Contributing Singapore Named entities to Wikidata

NKOS 2023

Chow Yaw Huah & Glenn Hong
National Library Board Singapore

Abstracts
The National Library Board of Singapore (NLB) has created thousands of authorised local names over the past 10 years for indexing of its resources. These names of entities comprise people, organisations and places and include their variant forms such as aliases, and previous or successive forms. Associated information about the entities, such as the awards of a person, latitude and longitude of a place or the founding year of an organisation, are also included. NLB ensures trustworthiness of the data through authoritative research and by cross-checking content gleaned from crowdsourced knowledge bases such as Wikipedia.

These names are published on international registries such as the Virtual International Authority File (VIAF). In making these names available locally and overseas, indexing of content resources can be improved through the consistency of names used and the context provided with the information in each named entity.

Another knowledge base where these named entities' data are exposed is the Wikidata knowledge base. Wikidata is a collaboratively edited knowledge base hosted by the Wikimedia Foundation. Wikidata serves as a central knowledge base for Wikimedia projects (eg. Wikipedia, Wiktionary, Wikinews, etc) as well as being a freely available open database of linked open data for other projects. It is a free knowledge base that can be read and edited by both humans and machines. Wikidata aims to represent everything in the universe in a way computers can understand. Denny Vrandečić, who founded Wikidata in 2012, saw the need for a place where humans and bots could share knowledge on more equal terms. Libraries participate in Wikidata to tap on their uses and applications of linked open data to advance and enrich discovery of locally curated collections on the global web.

Wikidata also allows for the integration of external identifiers. These identifiers, which were created and reside in outside sources, are reconciled in Wikidata. Adding these external identifiers to Wikidata items provides links that point to another set of data residing elsewhere, which aids in establishing connections and providing additional information about the items.

This is where NLB come into the picture as this capability fits into our work of creating Singapore-centric named entities of people, organisations, and places (eg. Zubir Said, Catherine Lim, People's Action Party, and Anderson Bridge). With Wikidata's reach to the Semantic Web community, NLB was exploring how to contribute our named entities and its related information to Wikidata as part of our bigger linked data discovery project to expose NLB's resources to the Semantic Web. This work of exposing our named entities in Wikidata forms part of the infrastructure required for discovery of NLB's resources eventually.

This presentation highlights the journey undertaken by NLB in understanding the nature and benefits of Wikidata, as well as the challenges of using online tools such as OpenRefine and QuickStatements to automate the work process of reconciling and contributing NLB's local named entities to Wikidata. There was a process of iterative learning on how to insert identifiers of NLB named entities in

Wikidata. The learning also included queries to the Wikidata community, and experimenting with Wikidata tools to enable searching, matching, loading, and creating records in Wikidata.