gmail.com

த. சீனிவாசன், நிறுவநர், கணியம் அறக்கட்டளை, கனடா.
T. Shrinivasan, Founder, Kaniyam Foundation, Canada. tshrinivasan@gmail.com

இரா. திலீப், மென்பொறியாளர், சென்னை.
R. Dhilip, Software Engineer, Chennai. dhilidhiva@gmail.com

முன்னுரை

தமிழில் கிடைக்கப்பெற்ற முதல் இலக்கணமாகிய தொல்காப்பியம் கி.மு.14இல் தொல்காப்பியரால் எழுதப்பெற்றது [23] [24]. இவரின் இந்த இலக்கணமும் மொழியியலும் இன்றைய தமிழ்க் கணிமைக்கு அடிப்படை என்பார் ந.தெய்வசுந்தரம் [29]. அதுபோல் தொல்காப்பியம், எழுத்ததிகாரம், சொல்லதிகாரம், பொருளதிகாரம் எனும் மூன்று அதிகாரங்களில் தமிழின் மொழியமைப்பை விளக்கியுள்ளது என்பார் க. பாலசுப்பிரமணியன் [23] [24]. முதலாவது அதிகாரமாகிய எழுத்ததிகாரம் ஒன்பது இயல்களில் தமிழ் மொழியின் எழுத்துக்களின் அறிமுகம், அவ்வெழுத்துக்கள் சொல்லாகும் முறை, சொல்லிற்குள் இருக்கக்கூடிய எழுத்தமைப்பு முறைகள், அந்தச் சொற்கள் புணரும் தன்மைகள் ஆகியவற்றையும், இரண்டாவது அதிகாரமாகிய சொல்லதிகாரம் ஒன்பது இயல்களுள் சொல்வகைகள், தொடரமைப்புகள் ஆகியவற்றையும், மூன்றாவது அதிகாரமாகிய பொருளதிகாரம் ஒன்பது இயல்களில் தமிழ் மொழியின் பொருண்மைகளையும் விளக்கியுள்ளன. அதுமட்டுமின்றி, இந்த மூன்று அதிகாரங்களையும் தொல்காப்பியர் ஒருவரே எழுதினார் என்பதையும் தம் ஆய்வில் விளக்கியுள்ளார் க. பாலசுப்பிரமணியன் [23] [24]. இந்த இலக்கணம் இன்றைய கணினி மொழித் தொழில்நுட்பத்திற்கு எழுதுவதுபோல் அமைந்திருப்பதை உணரமுடிகின்றது. அந்த அளவிற்கு நுட்பமாகத் தொல்காப்பியர் தொகை, வகை, விரி அடிப்படையில் விளக்கியுள்ளார். அதனைத் தொல்காப்பியம் - நூன்பமரபு (பைத்தான்வழி உரை) [30] [31] எனும் கட்டுரையின்வழி விளங்கிக் கொள்ளலாம்.

தொல்காப்பியர் எழுதிய கருத்தியல்கள் இன்றைய பைத்தான் மொழியில் முழுமையாக விதியாக்கங்களாக உருவாக்கப்படவில்லை. ஆகையால் மொழித் தொழில்நுட்ப அடிப்படையில் தொல்காப்பியத்திற்குப் பைத்தான் நிரல் உருவாக்குவது காலத்தின் தேவை என்பதை உணர்ந்து சில முயற்சிகள் மேற்கொள்ளப்பெற்றுள்ளன. அந்த முயற்சி தொழில்நுட்பப் பொறியியல் வல்லுநர்களுடன் இணைந்து மெய்ம்மயக்கம், மொழிமுதல், மொழியிறுதி ஆகிய கருத்தியல்களுக்குப் பைத்தான் நிரல் உருவாக்கப்பெற்றுப் பொது உரிமத்தில் வெளியிடவும் பெற்றது [25].

இவ்வாறு பைத்தானில் நிரலாக எழுதுவதால் என்ன நடந்துவிடப் போகின்றது எனத் தோன்றலாம். இங்கு ந.தெய்வசுந்தரத்தின் "தொல்காப்பியம், நன்னூல் போன்ற இலக்கண நூல்களும் இன்றைய மொழியியலாளர் ஆய்வுகளும் மனித மூளைக்காக உருவாக்கப்பட்டவையாகும். அவற்றை அப்படியே கணினிக்குக் கொடுத்தால், கணினியால் புரிந்துகொள்ள முடியாது. கணினியில் அமைந்துள்ள மின்னணுச் சில்லுகளுக்குப் புரியக்கூடிய கணினி நிரல்களாக (program) அவை மாற்றப்பட வேண்டும்" [1] எனும் கருத்துச் சுட்டிக்காட்டத் தக்கதாய் அமைந்துள்ளது.

இதனை உள்வாங்கியே நூன்மரபு, மொழிமரபு ஆகிய இரண்டு இயல்களில் தொல்காப்பியர் விளக்கியுள்ள மெய்ம்மயக்கக் கருத்தியல்களுக்கு உருவாக்கப்பெற்ற பைத்தான் நிரலாக்கத்தைக் குறிப்பிட்டுக் கூறலாம். இந்த மெய்ம்மயக்கக் கருத்தியலை விளக்குவதற்குத் தொல்காப்பியர் மொத்தம் 82 விதிகளுள் 12 விதிகள் அமைத்துக்கொண்டார். அவை, மெய்ம்மயக்கங்களைக் கூறும் விதிகளாக அமைந்திருந்தன. அந்த விதிகளுள் உள்ள கருத்தியல்களைத் தொகுத்து நோக்கும்பொழுது ஒன்பது வகை மெய்ம்மயக்க விதிகள் கிடைக்கப் பெறுகின்றன. அவை,

- "மெய்ம்மயக்கம்1": "ட்றல்ள்+கசப"
- "மெய்ம்மயக்கம்2": "ல்ள்+யவ"
- "மெய்ம்மயக்கம்3": "ங்ஞண்நம்ன்+இனவொலி(கசடதபற)"
- "மெய்ம்மயக்கம்4": "ண்ன்+கசஞபமயவ"
- "மெய்ம்மயக்கம்5": "ஞ்நம்வ்+ய"
- "மெய்ம்மயக்கம்6": "ம்+வ"
- "மெய்ம்மயக்கம்7": "ய்ர்த்+க ச த ப ஞ ந ம ய வ ங"
- "மெய்ம்மயக்கம்8": "ர்ழ் தவிர -> க்...ன் + க...ன"
- "மெய்ம்மயக்கம்9": "ர்ழ் குற்றொற்றாகா"

என்பன. இவை கணினிக்குப் புரியக்கூடிய வகையில் உருவாக்கப்பெற்ற நிரலாக்கச் சருக்கம் ஆகும். இவ்விதிகளை மொழித் தொழில்நுட்ப அடிப்படையில் ஆய்ந்து பார்க்க வேண்டும். அப்பொழுதுதான் தமிழ்மொழிச் சொற்களைக் கண்டறியும் விதிமுறைகளில் விடுபட்ட விதிகளை அறிய முடியும். ஏனென்றால் தமிழண்ணல் [26] இந்த விதிகள் தமிழ் மொழியின் சொற்களை இனங்காட்டக் கூடியவை என்கிறார். ஆனால், தற்சமம்,

தற்பவம் ஆகிய சொற்களுடனும் இவ்விதி பொருந்திப் போகின்றது. இருப்பினும் தற்போதைய கருதுகோள் தமிழண்ணல் கருத்துப்படியே இருக்கட்டும்.

ஆகையால், தமிழ்மொழிச் சொற்களின் அமைப்பில் உள்ள உள்ளடக்குமுறையைக் குறிக்கும் விதிகளாக இவை அமைகின்றன. அந்த விதிகளின்படி தமிழ்மொழியின் அமைப்புமுறையை எளிதில் கற்றுக்கொள்ள முடியும். அவ்வாறு கற்கும்பொழுது பிழையான சொற்களைத் தவிர்க்க இயலும். அதற்காகவே பைத்தான் மொழியில் தொல்காப்பியர் கூறிய **மெய்ம்மயக்கம், மொழிமுதல், மொழியிறுதி** ஆகியவற்றிற்கு எழுதப்பெற்ற பைத்தான் நிரலாக்கத்தை கித்துப்பில் [25] காணலாம். இது தொல்காப்பியக் குறுஞ்செயலி உருவாகுவதற்கும் [27] அடிப்படையாக அமைந்தது. அதுமட்டுமின்றி மொழித் தொழில்நுட்ப அடிப்படையில் ஆய்வுகள் மேற்கொள்வதற்கும் வழியை இது ஏற்படுத்தும் எனலாம்.

தூண்டுகோள்கள்

இந்த மாதிரியான எண்ணம் உருவாவதற்கு அடிப்படையாக சேக்கபியர் படைப்பின்மீதான ஆய்வுகளே காரணமாக அமைந்தன. தொல்காப்பியத்திற்குப் பைத்தான் நிரலாக்க (program based) அடிப்படையிலோ, இயந்திரவழிக் கற்றல் அடிப்படையிலோ (Machine Learning based) ஆய்வுகள் நிகழ்ந்துள்ளனவா எனக் கூகுள் தேடுபொறியில் (Google search engine) தேடிப் பார்த்தால் இல்லை என்ற நிலையே உள்ளது. சேக்கபியர் சார்ந்த படைப்புமீதான ஆய்வினைத் தேடிப்பார்த்தால் மிகுதியாக உள்ளன. அவற்றை shakespeare-dataset [12], shakespeare-dataset: complete works, plays, sonnets and poems of shakespeare [13], learning-hadoop-and-spark /shakespeare.raw [14], TensorFlow/Week3_Programming [15], machine-learning-experimentstext_generation_shakespeare_rnn [16], datasets/shakespeare [17], Shakespeare-database [18], shakespeare-PyPI [19], py-shakespeare-PyPI [20], Shakespearelang [21], Python NLTK Shakespeare corpus [22] எனவரும் ஆய்வுகள்வழி அறியலாம். இவை சான்றிற்காக மட்டுமே. ஆனால் இன்னும் பல ஆய்வுகள் உள்ளன. இது தொல்காப்பிய ஆய்வையும் இதுபோல் செய்தல் வேண்டும் எனத் தூண்டிவிடுவதாய் அமைகின்றது.

முன்னாய்வுகள்

கித்துப்பில் (GitHub) பதிவேற்றப்பெற்ற தொல்காப்பியவிதி நிரலாக்கத்தை [27] அடிப்படையாக வைத்துச் சில ஆய்வுகள் மேற்கொள்ளப்பெற்றன. அவை வருமாறு; Automatic Identification of Meimayakkam in Tamil Words Using Rule Based and Transfer Learning Approaches [5]. A development of the Meimayakkam second Rule Based on Tholkaappiyam and Nannul grammar concepts [6], DATA SCIENCE BASED CORPUS CREATION FOR THOLKAAPPPIYAM [7], App Development for Tholkaappiya Meymayakkam First rule

problem solve [8].

இதன் தொடர்ச்சியாக இந்த ஆய்வில் தொல்காப்பிய மெய்ம்மயக்கப் பைத்தான் நிரலாக்கத்தை இன்னும் செம்மைப்படுத்தும் முகமாக ஒரு இலக்கத்திற்கும் மேற்பட்ட சொல்வங்கியிலும் இந்த ஆய்விற்காகச் சிறுசிறு சொல்வங்கித் தரவுகளிலும் பயிற்சிக்காக உருவாக்கப்பெற்ற தரவுகளில் இந்த நிரலாக்கத்தை ஆய்விற்கு உட்படுத்தும்பொழுது ஏற்படும் நிரலாக்கச் சிக்கல்களைக் கண்டறிவது முதன்மை நோக்கமாக இவ்வாய்வு அமைகின்றது. அதனுடன் இந்த நிரலாக்கத்தை மேம்படுத்த இயலுமா எனவும், தமிழ் மொழியைக் காட்சிப்படுத்தம் செய்வதில் பைத்தான் அகராதிகளைப் பயன்படுத்த இயலுமா எனவும் கருதுகோள்களாக அமைத்துக் கொள்கின்றது இவ்வாய்வு.

பைத்தான் நிரல் (Python Program) உருவாக்கமும் செயல்பாடும்

பைத்தான் மொழியில் மெய்ம்மயக்கத்திற்கு நிரலாக்கம் எழுதப் பின்வரும் நிலையில் விதிகள் முதலில் வகுத்துக்கொள்ளப்பெற்றன.

- "மெய்ம்மயக்கம்1": "ட்றல்ள்+கசப"
- "மெய்ம்மயக்கம்2": "ல்ள்+யவ"
- "மெய்ம்மயக்கம்3": "ங்ஞ்ணந்ம்ன்+இனவொலி(கசடதபற)"
- "மெய்ம்மயக்கம்4": "ண்ன்+கசஞபமயவ"
- "மெய்ம்மயக்கம்5": "ஞ்ந்ம்வ்+ய"
- "மெய்ம்மயக்கம்6": "ம்+வ"
- "மெய்ம்மயக்கம்7": "ய்ர்த்+க ச த ப ஞ ந ம ய வ ங"
- "மெய்ம்மயக்கம்8": "ர்த் தவிர -> க...ன் + க...ன"
- "மெய்ம்மயக்கம்9": "ர்த் குற்றொற்றாகா"

இதன் அடிப்படையிலேயே தொல்காப்பியப் பைத்தான் நிரலாக்கமும் அமைந்துள்ளது. இதனைப் பின்வரும் நிரல்கள் மூலம் புரிந்துகொள்ளலாம்.

```
from meymayakkamfinal import *
```

இது meymayakkamfinal எனும் விதிநிரலை இங்கு அழைத்துச் சோதிப்பதற்குப் பயன்படுகின்றது.

```
import csv
```

இது csv எனும் கோப்பை ஆய்வதற்காகப் பயன்படுகின்றது.

```
ruleset={  
    "meymayakkam":(meymayakkam1,meymayakkam2,meymayakkam3,meymayakkam4,meymayakkam5,meymayakkam6,meymayakkam7,meymayakkam8,meymayakkam9),  
}
```

இது meymayakkamfinal-இல் உள்ள விதிகளை ஒரு அகராதி ruleset எனும் மாறிக்குள் பட்டியலாக உருவாக்கிக் கொள்கின்றது.

```
meymayakkam_names = {
```

```

"meymayakkam1": "ட்ற்ல்ள்+கசப",
"meymayakkam2": "ல்ள்+யவ",
"meymayakkam3": "ங்ஞ்ண்நம்ன்+இனவொலி(கசடதபற)",
"meymayakkam4": "ண்ன்+கசஞபமயவ",
"meymayakkam5": "ஞ்நம்வ்+ய",
"meymayakkam6": "ம்+வ",
"meymayakkam7": "ய்ர்த்+க ச த ப ஞ ந ம ய வ ங",
"meymayakkam8": "ர்த் தவிர -> க...ன் + க...ன்",
"meymayakkam9": "ர்த் குற்றொற்றாகா"
}

```

இது meymayakkam_names எனும் அகராதி மாறிக்குள் மெய்ம்மயக்கப் பெயர்களின் மதிப்புகள் பட்டியலாக உருவாக்கிக் கொள்கின்றது.

```

report={
'total_words':0,
'skipped_words':0,
'meymayakkam_correct':0
}

```

இந்த நிரலின் அடிப்படையில் ஆய்ந்து பார்க்கும் பொழுது பின்வரும் விளைவு கிடைக்கப் பெறுகின்றது. அதற்காகப் பெயர்ச்சொற்கள் கொண்ட தரவினை இந்த ஆணைத்தொடர் (Algorithm) பைத்தான் நிரலில் தந்து இயக்கும்பொழுது,

```

Report of nouns.csv
மொத்த சொற்களின் எண்ணிக்கை : 85269
சரி பார்க்காத ஓர் எழுத்துச் சொற்கள், ஈரெழுத்துச் சொற்களின் எண்ணிக்கை : 162
மெய்ம்மயக்க விதிப்படி உள்ள சொற்களின் எண்ணிக்கை : 24960

```

என அதன் விளைவைக் கணக்கிட்டுச் சொல்லிவிடுகின்றது. இது ஒரு தரவில் எவ்வளவு ஆய்வுசெய்யும் விதிப்படி உள்ள சொற்கள் உள்ளன என்பதை அறிய வழிவகை செய்து தருகின்றது. அதுபோல் தாவரச் சொற்றரவில் உள்ள சொற்களை ஆய்ந்து பார்க்கும்பொழுது,

```

Report of plant.csv
மொத்த சொற்களின் எண்ணிக்கை : 36
சரி பார்க்காத ஓர் எழுத்துச் சொற்கள், ஈரெழுத்துச் சொற்களின் எண்ணிக்கை : 14
மெய்ம்மயக்க விதிப்படி உள்ள சொற்களின் எண்ணிக்கை : 9

```

எனப் புள்ளிவிவரத்துடன் தருகின்றது. அந்தத் தரவுச் சொற்களைத் தனியாகவும் பிரித்து எடுத்துத் தருமாறு எழுதப்பெற்றுள்ளது. ஆகையால் ஒப்புநோக்கிக் கொள்ளவும் இது பயன்படும். அவ்வாறு ஒப்புநோக்கும்பொழுதுதான் இன்னும் தொழில்நுட்பம் புரிந்துகொள்ளுமாறு அந்த விதியை எப்படித் திருத்தலாம் என எழுத முடியும். அதன் அடிப்படையில் அடுத்தக்கட்ட நிரலாக்க மேம்பாடும் நிகழ்ந்தது. அதன் அடிப்படையில் அமைந்த தரவுகைளே பின்வரும் அட்டவணை-1இல் தரப்பெற்றுள்ளது.

அட்டவணை - 1, சரியான சொற்கள் - தவறான சொற்கள் பட்டியல்

S.No.	Correct Words	Correct Words	Incorrect Words
-------	---------------	---------------	-----------------

1	அரளிச்செடி	10. நாணல்	ஆவாரஞ்செடி
2	ஆமணக்கு	11. மக்காச்சோளம்	எருக்கஞ்செடி
3	எள்ளுச்செடி	12. வரகு	கம்பு
4	கற்றாழை	13. ரோஜா	கரும்பு
5	கள்ளி		காசித்தும்பை
6	காட்டாமணக்கு		கேழ்வரகு
7	குப்பைமேனி		சூரியகாந்தி
8	செடிப்பசலை		தும்பை
9	சோளம்		நந்தியாவட்டை

இப்படி ஒப்புநோக்கும் பொழுதுதான் அந்த விதியைத் திருத்தி அமைக்க வேண்டிய தேவையும் உணரப்பெற்றது. ஒரு குழந்தைக்குக் கற்றுக்கொடுப்பதுபோல் கணினிக்கு எழுதினால் என்ன என நோக்கும்பொழுது ஒரு மாற்றம் தோன்றியது. அது எழுத்துக்களின் அடிப்படையில் அதாவது ககரவரிசை முறையில் எழுதவேண்டும் என்பதாகும்.

விதியாக்கம் மாற்றம் பெறல்

அந்தக் கருதுகோளுக்கு ஏற்றாற்போல் தொல்காப்பியர் கூறிய மெய்ம்மயக்க விதி மெய்யெழுத்து வரிசைமுறையின்படி எழுதப்பெறுகின்றது. அதில் தொல்காப்பியர் சொன்ன கருத்துக்களைக் ககர வரிசைப்படி மாற்றி அமைத்துக் கொள்ளப்பெற்றது. அவை வருமாறு;

- "மெய்ம்மயக்கம்1": "க்+க"
- "மெய்ம்மயக்கம்2": "ங்+கங"
- "மெய்ம்மயக்கம்3": "ச்+ச"
- "மெய்ம்மயக்கம்4": "ஞ்+சஞய"
- "மெய்ம்மயக்கம்5": "ட்+கசடப"
- "மெய்ம்மயக்கம்6": "ண்+கசஞடணபமயவ"
- "மெய்ம்மயக்கம்7": "த்+த"
- "மெய்ம்மயக்கம்8": "ந்+தநய"
- "மெய்ம்மயக்கம்9": "ப்+ப"
- "மெய்ம்மயக்கம்10": "ம்+பமயவ"
- "மெய்ம்மயக்கம்11": "ய்+கசதபஞநமயவங"
- "மெய்ம்மயக்கம்12": "ர்+கசதபஞநமயவங"

"மெய்ம்மயக்கம்13": "ழ்+கசதபளுநமயவங்"

"மெய்ம்மயக்கம்14": "வ்வ"

"மெய்ம்மயக்கம்15": "ல்+கசபலயவ"

"மெய்ம்மயக்கம்16": "ள்+கசபளயவ"

"மெய்ம்மயக்கம்17": "ற்+கசபற"

"மெய்ம்மயக்கம்18": "ன்+கசருபமயவறன"

இவற்றை இன்னும் விரித்தும் எழுதலாம். அவ்வாறு எழுதும்பொழுது அது கணினிக்குப் புரியும்படி அமையும். அவை வருமாறு;

"மெய்ம்மயக்கம்1": "க்+க"

"மெய்ம்மயக்கம்2": "ங்+கங" -> ங் + க, ங் + ங

"மெய்ம்மயக்கம்3": "ச்+ச"

"மெய்ம்மயக்கம்4": "ஞ்+சஞய" -> ஞ் + ச, ஞ் + ஞ, ஞ் + ய

"மெய்ம்மயக்கம்5": "ட்+கசபட" -> ட் + க, ட் + ச, ட் + ட, ட் + ப

"மெய்ம்மயக்கம்6": "ண்+கசருடணபமயவ"

"ண்+க", "ண்+ச", "ண்+ஞ", "ண்+ட", "ண்+ண", "ண்+ப", "ண்+ம", "ண்+ய",
"ண்+வ"

"மெய்ம்மயக்கம்7": "த்+த"

"மெய்ம்மயக்கம்8": "ந்+தநய" -> "ந்+த", "ந்+ந", "ந்+ய"

"மெய்ம்மயக்கம்9": "ப்+ப"

"மெய்ம்மயக்கம்10": "ம்+பமயவ" -> "ம்+ப", "ம்+ம", "ம்+ய", "ம்+வ"

"மெய்ம்மயக்கம்11": "ய்+கசதபளுநமயவங்"

"ய்+க", "ய்+ங", "ய்+ச", "ய்+ஞ", "ய்+த", "ய்+ந", "ய்+ப", "ய்+ம", "ய்+ய", "ய்+வ"

"மெய்ம்மயக்கம்12": "ர்+கசதபளுநமயவங்"

"ர்+க", "ர்+ங", "ர்+ச", "ர்+ஞ", "ர்+த", "ர்+ந", "ர்+ப", "ர்+ம", "ர்+ய", "ர்+வ"

"மெய்ம்மயக்கம்13": "ழ்+கசதபளுநமயவங்"

"ழ்+க", "ழ்+ங", "ழ்+ச", "ழ்+ஞ", "ழ்+த", "ழ்+ந", "ழ்+ப", "ழ்+ம", "ழ்+ய", "ழ்+வ"

"மெய்ம்மயக்கம்14": "வ்வ"

"மெய்ம்மயக்கம்15": "ல்+கசபலயவ" -> "ல்+க", "ல்+ச", "ல்+ப", "ல்+ல", "ல்+ய", "ல்+வ"

"மெய்ம்மயக்கம்16": "ள்+கசபளயவ" -> ள்+க, ள்+ச, ள்+ப, ள்+ள, ள்+ய, ள்+வ

"மெய்ம்மயக்கம்17": "ற்+கசபற" -> ற்+க, ற்+ச, ற்+ப, ற்+ற

"மெய்ம்மயக்கம்18": "ன்+கசஞபமயவறன"

"ன்+க", "ன்+ச", "ன்+ஞ", "ன்+ப", "ன்+ம", "ன்+ய", "ன்+வ", "ன்+ற", "ன்+ன"

இவ்வாறு எழுதுவது எளிதில் தொழில்நுட்பத்திற்குப் புரியுமாறு அமையும். இவற்றுள் மெய்ம்மயக்கம் முதல் விதிக்கு (க் + க) மட்டும், பைத்தான் நிரலில் (Python Program) ஆணைத்தொடர் (Algorithm) உருவாக்கி இயக்கிப் பார்க்கும்பொழுது, பின்வரும் விளைவு கிடைக்கின்றது.

sathiyaraj@sathiyaraj-Aspire-ES1-521:~/ஆவணங்கள்/core_stable_Sathiyaraj/0தொல்காப் பியர்

_விதிகள்/tholkaappiyar_rules/meymayakkam_seperate_rules/Final_Meimayakkam_Update/T holkaappiyam_Rule_Development-/rule 1/16.5.2024\$ python3 one_ruletest.py

சொந்தரவில் உள்ள சொற்களின் எண்ணிக்கை : 85256
ஓரெழுத்து, ஈரெழுத்துச் சொற்களின் எண்ணிக்கை : 3870
விதி1-ற்குட்பட்ட சொற்களின் எண்ணிக்கை : 11022
விதி1-ற்குட்படாத சொற்களின் எண்ணிக்கை : 335
விதி1-ற்குப் பொருந்தாத சொற்களின் எண்ணிக்கை : 70029

இது எந்தளவிற்குத் துல்லியத்தைத் தரும் என்பதற்காகப் பின்வரும் sample_word_mixeddata.csv எனும் தரவு உருவாக்கப்பெற்று ஆய்வுசெய்து பார்க்கப்பெற்றது. அது பின்வரும் அட்டவணை - 2இல் இடம்பெற்றுள்ளது.

அட்டவணை - 2, sample_words

S.No.	sample_words	Correct Words	Incorrect Words
1	ககனாக்கிரகம்	ககனாக்கிரகம்	ககன்க்பிகிரகம்
2	ககன்க்பிகிரகம்	கக்கக்கெனல்	ககங்க்கன்கெனல்
3	கக்கக்கெனல்	கக்கசம்	ககன்கசம்
4	ககங்க்கெனல்	கக்கடி	ககனுகடி

5	கக்கசம்	கக்கதண்டம்	கக்பிகதண்டம்
6	கக்னகசம்		
7	கக்கடி		
8	கக்னுகடி		
9	கக்கதண்டம்		
10	கக்பிகதண்டம்		

எனவரும் தரவுச் சொந்தரவில் வரும் சிக்கலை அறிந்திட பின்வரும் சொல் பட்டியல் அட்டவணை - 3 உருவாக்கப்பட்டது. மேற்கண்ட அட்டவணை - 2இல் (sample_word_sameword.csv) இடம்பெறும் முதல் சொல்லாகிய 'ககனக்கிரகம்' என்பதன் மாற்றுச் சொல்லாக, அதில் உள்ள 'ன' எனும் எழுத்தை அகரவரிசைப்படி அமையுமாறு எழுதி உருவாக்கிக் கொள்ளப்பெற்றது.

அட்டவணை - 3, சொல் பட்டியல்

S.No.	sample_words
1	ககனாக்கிரகம்
2	ககனாக்கிரகம்
3	ககனிக்கிரகம்
4	ககனீக்கிரகம்
5	ககனுக்கிரகம்
6	ககனூக்கிரகம்
7	ககனெக்கிரகம்
8	ககனேக்கிரகம்
9	ககனைக்கிரகம்
10	ககனொக்கிரகம்
11	ககனோக்கிரகம்
12	ககனௌக்கிரகம்

இந்த அட்டவணை - 3-னைப் பைத்தான் நிரலாக்கத்தில் ஆய்வு உள்ளீடாகத் தரும்போது பின்வரும் விளைவு கிடைக்கப்பெறுகின்றது.

சொந்தரவில் உள்ள சொற்களின் எண்ணிக்கை: 12
ஒரெழுத்து, ஈரெழுத்துச் சொற்களின் எண்ணிக்கை: 0
விதி1-ற்குட்பட்ட சொற்களின் எண்ணிக்கை: 12
விதி1-ற்குட்படாத சொற்களின் எண்ணிக்கை: 0
விதி1-ற்குப் பொருந்தாத சொற்களின் எண்ணிக்கை: 0

இந்த விளைவின்படி அனைத்துச் சொற்களும் **விதி1-ற்குட்பட்ட சொற்கள்** எனத் தருவது பிழையான விளைவு என்பதை அட்டவணை - 3இல் உள்ள சொல் பட்டியலைப் பார்த்தாலே தெரியும். இன்னும் சில கருத்துக்களையும் உள்வாங்கிக் கொள்ள வேண்டும். ஏனெனில் ஒரு தரவகத்தில் ஒருகோடிச் சொற்கள் இருக்கின்றன என வைத்தால் துல்லியமாகச் சரியான சொற்களைக் கண்டறிய முடியுமா என ஐயம் தோன்றுகின்றது. ஆகவே, சில கேள்விகளும் கூடவே எழுகின்றன. அவை,

- மேற்கண்ட பட்டியல் போன்று பிழையான சொற்களை உருவாக்க இயலுமா?
- அவ்வாறு உருவாக்க வேறு ஏதேனும் நுட்பம் உள்ளதா?
- இயந்திரவழிக் கற்றல் மூலம் (Machine Learning Based) சாத்தியமா?
- செய்யறிவு நுட்பம்வழிச் (AI Based) சாத்தியமா?
- விதிமுறைவழிச் (Rule Based) சாத்தியமா?

என்பன. தற்போது செய்யறிவு மிக அதிகமாகப் பயன்படுத்தப்பட்டு வருகின்றது. அவற்றுள் கூகுள் செமினி செய்யறிவு (Gemini AI Tool) நுட்பத்தின் விளைவு எப்படி இருக்கின்றது என்பதை அறிவோம். அதற்காக மேலே உள்ள சொல்பட்டியலைத் தந்து, 'இதில் உள்ள பிழையான சொற்களைக் கண்டறிக' எனதக் கட்டளையாகத் தரும்பொழுது அது தரும் பதில் பின்வருமாறு அமைகின்றது.

தவறான சொற்கள்:

கீழ்க்கண்ட சொற்கள் தவறானவை:

ககனாக்கிரகம்
ககனிக்கிரகம்
ககனீக்கிரகம்
ககனெக்கிரகம்
ககனேக்கிரகம்
ககனௌக்கிரகம்

விளக்கம்:

"ககனம்" என்ற வார்த்தையின் சரியான வடிவம் "ஆகாயம்" என்பதாகும்.

"ஆகாயம்" என்ற வார்த்தையுடன் "கிரகம்" என்ற வார்த்தையை சேர்க்கும்போது, "ஆகாயக்கிரகம்" என்ற சரியான சொல் உருவாகிறது.

மேலே குறிப்பிட்ட தவறான சொற்களில், "க" என்ற எழுத்து அதிகமாகவோ அல்லது குறைவாகவோ பயன்படுத்தப்பட்டுள்ளது.

இந்தக் கூகுள் செமினி செய்யறிவு (Google Gemini AI) நுட்பத்தின் பதிலிலும் போதியளவு துல்லியம் இல்லை. ஆகையால், மெய்ம்மயக்கத்திற்கு மட்டும் விதி எழுதினால் போதாது என்பதையும், கூடுதலாகச்

- சொல்வளம்
- அகராதி வளம்

ஆகியன இருத்தல் வேண்டும் என்பதையும் உணரப் பெற்றது. அதனடிப்படையில் அணுகும்பொழுதுதான் சரியான மெய்ம்மயக்கச் சொற்களைக் கண்டறியும் நுட்பம் கிடைக்கும் எனும் தீர்வும் கிடைக்கபெற்றது. அதனை மெய்ப்பிக்கும் நிரல் மாற்றி [28] எழுதப்பட்டது. அந்த நிரலால் கிடைக்கும் விளைவு வருமாறு;

```
lenovo@lenovo-Lenovo-ideapad-110-15ACL:~/Documents/தொல்காப்பியம்_மெய்ம்மயக்கம்/  
First_Rule$ python3 one_ruletest.py
```

சொந்தரவில் உள்ள சொற்களின் எண்ணிக்கை: 12

ஓரெழுத்து, ஈரெழுத்துச் சொற்களின் எண்ணிக்கை: 0

விதி1-ற்குட்பட்ட சொற்களின் எண்ணிக்கை: 1

விதி1-ற்குட்படாத சொற்களின் எண்ணிக்கை: 11

விதி1-ற்குப் பொருந்தாத சொற்களின் எண்ணிக்கை: 0

இங்கு விதி ஒன்றை ஆய்விற்கு உட்படுத்தும் சொற்பட்டியல் தரப்பெற்றது. அதில் சரியான சொற்களும் பிழையான சொற்களும் இடம்பெறுமாறு அமைத்துக் கொள்ளப்பெற்றது. அதன்பின்பு, நிரலாக்கத்தில் இயங்குமாறு செய்தபோது மேற்கண்ட விளைவு கிடைக்கப்பெற்றது. அதற்காகச் சரியான சொற்பட்டியலும் வைத்துக் கொள்ளப்பெற்றது. அதில் 'ககனக்கிரகம்' என்ற சொல் மட்டும் தான் சரியான சொல் எனத் தனிக் கோப்பில் தரப்பட்டது என்பதையும் இங்குக் கவனத்தில்கொள்ள வேண்டும். ஆய்வுக்கு உட்படுத்தும் சொற்பட்டியலைச் சரிபார்த்த பின்பு, சரியான சொற்கள் எனக் கண்டறியும் சொற்களைப் பொருள்தரும் சொற்பட்டியலில் சரி என்ற சொற்கள் இருக்கின்றனவா என ஆய்வுசெய் என அதற்குக் கட்டளை இடப்பெற்றது. அது சரிபார்த்த பின்பு சரியான புள்ளிவிவரத்தையே இந்த நிரலாக்கம் வழங்குமாறு எழுதப்பெற்றுள்ளது. அந்த நிரலாக்கத்தை நேயக்கோ (neyakkoot) கித்துப்பில் (Github) [28] பார்க்கலாம்; ஆய்விற்கு உட்படுத்தியும் காணலாம்.

இந்தப் புள்ளி விவரத்தை வரைபடமாகக் காட்சிப்படுத்திப் பார்க்கும் பொழுதுதான் தெளிவாக ஆய்வின் போக்கினை அறியமுடியும். ஆகையால் பிளாட்லி அகராதி பயன்படுத்தப்பெற்றது. அதனைக் காட்சிப்படுத்திக் காட்ட பின்வருமாறு நிரல் எழுதிக் கொள்ளப்பெற்றது. அது வருமாறு;-

```
# பை படம் வரைதல்
```

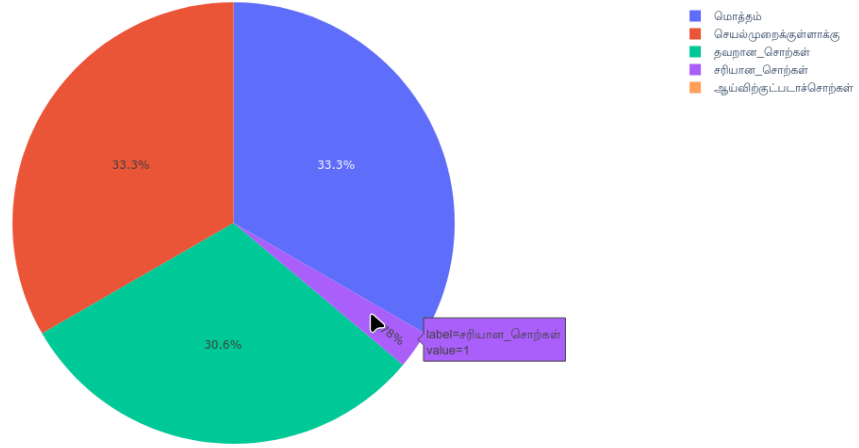
```
labels = ['மொத்தம்', 'செயல்முறைக்குள்ளாக்கு', 'சரியான_சொற்கள்',  
'தவறான_சொற்கள்', 'ஆய்விற்குட்படாச்சொற்கள்']
```

```
வரைபடமதிப்பீடு = [மொத்தம், செயல்முறைக்குள்ளாக்கு, சரியான_சொற்கள்,  
தவறான_சொற்கள், ஆய்விற்குட்படாச்சொற்கள்]
```

```
fig = px.pie(values=வரைபடமதிப்பீடு, names=labels)  
fig.show()
```

இந்த நிரல் விளைவின்படி பின்வரும் படம் - 1 தோன்றும். அதன்படம் வருமாறு:-

படம் - 1, பிளாட்லி வரைபடம்



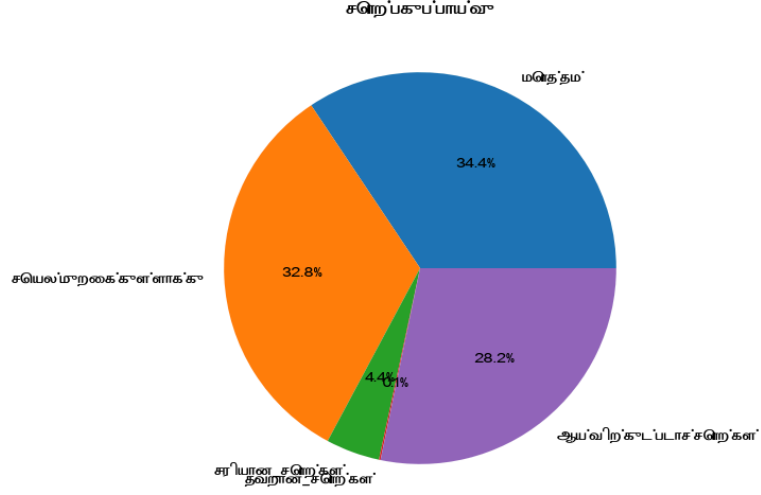
இவ்வாறு காட்சிப்படுத்திப் பார்க்கும் பைத்தான் அகராதிகளாக மேட்பிளாட்லிப், பிளாட்லி, சீபார்ன் போன்றன உண்டு. அவற்றுள் பெரும்பான்மை மேட்பிளாட்லிப் எனும் அகராதியே பயன்படுத்தப்பட்டு வருகின்றது. அதில் காட்சிப்படுத்திப் பார்க்கும்பொழுது சிக்கல் உருவானது. அதனைப் பின்வருமாறு விளக்கமாக அறிவோம்.

காட்சிப்படுத்தத்தில் சிக்கலும் தீர்வும்

ஆய்வுசெய்தவற்றைக் காட்சிப்படுத்திப் பார்க்கும்பொழுது தரவின் துல்லியத்தை இன்னும் மேம்படுத்துவதற்கான சூழல் உருவாகும். ஆகவே இயந்திரக் கற்றலில் காட்சிப்படுத்துவதற்கே தனி பைத்தான் அகராதிகள் நிரம்ப உள்ளன. அவற்றுள் மேட்பிளாட்லிப் (Matplotlib) பைத்தான் அகராதியைப் பயன்படுத்தி இயக்கிப் பார்க்கும்பொழுது தமிழ்மொழியைக் காட்சிப்படுத்துவதில் சிக்கல் வருகின்றது. அதனைக் காட்டும் படம்-2 வருமாறு;

படம்-2, தமிழ்ச்சொல் காட்சிப்பிழை,

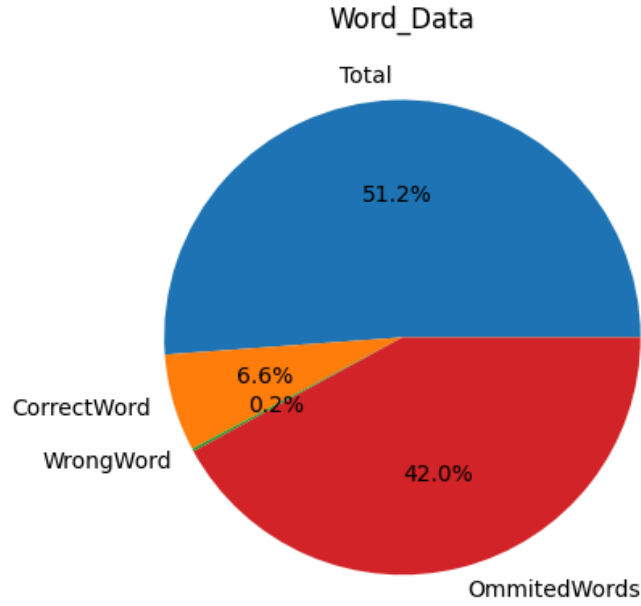
விதி1_ற்குட்பட்ட சொற்களின் எண்ணிக்கை: 525



இவ்வாறு வருவதால் ஆங்கிலத்தில் தகவல் தருமாறு மாற்றி எழுத வேண்டியுள்ளது. அதனைப் பின்வரும் படம்-3இல் பார்க்கலாம்.

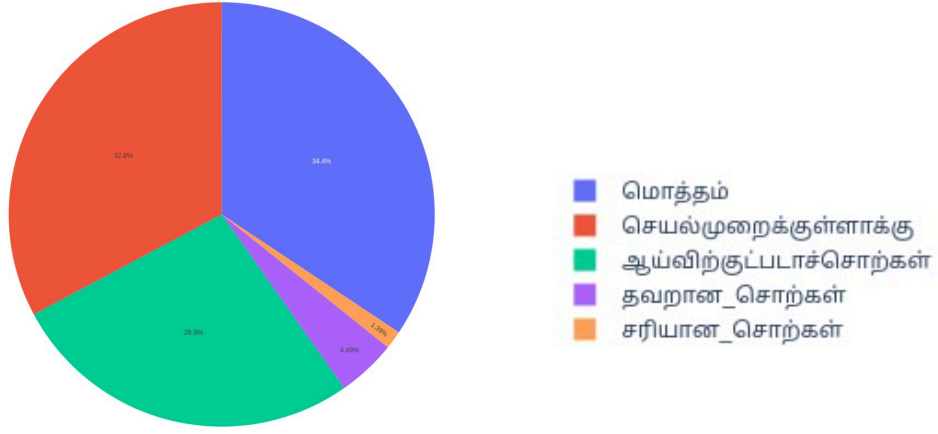
படம்-3, ஆங்கிலக் காட்சி நேர்த்தி

விதிபற்றுகுட்பட்ட சொற்களின் எண்ணிக்கை: 525



அதன்பிறகு வேறு எந்தப் பைத்தான் அகராதியைப் பயன்படுத்தலாமென ஆய்ந்து பார்க்கும்பொழுது பிளாட்லி (Plotly) அதற்கு ஏற்றத் தன்மையில் அமைந்து இருப்பதை அறியமுடிந்தது. அதனைப் பின்வரும் படம்-4இல் காணலாம்.

படம்-4, தமிழ்ச்சொல் காட்சி நேர்த்தி



இது அந்தச் சிக்கலிற்கான தீர்வாகவும் அமைந்தது.

முடிவுரையும் எதிர்காலச் செயல்பாடும்

இதுவரை விளக்கப்பட்ட கருத்தியல்களின் அடிப்படையில் நோக்கும்பொழுது இனிவரும் காலங்களில் தமிழில் இயந்திரவழிக் கற்றல் சார்ந்த ஆய்வுகள் பெருகி, இலக்கிய இலக்கணத் தரவுகள் மொழித் தொழில்நுட்பத்திற்கு ஏற்றாற்போல் உருமாறும் என்பதில் ஐயமில்லை. அவ்வாறு பெருகும் ஆய்வுகளுக்கு இது முன்னோடியாகவும் தூண்டுகோலாகவும் அமையும் என்பதில் எவ்வித மாற்றுக் கருத்தும் இல்லை.

தொல்காப்பிய மெய்ம்மயக்க விதிக்கு ஏற்படுத்தியிருக்கும் நிரல் உருவாக்கம் போன்று, ஒவ்வொரு இலக்கியத் தரவு உருவாக்கத்திற்கும் எழுதப்பெறும் பைத்தான் நிரல்கள் பொது உரிமத்தில் வைக்கப்பெற வேண்டும் எனவும், அதனை ஆய்வாளர்கள் ஆய்வறத்தோடு மேற்கோள் காண்பித்து எடுத்தாள வேண்டும் எனவும் இவ்வாய்வின் மூலம் வலியுறுத்தப்படுகின்றது.

துணைநின்றவை

1. தெய்வசுந்தரம் ந., 2021, மொழியியலும் கணினிமொழியியலும், சென்னை: அமுத நிலையம். ப.184.
2. த.சத்தியராஜ், தகவலுழவன், 17 அக்டோபர் 2022, விக்கிமூலமும் தமிழ் இயற்கை மொழி ஆய்வுகளும், கோயமுத்தூர்: இனம் பதிப்பகம்.
3. முனைவர் த.சத்தியராஜ், முனைவர் ரா.நித்யா, தகவலுழவன், 17 அக்டோபர் 2023, விக்கித்திட்டங்களில் பைத்தான் பயன்பாடு, கோயமுத்தூர் : இனம் பதிப்பகம்.
4. முனைவர் த.சத்தியராஜ், 2022, தமிழ் விக்கிமூலத்தில் குறுந்தொகைத் தரவு மேம்பாடு (E-content development for Kurunthogai resource in ta.Wikisource) Kalviyiyal Maanaaddu Aaivu Kovai 2021, Thoguthi-2.
5. Vinoth, A., Thangasamy, S., Nithya, R., Poovandran, G., Mounash, V., Subalalitha, C. N., ... & Jafer, K. S. (2023, December). **Automatic Identification of Meimayakkam in Tamil Words Using Rule Based and Transfer Learning Approaches.** In International

Conference on Speech and Language Technologies for Low-resource Languages (pp. 443-458). Cham: Springer Nature Switzerland.
https://link.springer.com/chapter/10.1007/978-3-031-58495-4_33)

6. Vinoth A, Sathiyaraj Thangasamy, John Paul Boopathi A, Poovandran G, **A development of the Meimayakkam second Rule Based on Tholkaappiyam and Nannul grammar concepts**, proceeding of ICSIEM 2024.
7. MAHALAKSHMI M., Charles Mahimainathan A., Dr. Vinoth A., Dr. Sathiyaraj Thangasamy, **DATA SCIENCE BASED CORPUS CREATION FOR THOLKAAPPPIYAM**, JNU, (2024)
8. Dr. Vinoth A., Dr. Sathiyaraj Thangasamy, Poovandran, G., **App Development for Tholkaappiya Meymayakkam First rule problem solve**, INFITT & Kumaraguru (2024)
9. Nedunchezhiyan, K.: Tholkappiyar period.
<https://newindian.activeboard.com/t59991225/topic-59991225/>. Accessed 14 மே 2024
10. கணியம், https://github.com/KaniyamFoundation/all_tamil_nouns, Accessed 14 மே 2024
11. Tamil Oneindia Homepage.
<https://tamil.oneindia.com/art-culture/essays/2010/0429-tolkappiyar-tamil-literature.html>. Accessed 14 மே 2024
12. shakespeare-dataset, GitHub - <https://github.com/topics/shakespeare-dataset>, Accessed 14 மே 2024
13. GitHub - cobanov/shakespeare-dataset: complete works, plays, sonnets and poems of shakespeare, <https://github.com/cobanov/shakespeare-dataset>, Accessed 14 மே 2024
14. learning-hadoop-and-spark/0b-Example-Datasets/shakespeare-davinci/shakespeare.raw - <https://github.com/lynnlangit/learning-hadoop-and-spark/blob/master/0b-Example-Datasets/shakespeare-davinci/shakespeare.raw>, Accessed 14 மே 2024
15. TensorFlow-2-for-Deep-Learning-Specialization/Customising-your-models-with-TensorFlow-2/Week3_Programming_Assignment.ipynb, https://github.com/doducthao/TensorFlow-2-for-Deep-Learning-Specialization/blob/master/Customising-your-models-with-TensorFlow-2/Week3_Programming_Assignment.ipynb, Accessed 14 மே 2024
16. machine-learning-experiments/experiments/text_generation_shakespeare_rnn/text_generation_shakespeare_rnn.ipynb, <https://github.com/trekhleb/machine-learning-experiments/blob/master/experiments/>

- text-generation-shakespeare-rnn/text-generation-shakespeare-rnn.ipynb, Accessed 14 மே 2024
17. [fedjax/fedjax/datasets/shakespeare.py](https://github.com/google/fedjax/blob/main/fedjax/datasets/shakespeare.py),
<https://github.com/google/fedjax/blob/main/fedjax/datasets/shakespeare.py>,
Accessed 14 மே 2024
 18. Shakespeare-database, <https://github.com/topics/shakespeare-database>, Accessed 14 மே 2024
 19. shakespeare-PyPI, <https://pypi.org/project/shakespeare/>, Accessed 14 மே 2024
 20. py-shakespeare-PyPI, <https://pypi.org/project/py-shakespeare/>, Accessed 14 மே 2024
 21. Shakesparelang, <https://shakesparelang.com/1.0/>, Accessed 14 மே 2024
 22. Python NLTK Shakespeare corpus,
<https://stackoverflow.com/questions/43721175/python-nltk-shakespeare-corpus>,
Accessed 14 மே 2024
 23. பாலசுப்பிரமணியன் க., 2017, தொல்காப்பிய இலக்கண மரபு, சென்னை - அரிமா நோக்கு.
 24. பாலசுப்பிரமணியன் க., 2015, தொல்காப்பியத்தின் ஒருமையும் முழுமையும், சென்னை - உலகத் தமிழாராய்ச்சி நிறுவனம்.
 25. Repository migrated to KanchiLUG's repository
<https://gitlab.com/kachilug/tamilrulepy>, Accessed 16 மே 2024
 26. தமிழண்ணல், 2008, தொல்காப்பியம் மூலமும் கருத்துரையும், மதுரை; மீனாட்சி புத்தக நிலையம். ப.12.
 27. தொல்காப்பியக் குறுஞ்செயலி,
https://github.com/neyakkoot/tholkaappiyam_mobil_app_development,
Accessed 16 சூன் 2024
 28. தொல்காப்பிய மெய்ம்மயக்கம்,
https://github.com/neyakkoot/tholkappiyam_meymayakkm_first_rule_developed,
Accessed 16 சூன் 2024
 29. தெய்வசுந்தரம் ந., 2021, மொழியும் தமிழ் இலக்கணமும், சென்னை: அமுத நிலையம். ப.99.
 30. சத்தியராஜ் தங்கச்சாமி, ஏப்பிரல் 2024, தொல்காப்பியம் - நூன்மரபு (பைத்தான்வழி உரை 1), இலக்கணம் காலாண்டுச் சஞ்சிகை 6, பூ 2; இதழ் 2, சித்திரை - ஆனி, ISSN: 2961-5712, பக்.22-28.

31. சத்தியராஜ் தங்கச்சாமி, ஆகஸ்ட் 2024, தொல்காப்பியம் - நூன்மரபு (பைத்தான்வழி உரை 2), இலக்கணம் காலாண்டுச் சஞ்சிகை 7, பூ 2; இதழ் 3, ஆடி - புரட்டாதி, ISSN: 2961-5712, பக்.11-14.