

தமிழுக்கான கட்டற்ற சந்திப் பிழை திருத்தி - உருவாக்கம், பயன்பாடுகள்

நித்யா துரைசாமி nithyadurai87@gmail.com

சுருக்கம் :

தமிழில் சொற்பிழைத் திருத்தியின் ஒரு முக்கிய கூறு, சந்திப் பிழைத் திருத்தி ஆகும். கட்டற்ற மென்பொருளாக, பைதான் நிரலாக்க மொழியில், ஒரு சொற்பிழைத்திருத்தி உருவாக்கும் நோக்கத்தில், அதன் முதல் படியாக, ஒரு சந்திப்பிழைத் திருத்தி உருவாக்கப்பட்டுள்ளது. எழுத்துகளை எளிதில் கையாளுவதற்கும், உயிர்மெய் எழுத்துகளை எளிதில் பிரிக்கவும் சேர்க்கவும் open-tamil என்ற கட்டற்ற மென்பொருள் நிரல் பயன்படுத்தப்பட்டுள்ளது.

இக்கட்டுரையில், சாதாரண பைதான் நிரலில் தமிழைக் கையாளுதலில் உள்ள சிக்கல்கள், open-tamil எவ்வாறு அவற்றைத் தீர்க்கிறது?, சந்திப்பிழைத் திருத்தி உருவாக்க உதவிய தமிழ் விதிகள், அவற்றை நிரலாக்கம் செய்த முறைகள் ஆகியவற்றை விரிவாகக் காணலாம்.

1. அறிமுகம் :

தமிழில் பல சொற்களை இணைத்து ஒன்றாக எழுதுவது சொற்றொடர் எனப்படும். அவ்வாறு சொற்றொடர்களை எழுதும்போது இரண்டு சொற்களுக்கு இடையில் ஒற்றெழுத்துக்களை சேர்க்க வேண்டிய இடத்தில் சேர்க்காமல் விட்டுவிடுவதும், தேவையில்லாத இடங்களில் சேர்ப்பதுமே சந்திப்பிழை எனப்படும். இத்தகைய பிழைகளை நீக்குவதற்கு முதலில் நாம் இதற்கான இலக்கணத்தை நன்கு புரிந்து கொள்ள வேண்டும். எந்தெந்த இடங்களில் சேர்க்க வேண்டும், எந்தெந்த இடங்களில் சேர்க்கக் கூடாது என்பதற்குப் பல்வேறு விதிகள் உள்ளன. அவை வல்லினம் மிகும் இடங்கள், வல்லினம் மிகா இடங்கள் என்று அழைக்கப்படுகின்றன. அவற்றை எடுத்து நிரலாக்கம் செய்வதன் மூலம் இத்தகைய பிழைகளை நாம் சுலபமாகக் களைந்து விடலாம்.

ஒரு சொற்றொடரில் முதலில் உள்ள வார்த்தை நிலைமொழி எனவும், அதைத் தொடர்ந்து எழுதப்படும் வார்த்தை வருமொழி எனவும் அழைக்கப்படும். நிலைமொழியைத் தொடர்ந்து வரும் வருமொழியின் முதல் எழுத்து வல்லின மெய்யைத் தழுவிய எழுத்துக்களாக இருப்பின் சூழ்க்கண்ட இலக்கண விதிகளைப் பொறுத்து அவ்வெழுத்து மிகும் அல்லது மிகாது. அதாவது வருமொழியின் முதல் எழுத்து க் ச் ட் த் ப் ற் போன்ற வல்லின மெய்யின் மூது ஏறி வரும் உயிர்மெய் எழுத்துக்களைப் பெற்றிருப்பின் அவை பின்வரும் விதிகளைப் பொறுத்து இவ்வெழுத்துக்களை இட வேண்டுமா வேண்டாமா என்பதை முடிவு செய்யும்.

வல்லினம் மிகா இடங்களுக்கான விதிகள்:

1. குறியீடுகள் (அட்டா! , 50\$, எது?)
2. சுட்டு, வினா அடியாகத் தோன்றியவை மற்றும் அஃறிணைப் பன்மை (எத்தனை பழங்கள்?, பல பசு)
3. வந்த, கண்ட, சொன்ன, வரும் என்பன போன்ற பெயரெச்சங்களோடு படி, ஆறு என்னும் சொற்கள் சேர்ந்து வரின் (கண்டவாறு சொன்னான்)
4. எண்ணுப்பெயர்கள் (ஐந்து சிறுவர்கள்)
5. ஒடு & ஓடு என உயிர் ஈறு கொண்டவை (கத்தியோடு நின்றான்)
6. 'கொண்டு' என்னும் சொல்லுருபு (கத்திகொண்டு குத்தினான்)
7. இல் என்பதோடு இருந்து என்னும் சொல்லுருபு (வீட்டிலிருந்து சென்றான்)

8. இன் என்பதோடு நின்று என்னும் சொல்லுருபு (வீட்டினின்று வெளியேறினான்)
9. ஆறாம் வேற்றுமைக்கு உரிய அது (எனது புத்தகம்)
10. 'உடைய' என்னும் சொல்லுருபு (என்னுடைய புத்தகம்)
- 11- 13. ண்டு, ந்து, ன்று என முடியும் மென்தொடர்க் குற்றியலுகர வினையெச்சங்கள் (கண்டு பேசினார், கொன்று குவித்தான்)
14. ய்து என முடியும் இடைத்தொடர்க் குற்றியலுகர வினையெச்சங்கள் (செய்து தந்தான்)
- 15 - 18. ஆத, இய, ஐய,ற்ற,ல்ல, ட்ட கின்ற, உம் ஆகிய விசுதிகள் பெற்று முடியும் மற்ற பெயரெச்சங்கள் (அழியாத கல்வி, பெரிய பெண், இன்றைய செய்தி, கற்ற சிறுவன்)
- 19 - 21. ஆக, அன, யுற,றுகு,க்கு ஆகிய விசுதிகள் பெற்று முடியும் மற்ற வினையெச்சங்கள்
22. இரட்டைக் கிளவி, அடுக்குத்தொடர்கள் (கல கல)
23. வியங்கோள் வினைமுற்று (வருக புலவரே)
24. பொதுப்பெயர்கள் (மகளிர் கல்லூரி)
25. கள், தல் விசுதிகள்
26. பல, சில, ஏவல் வினை (வா கலையரசி)
28. பொதுப்பெயர்கள், உயர்திணைப் பெயர்கள், வடமொழி சொற்கள்

வல்லினம் மிகும் இடங்களுக்கான விதிகள்:

1. ஓர் எழுத்துச் சொற்கள் (கை குழந்தை)
3. வன்தொடர்க் குற்றியலுகரம் (கற்று கொடுத்தான்)
4. மென்தொடர்க் குற்றியலுகரம் (குரங்கு குட்டி)
5. முற்றியலுகரம் - தனிக் குறில் அடுத்து வரும் உகரம் (பொது பணி)
6. உயிர்த்தொடர்க் குற்றியலுகரம் (விறகு கடை)
7. வல்லினமெய் அல்லாத பிற மெய்களின் மேல் ஏறி வருகின்ற உகரம் (தேர்வு கட்டணம்)
8. இரண்டாம் வேற்றுமை விரி (கனியை தின்றான்)
9. நான்காம் வேற்றுமை விரி (எனக்கு கொடு)
- 10 - 13. அகர, இகர, யகர மெய் ஈற்று வினையெச்சம் (வர சொன்னான், குறிஞ்சி தலைவன், தேங்காய் சட்னி, தயிர் குடம்)
14. ஈறுகெட்ட எதிர்மறைப் பெயரெச்சம் (தூரா சிக்கல்)
- 15 - 16. மகர இறுதி கெட்டு உயிர் ஈறாய் நிற்கும் சொற்கள் (மரம் கிளை)

கூழ்க்கண்ட வலைப்பதிவுகளிலிருந்து இலக்கண விதிகள் யாவும் எடுக்கப்பட்டுள்ளன.

<http://www.tamilvu.org/courses/degree/c021/c0214/html/c0214661.htm>

<https://pprakash.me/tamil/2016/05/16/va-mig-mik/>

நிரலாக்கம் :

open-tamil எனும் பைதான் library இந்த சந்திப்பிழை திருத்தி உருவாக்கத்தில் பெரும்பங்கு வகித்துள்ளது. அதாவது தமிழில் உள்ள எழுத்துக்களை தமிழ் நடையிலேயே பிரித்து எடுப்பதற்கு இதில் பல்வேறு functions உள்ளன. அவற்றின் பயன்பாடுகளைப் பின்வருமாறு காணலாம்.

- **get_letters()** எனும் function ஒரு வார்த்தையில் உள்ள எழுத்துக்களை list வடிவில் மாற்றி அவற்றை ஒவ்வொன்றாக index மூலம் அணுக உதவுகிறது. அதாவது "தொக்கு" எனும் வார்த்தையில் உள்ள எழுத்துக்களை இந்த function-ஆனது 'தொ', 'க்', 'கு' எனும் list வடிவில் பிரிக்கும். இதையே open-tamil அல்லாத வேறு ஏதாவதொன்றால் பிரித்தால் அவை முறையே 'ெ', 'த', 'ா', 'க', 'ு', 'க', 'ு' என்பது போன்ற list-ஆகப் பிரிக்கப்படும்.
- **splitMeiUyir()** எனும் function ஒரு உயிர்மெய் எழுத்தில் உள்ள உயிர் எழுத்தையும் மெய் எழுத்தையும் பிரித்து list வடிவில் சேமிக்கிறது. அதாவது இந்த list-ன் 0 எனும் index உயிர் எழுத்தையும், 1 எனும் index மெய் எழுத்தையும் குறிக்கும்.
- மேலும் உயிர் எழுத்துகள், மெய் எழுத்துகள், உயிர் மெய் எழுத்துகள், வடமொழிச்சொற்கள் போன்றவை நிரலில் எளிதில் பயன்படுத்தும் வகையில் பட்டியலாகக் கிடைக்கின்றன.

open-tamil நிரலைக்கொண்டு எழுதப்பட்ட இந்த சந்திப்பிழைத் திருத்தி, இப்போது open-tamil நிரலின் ஒரு பாகமாகவே இணைக்கப்பட்டுள்ளது.

பயன்பாடுகள் :-

open-tamil மற்றும் சந்திப்பிழைத்திருத்தி ஆகியவை பைதான் library ஆக இருக்கின்றன. இதனால் பொது மக்கள் பயன்படுத்த இயலாது. ஆனால் இந்த நிரல்களைப் பயன்படுத்தி, பல்வேறு மென்பொருட்களை எளிதில் உருவாக்கலாம்.

உதாரணமாக, open-tamil ன் பல்வேறு பயன்பாடுகளை விளக்கும் வகையில் <http://tamilpesu.us> என்ற கட்டற்ற இணைய செயலி உருவாக்கப் பட்டுள்ளது. இதன் ஒரு பகுதியாக சந்திப்பிழைத்திருத்தியும் உள்ளது. Python, Django Framework கொண்டு இந்த மென்பொருள் உருவாக்கப்பட்டுள்ளது. இதன் எளிய இடைமுகப்பில், பல்வேறு மொழியியல் கருவிகள் இணைக்கப்பட்டு வருகின்றன. இது API வசதியும் கொண்டிருப்பதால், பல்வேறு செயலிகள், மென்பொருட்கள், கைபேசி செயலிகள் மூலம் எளிதில் இணைத்து பயன்படுத்த இயலும்.

The screenshot shows a web browser window with the URL <http://tamilpesu.us/sandhi-check/>. The page title is "Open Tamil Web" and it includes a search bar and navigation links for "Eng" and "Documentation". The main content area is titled "தமிழ் சந்திப்பிழை திருத்தி" (Tamil Sandhi Checker). Below the title, it states: "Tamil Sandhi Checker is project by Mrs. Nithya Shrinivasan, distributed under GNU GPLv3 terms. Tamil Sandhi Checker is an ongoing project to provide automatic detection and correction of 'sandhi' grammatical errors in writing Tamil. Read more about this project at [here on github](#)." There is a text input field labeled "சங்க தமிழ்" (Sangam Tamil) and a list of features on the left sidebar: "பயன்பாடுகள்" (Applications), "Text to Speech Synthesizer", "சொல் தேடல்" (Word Search), "தமிழ் எண்கள்" (Tamil Numbers), "தமிழ் சந்திப்பிழை திருத்தி" (Tamil Sandhi Checker), "பெருக்கல் அட்டவணை" (Multiplication Table), and "மற்றவை" (Others).

அடுத்த செயல் திட்டங்கள் :

1. பல்வேறு இன்னும் பல்வேறு விதிகளை சேர்த்தல்
2. LibreOffice Plugin, MS Office Plugin ஆக வழங்குதல்
3. விக்கிப்பீடியா கட்டுரைகள் அனைத்தையும் சேர்த்து, தானியக்கமாகத் திருத்துதல்
4. Firefox, Chrome உலாவிகளுக்கு Plugin வழங்குதல்
5. Python மொழியில் இருந்து விலக்கி, Rule Engine ஆக உருவாக்கி, எந்த நிரலாக்க மொழியிலும் பயன்படுத்துமாறு உருவாக்குதல்

முடிவுரை :

open-tamil என்ற நிரல் தொகுப்பு, கட்டற்ற நிரலாக மூல நிரலுடன் வெளியிடப்பட்டதால், அதன் மூலம் சந்திப்பிழைத் திருத்தி, எளிதில் உருவாக்கப்பட்டது. இணைய செயலியாகவும் வெளியிடப்பட்டது. தமிழ் இலக்கணத்தின் தெளிவான விதிகள், இது போன்ற மொழியியல் கருவிகள் உருவாக்குவதற்கு உதவுகின்றன. பலரும் இணைந்து, பல்வேறு கட்டற்ற மொழிக்கருவிகளை உருவாக்கி, தமிழின் பயன்பாட்டை, அடுத்த தலைமுறைகளுக்கு எடுத்துச் செல்வோம்.

உசாத்துணைகள் :

1. open-tamil மூலநிரல் - <https://github.com/Ezhil-Language-Foundation/open-tamil>
2. tamilpesu.us மூலநிரல் - <https://github.com/abuvanth/open-tamil-web-interface>
3. சந்திப்பிழைத்திருத்தி மூலநிரல் - <https://github.com/nithyadurai87/tamil-sandhi-checker>
4. சந்திப்பிழைத்திருத்தி அறிமுகக் காணொளி - <https://www.youtube.com/watch?v=eC82S7wOr3E&t=916s>