NF : Project : Tamil OCR Implementation

செயற்றிட்ட தலைப்பு	NF : Project : Tamil OCR Implementation
செயற்றிட்ட இடம்	இலங்கை
நிறைவேற்றும் அமைப்பு	நூலக நிறுவனம்
முக்கிய பங்குத்தாரர்கள்	நூலகம் நிறுவனம்
காலம்	Jan 2016 - Dec 2016

நூலக நிறுவனம்

நூலக நிறுவனமானது இலங்கைத் தமிழ் பேசும் சமூகங்களின் அறிவுத் தொகுதிகளையும் மரபுரிமைகளையும் ஆவணப்படுத்துவதில் ஈடுபட்டு வரும் ஒரு தன்னார்வத் தொண்டு நிறுவனமாகும். 2005 முதல் திறந்த, கூட்டுச் செயற்பாட்டை முன்னிறுத்திச் செயற்பட்டு வரும் நூலக நிறுவனம் இலங்கைத் தமிழ் பேசும் சமூகங்கள் தொடர்பான ஒரு முதன்மையான உசாத்துணைத் திரட்டினை உருவாக்கியுள்ளது. பல்வேறு செயற்றிட்டங்கள் மூலம் ஆவணப்படுத்தற் பணிகளை நூலக நிறுவனம் முன்னெடுத்து வருகிறது.

செயற்றிட்டச் சுருக்கம்

நூலக நிறுவனத்தின் ஊழியர்களுக்கும், தன்னார்வலர்களுக்கும், இந்நிறுவனத்தின் செயற்திட்டங்களில் ஒன்றான, நூலக நிறுவன தமிழ் எழுத்துணரியாக்கச் செயலாக்கம் (NF Tamil OCR Implementation) என்ற செயற்திட்டம் பற்றிய அறிமுகத்தை வழங்குவதும், அந்தச் செயற்திட்டத்தை முன்னெடுப்பதற்குத் தேவையான தகவல்களைத் தொகுப்பதும் ஆகும். இந்த ஆவணம் தொழினுட்ப அணி, ஊழியர்கள், தன்னார்வலர்கள் ஆகியோருக்கும் பயன்படும்.

இந்த ஆவணத்தின் வாசகர்கள்

நூலக நிறுவனத்தின் ஊழியர்களுக்கும், தன்னார்வலர்களுக்கும், இந்நிறுவனத்தின் செயற்திட்டங்களில் ஒன்றான, நூலக நிறுவன தமிழ் எழுத்துணரியாக்கச் செயலாக்கம் (NF Tamil OCR Implementation) என்ற செயற்திட்டம் பற்றிய அறிமுகத்தை வழங்குவதும், அந்தச் செயற்திட்டத்தை முன்னெடுப்பதற்குத் தேவையான தகவல்களைத் தொகுப்பதும் ஆகும். இந்த ஆவணம் தொழினுட்ப அணி, ஊழியர்கள், தன்னார்வலர்கள் ஆகியோருக்கும் பயன்படும்.

இலக்குகளும் நோக்கங்களும்

நூலகத்தின் உள்ளடக்கம் தற்போது pdf வடிவிலேயே உள்ளன. தமிழில் எழுத்துணரியாக்கம் செய்வதற்கான நுட்பம் கூகிள் ஆவணக api ஊடாகவும், சூறீனிவாசன் உருவாக்கிய OCR4wikisource சாத்தியமாக்கியது. இவற்றைப் பயன்படுத்தியும், மேலதிக கருவியாக்கம் ஊடகவும் நூலக உள்ளடக்கத்தை எழுத்துணரியாக்கம் செய்து html வடிவில் பகிர்வதே இந்தச் செயற்திட்டத்தின் நோக்கம் ஆகும். எழுத்துணரியாக்கம் உள்ளடக்கத்தை பயனர்கள் இலகுவாகத் தேட, தரவிறக்க, பயன்படுத்த உதவுகின்றது.

முக்கிய பங்கேற்பாளர்கள்

- நூலகப் பயனர்கள்
- நூலக நுட்பச் செயலாக்கம்
- நூலக தொடர்பாடல் செயலாக்கம்
- தமிழ்க் கணிமை ஆர்வலர்கள்

நூலக நிறுவன நோக்கங்கள்/வியூகங்களுடன இணைவு

எண்ணிம நூலக, ஆவணக உள்ளடக்கத்தை பயனர்களுக்குக் அணுக்கப்படுத்தல் நூலக நிறுவனத்தின் முக்கிய நோக்கங்களில் ஒன்றாகும். தேட, தரவிறக்க, பயன்படுத்த எழுத்துணரியாகக்ம் உதவிசெய்து அந்த நோக்கத்தைச் செயற்படுத்த உதவும்

நடிபங்குகளும் பொறுப்புக்களும்

நடிபங்கு/Role	பொறுப்பு/Responsibility	அறிக்கையிடல்/Reporting
Staff Coordinator - Gajani (gajani31@gmail.com)	அறிக்கையிடல்	RB/வழிகாட்டுநர் சபை
Project Manager/Coordinator - Thakaval-Uzhavan (tha.uzhavan@gmail.com)	செயற்திட்டத்தை நிறைவேற்றல், மேலாண்மைச் செய்தல்,	Staff Coordinator/to RB if required
Project Designer - NF Technology/Natkeeran (natkeeran@gmail.com)	Project Documentation, Design, Evaluation	RB/வழிகாட்டுநர் சபை
Project Oversight - RB/வழிகாட்டுநர் சபை	Project Documentation, Design, Evaluation	
Developer/DevOps - Thakaval-Uzhavan (tha.uzhavan@gmail.com), Natkeeran (natkeeran@gmail.com)	Preparation of input data files; Preparation of html by executing the scripts; uploading to server, creating wiki page; creating sitemap	Project Manager
Subject Matter Experts - Shrinivasan T, Sundar Lakshmanan	Provide input/support with regards to Tamil OCR tooling	

செயற்பரப்பு செயற்பாடுகளும்

- Develop tooling to do OCR implementation using Google OCR engine.
- The goal is to automate as much of the work as possible.
- Develop input files, execute scripts

திட்ட முடிவு வரையறை

- **P**hase 1 of this project involves creation of the scripts to do batch ocr. The scripts must be released to the public under FOSS license GPL.
- Phase 1 of this project aims to ocr 5000 document**s**

கால அட்டவணையும் மைல்கற்களும்

- Jan 2016 - Dec 2016

நிதி வள மூலங்கள்

- Natkeeran has committed to fund the initial phase.
- The initial pilot phase provided a stipend of 10 000 Indian Rupees for Thakaval-Uzhavan.

சீர்தரங்களும் தரக்கட்டுப்பாடும்

We don't control the Google OCR quality. It has been evaluated as very useful and practical. It can be assumed that the quality of the pdf provided contributes to the quality.

இடர் மேலாண்மை

The load to the noolaham.net server may increase.

Issue/Change Management

எழுத்துணரியாக்கத் தொழினுட்பங்களையும், தொடர்ந்து அவற்றில் ஏற்படுத்தப்படும் மாற்றங்களையும், இந்த github பக்கத்தில் அவ்வப்போது இற்றைப்படுத்தப்படுகின்றன.