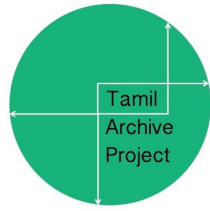


படிப்பகம் 



விக்கிமூலம்

நூலக நிறுவனம்  
Noolaham  
Foundation



Singapore  
Tamil



தமிழ் மரபு அறக்கட்டளை



தமிழ் எண்ணிம  
நூலகங்களுக்கா  
ன தமிழ்க்  
கணிமைத்  
தேவைகள்

இ. நற்கீரன்  
சனவரி 18, 2020

# உள்ளடக்கங்களை தமிழ் மொழியில் ...

- கண்டுபிடித்தல்/உலாவுதல் - Discovering/Browsing
- தேடல் - Search
- அனைவருக்கும் அணுகக்கப்படுத்தல் - Making it Accessible to All
- பகுத்தாய்தல், - Analyze


The screenshot shows the National Digital Library website. The search bar contains the text 'தமிழ்' (Tamil). Below the search bar, it displays 'Showing 1 out of a total of 1 results. (0.001 seconds)'. The result is a book titled 'தமிழ் நூல் முதலமைச்சர் செகரெட்டர் (தமிழ்)'. The book details include: Author: சி.வி. வெங்கடேசுவரன், Subject: Tamil Language, Date Issued: 2005, 2013, and Has Files: 36/31.

The screenshot shows the National Digital Library website with search results for 'தமிழ்'. The results are displayed in a list format. The first result is a book titled 'உரை சிவிலிங்கம் தமிழ் அளவளவும் / Urair Sivalingam Tamil alyalavum'. The book details include: Author: Kavitha Ciciriyappan, Book, 2014, 133 pages, and 3 copies available. The second result is a book titled 'தமிழ் ஒர் அறிமுகம் / Tamil, Or arimukam = Tamil, an introduction'. The book details include: Author: Vignesh Srinivasan, Book, 2018, 1 volume, and 3 copies available.

The screenshot shows the National Digital Library website with search results for 'தமிழ்'. The results are displayed in a list format. The first result is a book titled 'உள்ளடக்க வகை'. The book details include: Author: தமிழ் நூல் முதலமைச்சர் செகரெட்டர், Book, 1994, and 3 copies available. The second result is a book titled 'குறிச்சொற்கள்'. The book details include: Author: தமிழ் நூல் முதலமைச்சர் செகரெட்டர், Book, 1994, and 3 copies available. The third result is a book titled 'தமிழ் ஒர் அறிமுகம் / Tamil, Or arimukam = Tamil, an introduction'. The book details include: Author: Vignesh Srinivasan, Book, 2018, 1 volume, and 3 copies available.

# பன்மொழி மீதரவு

- தமிழ் ஒருங்குறியில்
- தமிழில் முழுமையான மீதரவு
- தமிழில் தேடல்
- முழு உரைத் தேடல்
- சமமான பன்மொழி மீதரவு
- எழுத்துப்பெயர்ப்பு
- அவசியமில்லை


 **EUSL Research Gateway**  
Library, Eastern University, Sri Lanka Login

Research Gateway - EUSL → Thesis Collection → Undergraduate Project Report → Faculty of Arts & Culture → View Item

Show simple item record

dc.contributor.author	பஞ்சலிங்கம், ரமணிமணி	
dc.date.accessioned	2019-11-25T09:30:21Z	
dc.date.available	2019-11-25T09:30:21Z	
dc.date.issued	2018	
dc.identifier.uri	http://www.digital.lib.esn.ac.lk/handle/123456789/4106	
dc.language.iso	ta	en_US
dc.publisher	கலை கலாசார பீடம்	en_US
dc.subject	வரலாறு நூல்-History Books	en_US
dc.subject	பண்டைய காலம்-Ancient Time	en_US
dc.subject	இலங்கை தமிழர் - Sri Lanka Tamils	en_US
dc.title	வரலாற்று நூல்களில் இலங்கைத் தமிழர் ஆற்றும்-இன்றும்	en_US
dc.identifier.sslno	FAC697	en_US

Files in this item

 **Name:** FAC697.pdf [View/Open](#)  
**Size:** 357.4Kb  
**Format:** PDF

This item appears in the following Collection(s)

- [Faculty of Arts & Culture \[705\]](#)

[Show simple item record](#)

**Search Research Gateway**

Search Research Gateway  
 This Collection

**Browse**

All of Research Gateway  
[Communities & Collections](#)  
[By Issue Date](#)  
[Authors](#)  
[Titles](#)  
[Subjects](#)

This Collection  
[By Issue Date](#)  
[Authors](#)  
[Titles](#)  
[Subjects](#)

**My Account**

[Login](#)  
[Register](#)



# தமிழ் எழுத்துணரி (Tamil OCR)

- தேடல், பகுப்பாய்வுக்கு அவசியம்
- கட்டற்ற மென்பொருள் - Tesseract 4
- கட்டண மென்பொருள் - கூகிள் எழுத்துணரி API

பக்கம்: கவியரங்கில் முடியரசன் (இரண்டாம் பதிப்பு).pdf/17 தொகுக்கப்படுகிறது

https://ta.wikisource.org/s/8lne **Tesseract OCR Button**

தலைப்பு வடிவமைப்பு வடிவமைப்பு OCR மேம்பட்ட சிறப்பு எழுத்துருக்கள் உதவி மெய்ப்பு பார்க்கும் கருவிகள்

வேலு (சேர்த்துக்கொள்ளாதது):

பக்க நடுப்பகுதி (ஒருங்கிணைவுக்கு சேர்க்கப்பட வேண்டாம்):

14

தமிழறிவைத் தத்து சிறப்பித்தது. அக்கல்லூரியில் நிகழும் அறிஞர் பெருமக்களின் உரைகள், அப்போது நிகழ்ந்த உரையாடல்கள் கவிஞரின் உள்ளத்தில் ஆழ்ந்த மொழிப் பற்றையும் இனப்பற்றையும் கிளர்ந்தெழச் செய்தன.

முடியரசன் தமது 21-ஆம் வயதில் எழுதிய, 'சாதி என்பது நமக்கு ஏனோ?' என்ற கவிதையே அவரது முதல் கவிதையாகும். இது திராவிட நாடு' இதழில், பெரியகுளம் துரைராசு' என்ற பெயரில் வெளியாயிற்று.

1940-ஆம் ஆண்டில் தந்தை பெரியாரின் தன் மான இயக்கத்தில் தொடர்பு கொண்டார். இத் தொடர்பு அவருடைய ஆளுமையை வெற்றிபெறச் செய்த போதிலும் 1943-ஆம் ஆண்டு வித்துவான் தேர்வில் தோல்வியுறச் செய்தது. அவர் தோல்வியுறவில்லை; தோல்வியுறாமலு செய்ப்பட்டார்.

இடைமீலே நவாயு டி. எசு. இராசமாணிக்கம் நாடகக் குழு தம்பால் பணியாற்ற வருமாறு அழைத்தது. சென்றார். அங்கிருந்த சிறைவாழ்க்கையும் மதவழிபாட்டு முறைகளும் வெறுப்பை

14

தமிழறிவைத் தத்து சிறப்பித்தது. அக்கல்லூரியில் நிகழும் அறிஞர் பெருமக்களின் உரைகள், அப்போது நிகழ்ந்த உரையாடல்கள் கவிஞரின் உள்ளத்தில் ஆழ்ந்த மொழிப் பற்றையும் இனப்பற்றையும் கிளர்ந்தெழச் செய்தன.

முடியரசன் தமது 21-ஆம் வயதில் எழுதிய, 'சாதி என்பது நமக்கு ஏனோ?' என்ற கவிதையே அவரது முதல் கவிதையாகும். இது 'திராவிட நாடு' இதழில், 'பெரியகுளம் துரைராசு' என்ற பெயரில் வெளியாயிற்று.

1940-ஆம் ஆண்டில் தந்தை பெரியாரின் தன் மான இயக்கத்தில் தொடர்பு கொண்டார். இத் தொடர்பு அவருடைய ஆளுமையை வெற்றிபெறச் செய்த போதிலும் 1943-ஆம் ஆண்டு வித்துவான் தேர்வில் தோல்வியுறச் செய்தது. அவர் தோல்வியுறவில்லை; தோல்வியுறாமலு செய்ப்பட்டார்.

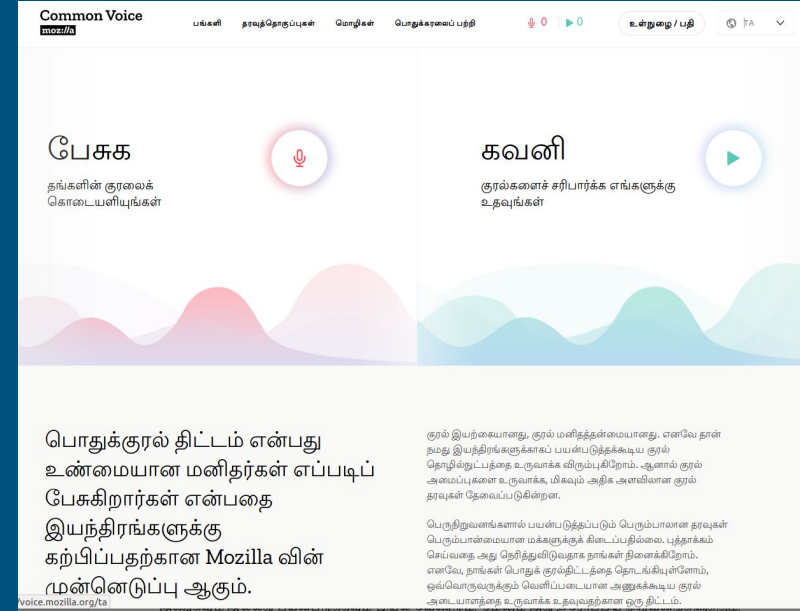
# தமிழ் உரை ஒலி மாற்றி (Speech Synthesis)

- கண்தெரியாதோருக்கு உள்ளடக்கங்களை அணுகக்கப்படுத்தல்
- தமிழ் திரை வாசிப்பான்கள் (Tamil Screen Readers)
- உயர்தர உரை ஒலி மாற்றி (Quality Tamil Speech Synthesis)
- தமிழ் ஒலி நூற்கள் (Tamil Audio Books)
- ஒரு படைப்பை ஒலி வடிவில் இலகுவாக மாற்றுவதற்கான Tooling, வழிகாட்ட



# பேச்சணரி (Speech to Text)

- காது கேட்பாதோருக்கு உள்ளடக்கங்களை அணுக்கப்படுத்தல்.
- தேடல், பகுப்பாய்வுக்கு ஏதுவாக்கல்.
- மேற்குநாடுகளில் இது சட்டம் ஆகும்.
- கட்டற்ற மென்பொருள் - மொட்சிலா பொதுக்குரல் கிட்டம் (கொடக்க நிலை)



The screenshot shows the Common Voice Mozilla website interface. The page is titled "Common Voice mozilla" and has a navigation bar with links for "பக்களி", "தரவுத்தொகுப்புகள்", "மொழிகள்", and "பொதுக்களம் பற்றி". There are also icons for "உள்நுழை / பதி" and "FA".

The main content area is divided into two columns. The left column is titled "பேசுக" (Speak) and features a microphone icon. Below it, the text reads: "தங்களின் குரலைக் கொண்டையளியுங்கள்" (Share your voice). The right column is titled "கவனி" (Listen) and features a play button icon. Below it, the text reads: "குரல்களைச் சரிபார்க்க எங்களுக்கு உதவுங்கள்" (Help us check the recordings).

Below the main content area, there is a section with text in Tamil. The text reads: "பொதுக்குரல் திட்டம் என்பது உண்மையான மனிதர்கள் எப்படிப் பேசுகிறார்கள் என்பதை இயந்திரங்களுக்கு கற்பிப்பதற்கான Mozilla வின் மன்னெடுப்பு ஆகும்." (The Common Voice project is a Mozilla initiative to teach machines how real people speak. It is a public domain project.)

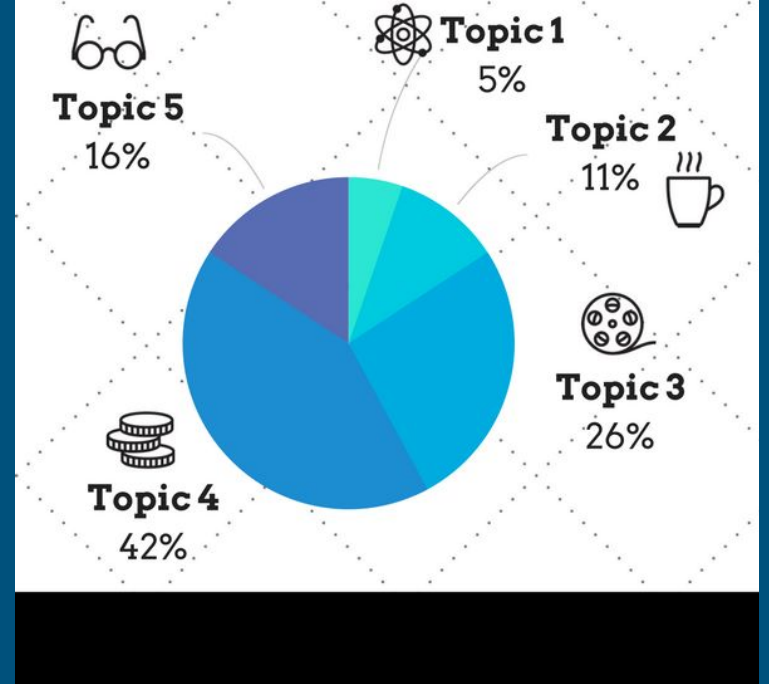
There are also two columns of smaller text in Tamil. The left column reads: "குரல் இயற்கையானது, குரல் மனிதர்கள்மையானது. எனவே தான் தமது இயந்திரங்களுக்காகப் பயன்படுத்தக்கூடிய குரல் தொழில்நுட்பத்தை உருவாக்க விரும்புகிறோம். ஆனால் குரல் அளவப்புகளை உருவாக்க, மிகவும் அதிக அளவிலான குரல் தரவுகள் தேவைப்படுகின்றன." (Voice is natural, voice is human. That's why we want to build voice technology for our machines. But to build voice technology, we need a lot of voice data. We need a lot of voice data.)

The right column reads: "பெருநிறுவனங்களால் பயன்படுத்தப்படும் பெரும்பாலான தரவுகள் பெரும்பான்மையான மக்களுக்குக் கிடைப்பதில்லை. புத்தாக்கம் செய்வதை அது நெறித்துவிரிவாக தங்கள் நினைக்கிறோம். எனவே, தங்கள் பொதுக் குரல்திட்டத்தை தொடங்கியுள்ளோம், ஒவ்வொருவருக்கும் வெளிப்படையான அணுகக்கூடிய குரல் அளவியாளத்தை உருவாக்க உதவுவதற்கான ஒரு திட்டம்." (Most of the data used by large companies is not available to most people. We want to build it differently. That's why we started our public domain project, we want to build a public domain project, we want to build a public domain project.)

# இயந்திரக் கற்றல்

நூலகங்களில் இயந்திரக் கற்றல் மீதரவு உருவாக்கம், மேம்பாட்டுக்கு (metadata enrichment) உதவுகிறது.

- வகைப்படுத்தல் - Classification
- துறை/குறிசொற்கள் கண்டுபிடிப்பு - Topic Modeling
- ஆவணச் சுருக்க உருவாக்கம் - Automatic summarization
- பெயரிடப்பட்ட உரிபொருள் கண்டுபிடிப்பு - Named Entity

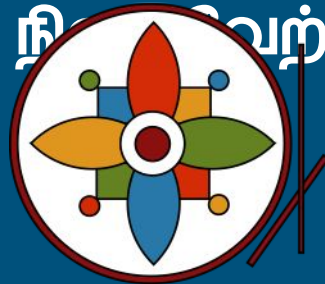






# எண்ணிம நூலகங்களும் தமிழ்க் கணிமையும்

- எண்ணிம நூலகங்கள் தமிழ்க் கணிமையை பயனர்களிடம் எடுத்துச் செல்வதற்கு முக்கிய ஒரு துறை.
- தமிழ்க் கணிமைக்குத் தேவையான தரவு வளங்களைத் தர வல்லன.
- தமிழில்/பன்மொழியில் மீதரவு, இடைமுகம், ஊடாட்டம் என்பது நாம் எதிர்பார்க்க வேண்டியது, வேண்ட வேண்டியது, நாமே நினைவற்ற வேண்டியது.



அனைவருக்காகவும்



நூலகங்கள்