

கணினி வழி திருக்குறள் தரவக உருவாக்கம்

பேரா. பெ. காளியானந்தம்

பகுதி நேர முனைவர் பட்ட ஆய்வாளர்
தமிழ்ப்பண்பாட்டு மையம்
அழகப்பா பல்கலைக்கழகம், காரைக்குடி
சிவகங்கை மாவட்டம்

முனைவர் கா. கணநாதன்

இணைப்பேராசிரியர்
தமிழ்த்துறை, தமிழ்ப்பண்பாட்டு மையம்
அழகப்பா பல்கலைக்கழகம், காரைக்குடி
சிவகங்கை மாவட்டம்

மலர்: 11

சிறப்பிதழ்: 1

மாதம்: ஏப்ரல்

வருடம்: 2024

P-ISSN: 2321-788X

E-ISSN: 2582-0397

DOI:

[https://doi.](https://doi.org/10.34293/sijash.v11iS1-April.7741)

[org/10.34293/sijash.](https://doi.org/10.34293/sijash.v11iS1-April.7741)

[v11iS1-April.7741](https://doi.org/10.34293/sijash.v11iS1-April.7741)

முன்னுரை

இருபத்து ஒன்றாம் நூற்றாண்டின் இணையற்ற அறிவியல் சாதனமாகக் கணினியில் நுல்களைப் பதிவு செய்து பாதுகாக்கும் பணி உலகெங்கிலும் எல்லா மொழிகளிலும் வெகு வேகமாக நடை பெற்று வருகிறது. மின்னூல் என்பது நூல் ஒன்றினுடைய மின்னணுவியல் அல்லது எண்முறை பதிப்பாகும். மின்நூலானது பொதுவாக பதிப்பாளர்களால், மின்னணுவியல், எண்முறை ஊடகங்கள் மூலம் தமது புத்தகங்களை விநியோகிப்பதற்காக உருவாக்கப்படுகிறது.

இவை வெற்று உரை வடிவமாகவோ நூல் சம்பந்தப்பட்ட சிறப்பு தகவல்களை தம்மகத்தே கொண்டவையாகவோ இருக்கும்.

தரவகம்

தரவகம் என்பது மின்னூரைகள் அல்லது மின்நூல்களின் தொகுப்பு எனலாம். ஒழுங்குமுறையுடன் அதிக அளவில் தேர்வு செய்யப்பட்டு கணினியில் சேமிக்கப்பட்ட இயற்கையான நடைகளை உடைய உரைகள். தரவகம் ஒரு குறிப்பிட்ட மொழியின் சொற்கள் அல்லது துறைச் சொற்களை (Language varieties or Register) உள்ளடக்கியதாக இருக்கவேண்டும். தரவகத்தில் இரண்டு முக்கிய காரணிகள் உள்ளன. 1. தரவக வடிவமைப்பு 2. அளவு இவ்விரு காரணிகளும் தரவகத்தின் தரத்தை உறுதி செய்கின்றன. தரவகத்தை இரு பெரும் பிரிவுகளாகப் பிரிக்கலாம். 1. உரைத்தரவகம் (Text corpus), 2. பேச்சுத் தரவகம் (Speech corpus).

இத்தரவக உருவாக்கம் சொல் மற்றும் இலக்கணத் திருத்தி, தேடு பொறிகள், மொழிசார்க் கருவிகள் போன்றவற்றுக்கு மிகவும் பயன்படக்கூடியன.

கணினி மற்றும் இணைய மொழி பயன்பாட்டில் பலநிலைகள் இனிமேல்தான் அடையப்பட வேண்டும். இவற்றைத் தரவகத்தை உருவாக்குவதன் மூலம் எளிதாக அடைய முடியும். தரவகங்கள் அடிப்படை சொல் திருத்திகளிருந்து எந்திர மொழி மாற்றங்கள் வரை எதிர்கொள்ளப்படும் சிக்கல்களுக்குத் தீர்வளிக்கக் கூடியவை. இதுவரை கணினியில் சங்க இலக்கியங்களின் பயன்பாடு ஆரம்ப நிலையிலேயே உள்ளது. இக்கட்டுரை சங்க இலக்கியங்களில் ஒன்றான திருக்குறள் தரவக உருவாக்கம் பற்றி விளக்குகிறது.

தரவக மொழியியல்

தரவக மொழியியல் என்பது மொழியை எவ்வாறு பயன்படுத்துகிறார்கள் என்று படிப்பது. அதாவது மொழியும் அதன் பயன்பாடும் என்று கூறலாம். ஒரு மொழி எவ்வாறு பேசப்படுகிறது மற்றும் எழுதப்படுகிறது என்று ஆராய்வது. மேலும் தரவக மொழியியல், மொழியின் கூறுகளாகிய சொற்கள், சொல்லுருவாக்கம், இலக்கணம், சொற்றொடர் அமைப்பு சொற்பொருள் அமைப்பு, உரையாடல் அமைப்புப் போன்றவற்றை மொழிப் பயன்பாட்டின் வழி ஆராய்வது. தரவக மொழியியலைப் பற்றி அறிந்து கொள்ள சொல்லடைவும் தொடரடைவும் இன்றியமையாததாக அமைகிறது.

தரவகத்தின் பயன்கள்

கல்வி நுட்பவியல், மொழி நுட்பவியல் துறைகளுக்கும், மின் அகராதி, சொல் திருத்தி இலக்கணத்திருத்தி, பேச்சு மொழியிலிருந்து எழுத்து மொழியை உருவாக்கும் கருவி. எழுத்து மொழியிலிருந்து பேச்சு மொழியை உருவாக்கும் கருவி, மொழிபெயர்ப்பு கருவி போன்ற மென்பொருள் உருவாக்கத்திற்கும் பயன்படுகிறது.

ஒரு சொல் அதன் தொடரியல் சூழலுக்கேற்பப் பொருள் மாற்றம் பெறுவதைத் தரவகம் மூலம் அறியலாம்

மொழியியல் விதிகள், கோட்பாடுகள் சரியானவைதானா என்பதைத் தரவகம் மூலம் அறியலாம்.

வரலாற்று முறை இலக்கணம், வரலாற்று முறை அகராதி தயாரிப்பதற்குத் பயன்படுகிறது.

சீர், உருபன், சொல், தொடர் முதலியவற்றைப் பிரிக்கப் பயன்படும் கருவிகளை உருவாக்குவதற்கும், வேர்ச் சொல்லாய்வுகள், பல்வேறு அகராதிகள், பொருட்களஞ்சியங்கள் தயாரிப்பதற்கும் தரவகம் பயன்படுகின்றது.

திருக்குறள்

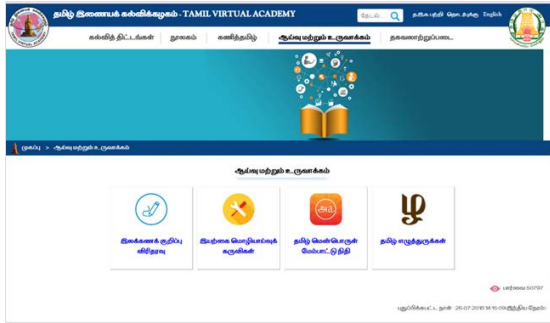
பதினெண்கீழ்க்கணக்கு நூல்களில் காலத்தால் முற்பட்டது திருக்குறள். சங்கம் மருவிய காலத்தில் தொடக்க நிலையில் திருக்குறள் தோன்றியது எனலாம். குறள் வெண்பாக்களால் ஆகிய திருக்குறள் 1330 குறள்களைக் கொண்டது. 133 அதிகாரங்களையும், அறத்துப்பால், பொருட்பால், காமத்துப்பால் என முப்பாலினையும் கொண்டது. தமிழ் இலக்கியங்களில் மிகவும் குறுகிய அடிகளைக் கொண்ட ஒரே நூல் இதுவே. இருளகற்றும் விளக்காய் அறியாமையை நீக்கி அறிவு ஒளியைத் தருவது திருக்குறளே எனச் சான்றோர் பலரும் உயர்த்தி ஏத்தும் புகழினைக் கொண்டது. தமிழரின் அற வாழ்வைக் காட்டும் நூல்களில் திருக்குறளின் பங்கு மிகவும் உயரியது திருக்குறள். ஒரு சமூகத்தளத்தில் நின்று பாடப்பட்டது. எனினும், உலகனைத்திற்கும் பொதுவான தன்மைகள் அந்நூலில் மிகுதியாயுள்ளன. கடுமையான குறிக்கோள்களை நடைமுறைக்கு ஒவ்வாத நிலையிலிருந்து பேசாமல் வாழ்க்கையை உயர்ந்தவனாகவும் இனியதாகவும் ஆக்கிக் கொள்வதற்குப் பயன்படும் ஒரு பண்பாட்டு வழிகாட்டியாகத் திருக்குறள் காட்டும் அறநெறிக் கோட்பாடுகள்

விளங்குகின்றன.. இதில் சொல்லப்படாதது எதுவுமில்லை என்னும் பெருமை கொண்டது. அத்துடன் எல்லாச் சமயங்களுக்கும் பொதுவாக விளங்குகின்றது. இப்படிப்பட்ட திருக்குறளை எழுதிய திருவள்ளுவர் உலகின் மிகச்சிறந்த பொது மறையாளர் என்று அழைக்கப் படுகின்றார்.

The screenshot shows the Tamil Virtual Academy website interface. At the top, there is a header with the academy's name in Tamil and English, along with navigation links. Below the header, there is a search bar and a language selector. The main content area displays a grid of Tamil text, likely a list of words or phrases, arranged in multiple columns and rows. The text is presented in a clean, readable font against a light background.

சங்க இலக்கியத் தரவகம்

தமிழ் இணையக் கல்விக்கழகம் தமிழ் மொழிக்கான மென் பொருள்களை உருவாக்கி அவற்றை இணையத்தில் வெளியிட்டு வருகிறது. இவற்றில் சங்க இலக்கியங்களுக்கான தரவகம் அதற்குரியத் தரவகத்தைப் பயன்படுத்தி பல்வேறு மென்பொருள்களை உருவாக்கி இணையத்தில் அளித்து வருகிறது.



(படம் 1: தமிழ் இணையக் கல்விக்கழக கணித்தமிழ்)

இவற்றைப் பதிவிறக்கம் செய்து தரவக உருவாக்கம், தரவக ஆய்வுகள், சொல் ஆய்வுகள், தொடரடைவுகள், சொல்லடைவுகள் போன்றவற்றை உருவாக்கிக்கொள்ளலாம், இக்கருவிகளைப் பயன்படுத்தித் திருக்குறள் (படம் 2) தொடரடைவு மற்றும் சொல்லடைவு போன்ற ஆய்வுகளை மேற்கொள்ளலாம் இலவசமாக பதிவிறக்கம் செய்து கொள்ளலாம்.

படம் 2: திருக்குறள் தரவகம்

திருக்குறள் தரவகம்

தரவக உருவாக்கம்

```
<?xml version="1.0" ?>?<?xml-stylesheet type="text/css"?>
<html><body><text>
<h1>திருக்குறள்</h1>
<h2>Tamil Text</h2>
</h3><h4>Compiled by Thiruvalluvar</h4>
<Tamil Text>
<unitid>01</unitid>
<meta>
<text>திருக்குறள்</text>
<language> தமிழ்</language>
<edi>திருவள்ளுவர்</edi>
<Total>1330-prose</Total>
<குறள்-001>
அகர முதல எழுத்தெல்லாம் ஆதி பகவன் முதற்றே
உலகு</குறள்-001>
<குறள்-002>கற்றதனால் ஆய பயனென் கொல்வால-
றிவன் நற்றாள் தொழா ஆர் எனின்</குறள்-002>
<குறள்-003>
மலர்மிசை ஏகினான் மாணடி சேர்ந்தார் நிலமிசை
நீடுவாழ்வார் </குறள்-003>
</p>
</text>
```

செம்மொழித் தரவகம்

செம்மொழித் தமிழாய்வு மத்திய நிறுவனம் சங்க இலக்கியங்களுக்கான தரவகங்களை உருவாக்கி இணையத்தில் அளித்து வருகிறது.



படம் 3: செம்மொழித் தரவகம்

முடிவுரை

திருக்குறள் தரவகம் உருவாக்கும் போது மூலபாடம், சந்தி பிரித்த பாடம் ஆகியவற்றை உருவாக்க வேண்டும் உருவாக்கும் போது எவ்வகையான சொற்பிரிப்பு முறையை பின்பற்றுவது என்பதை முடிவு செய்தல் வேண்டும். பொருண்மை அடிப்படையிலே அல்லது

மொழியியல் அடிப்படையிலே சொற்பிரிப்பு நெறிமுறையை உருவாக்கி அதன் அடிப்படையில் தரவகம் உருவாக்கப்பட வேண்டும். இவ்வாறு உருவாக்கப்படும் திருக்குறள் தரவகம் இயற்கை மொழி ஆய்வில் சங்க இலக்கியப் பயன்பாடுகளுக்கு முக்கிய பங்கு அளிக்கும்.

துணை நூல்கள்

1. <https://www.tamilvu.org/>
2. <https://www.tamilvu.org/>
3. [t%E0%A E%95%E0%AE%A3%E0%AE%BF%E0 %AE%A4%E0%AF%8D%E0%AE%A4%E0%AE %AE%E0%AE%BF%E0%AE%B4%E0%AF%8D](#)
4. Akilan R., Naganathan E R., ‘‘Morphological Analyzer for Classical Tamil texts a rule&based approach’’, ARPN Journal of Engineering and Applied Sciences Volume : 10, No: 20, ISSN : 1819&6608, 2015
5. <https://ctl.t.cict.in/>