Daniel Orongo Nyangweso

danielorongo@gmail.com
https://orcid.org/0000-0003-0978-8327
PhD School of Earth Sciences,
ELTE Eötvös Loránd University
Budapest, Hungary

Mátyás Gede

gedematyas@gmail.com
https://orcid.org/0000-0002-8639-2812
Institute of Cartography and Geoinformatics
ELTE Eötvös Loránd University
Budapest, Hungary

https://doi.org/10.4467/K7501.45/22.23.18067

Use of Toponym Ontologies in Web Gazetteer Entities

Abstract

Toponyms presented in web gazetteers have various applications such as land registration, defense mapping, cultural heritage, memory inscription and topographic representation. This study aims to demonstrate the application of geographical ontology in the study of toponyms in a multilingual space of individual entities or classes in gazetteers and user interactions. The display of toponyms on the web platforms must be appealing and easy to read for ease of comparison of entities. To capture user attention, processes that users are familiar with and knowledge systems for geographic information retrieval must be used to efficiently present toponym information. In reviewing the user experience, the gazetteer service was evaluated based on the language, the country, acquisition, platform, browser, and operating system. Content, type, location, regularity, extinction, and ethnographic associations in Kenya were integrated as part of mappings to form an integrative approach in dealing with open data. The study's findings show that users' perceptions of toponym gazetteer services are influenced by the visual features and information given. The results of this study may help in the future design of gazetteer entities that incorporate toponymy aspects in dealing with the increasing complexity of searching for entities and retrieval geographic information gazetteer.

Keywords
gazetteer, geographic ontology, toponyms

1. Introduction

Gazetteers are defined as an index of geographical directory usually used together with an atlas or maps with descriptions of features in terms of location and associated descriptive geographic information (Goodchild & Hill, 2008; Hill, 2009). Web gazetteers are those gazetteers showing the features described in indexed geographic databases and made available online as web maps such as on the Geonames and OpenStreetMap platforms. This article aims to demonstrate the application of geographical ontology in studying toponyms interactively in a multilingual space of individual entities or classes in gazetteers. The article presumably fills the gap in understanding the usage of information access and retrieval by users. To capture user attention, processes that users are familiar with and knowledge systems for geographic information retrieval ontologies must be used to efficiently present toponym information. Ontologies generally define terms and rules used and shared by related users within a domain, and their relationships in gazetteers. In knowledge sharing, an ontology is said to be "an explicit specification of a conceptualization" (Guarino, 1997, p. 295). An ontology can be deployed as a basic ontology with concepts, a gazetteer, or a semantic network. A semantic network (a term from psychology) is a graph with vertices representing concepts and edges indicating semantic connections between concepts (Rada et al., 1989). Unlike ontologies, semantic networks show semantic relationships between concepts that have linguistic characteristics or terms. Gazetteer entities are objects in ontologies for linked gazetteers in which entities describe toponyms. An entity is an object that occupies a space in the real world, occurs independently or as part of another object, and is described by a set of attributes (Watt & Eng, 2014). The entity object may be physical or conceptual. Entity classes are selected based on the strength of relationships among them to derive relationships to another entity by assigning primary and foreign keys. There are two types of entities, independent and dependent. Independent entities are kernels, building blocks of a database with a primary key that can be simple or composite and does not depend on another entity. Dependent or derived entities connect to kernels and exist by depending on two or more tables with many-to-many relationships and are connected to at least two foreign keys.

Applications of gazetteers include land registration, defence mapping, cultural heritage, memorial inscriptions and topographic display (Acheson et al., 2017). Accessing web gazetteer entities requires computer interaction by accessing entity descriptions in a visualized map interface. Measuring usage statistics of a web gazetteer can be done using the Google Analytics tool or a web crawler (Hu et al, 2014).

The complexity of web-based ontologies depends on the spatial knowledge of the users interacting with the interactive media to influence the choice of service offered by the web gazetteer, which ranges from quality of design, to ease of use, to simplicity to improve usability. However, most ontologies offer the discovery of geographic information, where the design of ontologies is based on the knowledge of users in the discovery of entities (Mechouche et al., 2013) on a need basis, after an interaction by artificial intelligence or computer system. The discovery of these entities from gazetteer on the web depends on a star scheme ranking (Berners-Lee, 2010).

This paper aims to answer the question of what usage experiences have been made concerning interest in gazetteers in terms of web pages visited, dwell time and range of access over time.

The process of selecting the tool for evaluation considered the query support of semantic tools, query of entity classes, schema and file export, visualization of external data as links, editing of class, external data sources' storage retrieval through Application Programming interface (API) and feature class visualizations. The parameters served as minimum requirements for ontologies that played a role in the study of user experiences on gazetteer websites and the selection process.

However, during the review, it was found that some ontology editing software does not support saving changes to files but creates new ontology files when editing is done over the web, such as the web WebVowl software and the Fluent Ontorion Editor 2015, which require saving a new file. In a study of problems and approaches to ontology development (Abdelmoty et al., 2007), OWL's XML (Extensible Markup Language) was found to be inefficient in dealing with large geographic datasets because Ontology Web Language (OWL) and Resource Description Frameworks (RDF) do not support geospatial data processing (due to the assumption of non-unique names) and therefore are not suitable for checking constraints on spatial data.

The rest of the article is structured as follows. **Section 2** presents a global review on the use of existing geographic ontologies for gazetteer ontology services and entities, as well as for some selected countries. **Section 3** on materials and techniques explores the integration of toponym gazetteers using mobile and online apps. The user–gazetteer interaction metrics discussed in **Section 4** were examined using analysis of user interaction when accessing the "Toponyms Gazetteer" website (https://mapearth.co.ke). **Section 5** discusses the advantages and limitations of using toponym ontologies in building ontologies for Gazetteers and related domains.

2. Gazetteer ontologies and entities overview

The use of ontologies in gazetteers has been a research topic in recent years. A review of the Alexandria Digital Library, GeoNames, and Getty Thesaurus databases revealed difficulties in mapping feature-type thesauri to ontologies or gazetteers (Janowicz & Keßler, 2008). Therefore, existing web gazetteers do not provide optimal support for the use of standards to regulate open standards (Frederico T. et al., 2002) for querying and proposed the use of distributed feature type schema to leverage the full functionality of gazetteers. A typical example of a linkable ontology gazetteer is the Portuguese Geo-Net PT 02 ontology (Lopez-Pellicer et al., 2010). The Geo-Net PT O2 ontology has names, types, relations, and footprints as entities. The recently updated Yet Another Great Ontology (YAGO) (Suchanek et al., 2007; Pellissier Tanon et al., 2020), draws its data from Wikipedia and WordNet in addition to GeoNames. In the linked open data cloud, YAGO is connected to DBpedia and Freebase, although Freebase has ceased its operations and was acquired by Google in 2016 (Chah, 2018). WordNet is a lexical database that uses computational linguistics and natural language processing to express conceptual relationships (Miller, 1995) such as synonymy, hyperonymy, hyponymy, antonymy, meronymy and other relationships for related place names.

Probing the ontology characteristics provides information about the usage, the category of users, the type of services offered, and the entities that are accessible in a particular gazetteer. An overview of ontology gazetteer characteristics fully reveals the usage, category of users, type of service and entities accessed for free (as shown in Table 1) using reasoners such as 'HermiT' in Protégé or Fluent Editor's 'Who-Or-What' to describe the logics in ontologies. An overview of selected gazetteer and ontology projects that were started as open-source projects (Ballatore et al., 2013).

A review of some open-source ontology studies revealed that the ontologies had potential, but dwindling returns, outside support to ensure sustainability, and shrinking funding making some projects cease operations or be taken over by others, such as Freebase (Chah, 2018) acquired by Google in 2016 (which offers full-text search). Others, such as the OpenCyc project, which began in 2002 (Matuszek et al., 2006), ceased operations in 2017 and shifted their basic license distribution model to commercial Cyc in 2017. Some of the knowledge bases contain the basics of ontology while others are semantic networks. OpenCyc was previously offered as an API, and a data dump under the Apache and Creative Commons open-source licenses. The API helps two applications to connect. Currently, available distribution for most APIs is a research license or a commercial license which gives access to Cyc's latest technology.

The representations of feature types in most gazetteers differ owing to differing subjects in the feature type ontology mappings, coverage, and fundamental characteristics of the gazetteers utilized in each domain area. To solve this difficulty, geographical signatures are proposed (Qian et al., 2016; Zhu et al., 2016) using top-down ontology analysis and bottom-up data-driven approaches.

Table 1 displays common ontologies from a global perspective, whereas Table 2 is for states or nations, with the themes or attributes covered in online resources in both tables. The feature types mapped on web platforms may vary depending on the type of feature object being mapped and the distribution licenses. WordNet, which started in 1985, is incomplete for some spoken languages or countries and has some access restrictions. It is a semantic network dictionary thesaurus.

Table 1. Gazetteer ontologies services and entities

Ontology type	Year	(active/ inactive)	Category of users	Data format	authors	Content	General entities
GeoNames	1994	Active	World	OWL/ RDF	GeoNames	6,774 axioms, 663 feature codes; in 9 properties, 19 classes, 702 class assertions	Place name attributes classes for whole world
Geographic Ontology	2012	Active	World	OWL/ RDF	BCN geographies	6,914 axioms, 32 classes, 21 objects, 19 data.702 assertions	Electoral entities, place, region, division, municipality, commune, country
Thesaurus Geographical names	2014	Active	World	meta/OWL	Getty Institute	4,579 axioms, 33 classes, 348 objects, 10 data and 82 class assertions	Name, place type, nation
DBpedia	2008	Active	World	OWL/ RDF	DBpedia	1 billion datasets, 62 million links, 50,818 axioms, 1,617 classes, 1,251 object property and 1,957 data properties, 1 class assertion	Persons, places, works, organizations, species, plants and diseases
ConceptNet	1999	Active	World	JSON-LD API, semantic network	MIT media lab	1.6M+ assertions, 700k natural language sentences	Words and phrases and the common-sense relationships as used
OpenCyc	1984	Active/ Cyc	World	RDF/ OWL	(Matuszek et al., 2006)	1.5 million terms, 416k collections, 42.5 predicates, 1M+ entities, 239k concepts, 60k classes, 2.1 million facts	Database, an inference engine for data analysis and modification, relations, attributes, fields, properties, functions
YAGO	2006	Active	World	RDF	(Pellissier Tanon et al., 2020)	2B type-consistent triples/facts, 64M+ entities	People, cities, countries, movies, and organizations

Ontology type	Year	(active/ inactive)	Category of users	Data format	authors	Content	General entities
GEONet/ GeoNet Names Server(GNS)	2006	Active	World (except US)	OWL/XML	(Salayandia et al., 2006)	Gravity, aeromagnetic, electromagnetic, seismic reflection, broadband seismic, seismicity/earthquake, heat flow, remote sensing (multispectral), drill hole data, geologic maps, faults, geochemistry/petrology, crustal stress and strain, digital elevation models, and remote sensing	Primary administrative division codes and names, geopolitical entity codes and names, feature designation (feature type) codes and names, etc
OpenStreetMap (OSMonto[1]) tags	2004	Active	World	XML/ OWL	Mihai Codescu, Gregor Horsinka, Oliver Kutz, Till Mossakowski and Rafaela Rau	Uses wordNet vocabulary	Address, a location or a route composed of an origin, a destination and possibly intermediary activities
OSM semantic network	2013	Active	World	XML	OSM	1,217 axioms, 466 classes, 29 Objects and 27 data with no class assertions.	LOD, WordNet, LinkedGeoData, TagInfo, OSM vector data
DataTypes[2]	2014	Active	World	JSON/RDF	*QUDT.org	11,188 axioms, 661 classes and 1,189 class assertions	Standards to quantify data expressed in RDF and JSON

*Quantities, Units, Dimensions and Types (QUDT) schema data type ontology deals with scalar types such as numbers, characters and Boolean and structured data types of arrays, lists, trees, and others including coordinate systems.

[1] https://wiki.openstreetmap.org/wiki/OSMonto
[2] http://qudt.org/

Table 2. Gazetteer ontologies services and entities

Ontology	Usage interface functionalities	Coverage	File Type, Nature of service	Authors	Content	General entities
GeoFabrik[3]	Drainage features	Australia	RDF/ turtle	Nicholas Car CSIRO	1,098 axioms, 86 classes, 43 object and 10 data property; no class assertions	Catchment, river drainage, division, and region
**Ordnance 50k	Topographic mapping	UK	OWL,	British Ordnance survey	Gazetteer/ontology 13 classes, 2 properties, 6 data objects	Roads, footpaths, woods, buildings, rivers, and streams
GeoNet PT02	Administrative features and place names	Portugal	URIs, RDF and OWL	University of Lisbon, (Lopez-Pellicer et al., 2010)	Uses GeoNet vocabulary, 701k concepts defined using 81 feature types	Name, type, relationships and footprint
OS Open names[4]	Roads, settlements, postcode's locations	Great Britain	CSV, GeoPackage, GML and API	Ordinance Survey	2.5 million locations	Place names, postcodes, road names and road numbers
SwissNames 3D[5&6]	2010	Switzerland and Principality of Liechtenstein	OWL	ESRI GD, Shapefile and CSV	400k geographical names mapped in 61 classes and 70 million 3D objects	Transport infrastructure, sports facilities, field and local names for buildings, houses, lakes, sites, terrain features of landscape and regions
Geographic Names Information System[7](GNIS)	1994	US	Database	USGS (United States Geological Survey)	2 million geographic names	Populated places, schools, lakes, streams, valleys, and ridges.

[3] https://github.com/CSIRO-enviro-informatics/geofabric-ont/
[4] https://www.ordnancesurvey.co.uk/business-government/products/open-map-names
[5] http://www.swisstopo.ch/
[6] https://www.swisstopo.admin.ch/en/maps-data-online/maps-geodata-online/3d-viewer.html
[7] https://geonames.nga.mil/gns/html/

Ontology	Usage interface functionalities	Coverage	File Type, Nature of service	Authors	Content	General entities
***GNIS-LD	2018	US	Turtle	(Regalia et al., 2018)	37 million, triples, 6.7 million named subject nodes of GNIS-id, 2.27 GNIS features, 2 million same as relations to GeoNames, 503k same as relations to DBpedia, 494.7 alternate names, 66 feature types	Church, stream, school, populated places, locale, building, cemetery, reservoir and summit etc.
Surface water[8]	Surface water	US	OWL	USGS	1,228 axioms, 93 lasses, 27 object and 14 data counts	Boundary, landform, named point, structure, surface water and transportation
An ontology of German place names[9]	2005	Germany	OWL	(Nagel, 2005)	17k entries, 21k lemmata, 18k toponyms, 1k adjectives, 1kn male and female and 1.2k classifiers)	Linguistic features of place names, associates place names to geographic entities and relationships of entities
Clinga[10] (Chinese linked geographical dataset)	2002/Active	China	OWL	Nanjing University	137 classes, 1,578 axioms, 54 objects	Geographical entities in China (e.g., cities, mountains) and their relations (e.g., capital-of)

** In addition to the 50k Gazetteer of the United Kingdom ontology, there are subsets of ontologies such as Postal Code, Administrative Geography and Civil Voting area, Geometric and Spatial Linkages, all of which function in tandem as similar outputs of the Ordinance Survey ontologies.
*** GNIS feature classes for domestic names are no longer available online but they are archived. The feature types affected are Airport, Bridge, Building, Cemetery, Church, Dam, Forest, Harbor, Hospital, Mine, Oilfield, Park, Post Office, Reserve, School, Tower, Trail, Tunnel, and Well.

8 https://www.usgs.gov/media/files/surface-water-ontology
9 https://journals.openedition.org/corela/1178?lang=en#tocto1n3, https://doi.org/10.4000/corela.1178
10 http://openkg.cn/dataset/clinga

3. Materials and methods

The data collection and pre-processing were presented using pre-developed tools of a mobile application and a web application to demonstrate access and usage of the toponym gazetteer. The GeoNames, DIVA-GIS, and OpenStreetMap (OSM) data dumps available for download were initially obtained from the hosting website before being hosted on a pre-developed website as an open-source platform for publishing geographical names (Nyangweso & Gede, 2021). Shell scripts and command-line tools running on an internet Linode server were used to parse the data dumps. The server specs included 1 CPU core, 25 GB storage, Secure Shell (SSH) protocol access, and 1GB RAM. The data collection and data operation rules were packaged using the Django framework, which is based on Python, as a publicly accessible gazetteer at https://mapearth.co.ke and use analytics tool. The goal of employing a Google analytics tool was to evaluate the ontology's usage experiences based on the use of the gazetteer service and user category, in terms of resources utilized to access, nature, and users' ease of comprehension of the information given or extracted from gazetteers. Other open-source technologies were utilized to combine data collection, storage, and visualization, as well as to enable continuous integration and development of its services.

To assess the usage experiences, we first adopted an ontology from GeoNames, and then we selected the important entities to accommodate essential attribute fields for characterizing the gazetteer. Protégé, WebVowl and Ontorion Fluent Editor were the three ontology editors used to edit and update the toponym ontology after export to include links and information associated with the toponym gazetteer using a pre-developed web-map and mobile apps that were used to build two sets of data tables.

The data tables for the web and a mobile application were incorporated into the PostgreSQL database as affiliation tables, which users may query, search, or utilize for place name relation mapping.

The evaluation criteria for user interactions with toponym gazetteer entities included the use of elements such as the country it was accessed from, the language of engagement, the user acquisition, the platform of acquisition or engagement, and the browser used in conjunction with the operating system. Due to the limited scope of the geo-parsed place-names investigation, other user preferences such as time and pages viewed were not included in the evaluation.

3.1. Methodology flow diagram

The methodology shown in Figure 1 incorporates volunteered geographic information (VGI) data assembly, ontology linking and relation modeling to enable geo-parsing of the data dumps to enable visualizing the place names on a map, export data into XLS, CSV, ODS, JSON and GeoJSON data formats for expression toponym footprints. The user interaction of the gazetteer was assessed using Google Analytics tool. Tools incorporated in the web application include the addition of new place names to the gazetteer and visualizing of the added place names for approval through the administration site of the website.

In this context, data dump geo-parsing entails computer-aided geographical grounding of the place-name language and automatically correlating place-name texts with coordinate location tags.

3.2. Toponyms ontology evaluation and testing

Our article also assesses the performance of a gazetteer among users by tying it to the Google Analytics tool via the website https://mapearth.co.ke, as well as providing insight into the present state of the ontology discussion on a global and country-by-country basis. The data was tested on performance in previewing the contents in the toponym gazetteer, export, query and search data on a database. A mobile and online map application may also be used to acquire fresh data while stationed away from the feature itself.

Each data point in the database has a Uniform Resource Locator (URL) link to open its geospatial location though it is not connected to the linked open data (LOD) cloud due to limited security resources on our test website server. However, the integrated PostgreSQL VGI data dumps and new data generated from toponym mobile app and web application platforms are accessed through SSH scripts.

The used toponyms in a gazetteer can be categorized based on one-to-one, one-to-many and many-to-many (for related places) to perform relation mappings (Chen et al., 2017).

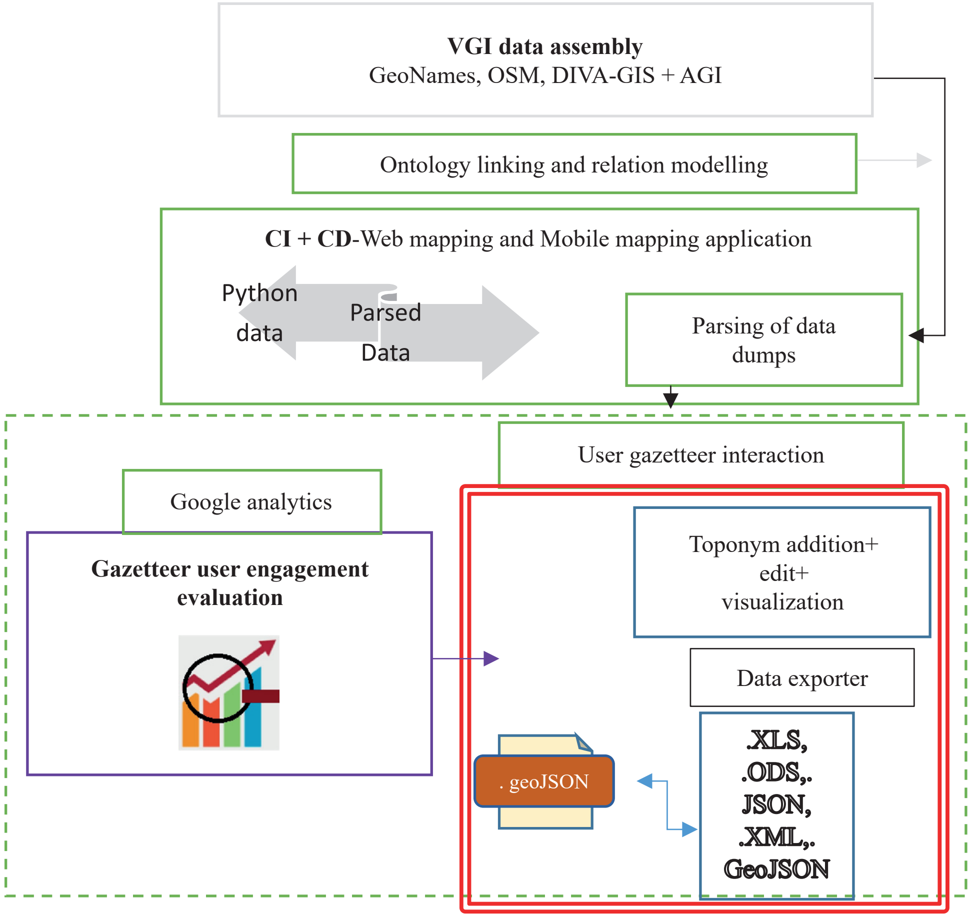

Figure 1. Methodology flow diagram

Source: Nyangweso & Gede, 2021, p. 26.

Table 3. The affiliations for relations in ontology structures and elements, as well as the required contents of the toponym gazetteer affiliation table, as used in characterizing types of the entities relations for Ontogazetteer (Machado et al. 2011)

Ontology component	Definition	Equivalent in Gazetteer	Example
Concept	The meaning of a term	Place	Tana (a term that is related to another place in the world)
Synonymy relationship	The meanings of the two terms A and B are nearly identical	Abbreviations historical name, alternative names and nicknames, acronyms, language differences, spelling differences	Nyansiongo-Kijauri, Kisii-Bosongo (Local names) North Horr (NH), Chepilat/Chebilate – language variation
Homonymy relationship	Term A is written precisely the same as term B, yet their meanings are different	Places (of the same or distinct categories) with coinciding names, places whose names coincide with other things are referred to as ambiguous names	Kisii town and tribe (names of more than one thing), Makongeni and Maji Mazuri, Thika road and Thika Town Mwembe (ambiguous names), Siriba personal name and town
Hypernymy relationship	A has a broader definition than term B	A location at a higher tier of a territorial hierarchy	Kenya and its 47 counties
Hyponymy relationship	Term A has a more limited definition than term B	A location at a higher tier of a territorial hierarchy	Kasarani estate neighborhoods and Nairobi city
Association relationship	Term A is linked to term B, implying that they have a semantic relationship	Semantically related places	Cities along a route; maize-producing counties; historical places; Magadi and Tabaka soapstone regions

Relations for the different elements of the gazetteer were represented using arrow charts connecting the domain and range elements. A mathematical function is a set of ordered pairs such as {(0, 1), (10, 24), (11, 16)} where coordinates for locations can be a function for longitude and latitude. In mathematics, the difference between a function and a relation is that a function has just one y-value for each x value unlike the ontology components (Table 3), where there is supertype and subtype such as for the case of hypernym is higher than hyponym in the category rank of semantic relations.

Machine learning and modeling are used in hypothetical quasi-natural-language queries about spatial relations between line features and areal features to define strengths of possible spatial relations, including testing and ranking in the strength of agreement of relations in the ranges of levels 1 to

5 or percentages of 1 to 100 (Mark & Egenhofer, 1994). The study levels of one to five were selected for toponyms exhibiting the relations for illustration from the compiled database.

3.3. Toponym gazetteer entities, attributes, and relations

The database of the data captured was used to generate an affiliation table for each entity, either as new data or from volunteer sources with each having its own Uniform Resource Identifier (URI) link and a unique ID.

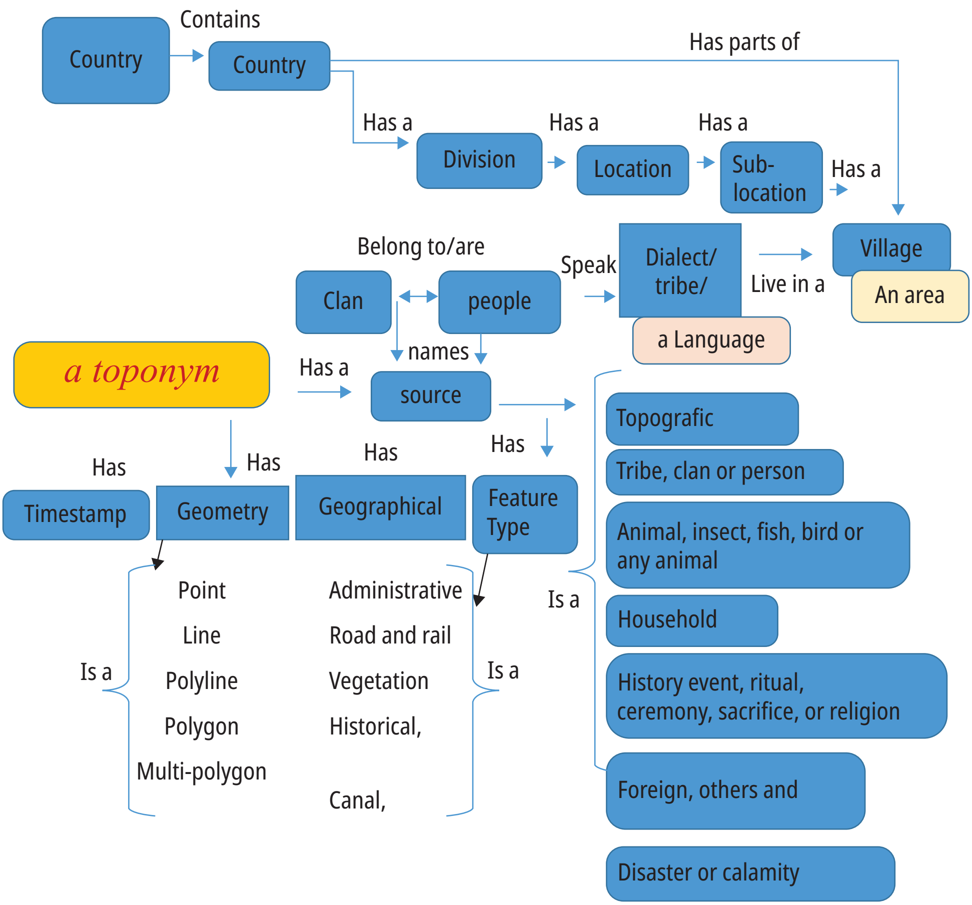

Figure 2. Gazetteer entity relationships

Source: own work.

Shapes defined by a point, line, polygon and mult-polygon indicate features in gazetteers which are classified into administrative, road and rail, vegetation, historical tourist and sport, and canal or water feature types. The said feature types are identified or described by toponyms given using tribe, clan or personal classification identities. These descriptions form classes bearing topographic, tribe, clan, person, animal, insect, fish, household, historical events or foreign identities (Figure 2). All these entities are contained in a country (within a specific hierarchy of administration classes) to express relatedness and disjointedness in web gazetteer ontologies in the features (identified by given names) in a particular gazetteer entity relations service, using ontologies. The entity class of tribes and clans were modelled to be attributed to classify the toponyms' origin and assign the source entities of the already named spaces as per the geographical location shown by the open-source gazetteer data using the bounded relations contained in the country attributes and the toponym mapped features (Figure 2), using specifically identified languages.

3.4. The schema Relationships of toponym ontology

Spatial schema = (R, A, D, T, attr, pkey, fkey, type)

Where, *R* = relationships, *A* = attributes, *D* = dependency of inclusion, *T* = data types, *attr* = property category or class, *pkey* = primary key, *fkey* = foreignkey, *type* = all dataset functions and attributes.

The schema relation aids in the explanation of database contents, allowing ontology reasoners to interpret the structure of the attributes described uniquely, eliminating logical contradictions.

3.5. Relationships: Toponym gazetteer

Figure 2 shows the relations used in the context of the paper where an ethnic group is composed of a tribe or dialect. The tribe has clans. The clans are composed of people who have names. The clan has a name and gives or provides toponyms of places. The clan has relations to others. The toponyms are entities of geometric objects having a geographical location, historical association or derivation, time when it came up and there are reasons for its continued use or change as attributes. An entity has an assertion or proof of its use or existence uniquely, by a known group within a region or area. There can be differences by which different languages express spatial concepts of space using language boundaries and relations (Talmy, 1983), hence the need to include representative typical examples of affected toponyms for illustration of the relations.

4. User engagement assessment on the toponym gazetteer and discussion

Acquisition is the means through which visitors access websites or how viewers are acquired or referred to download an app as shown in Figure 5. Put simply, it is the process of how people visit the website or download mobile apps. Figures 3–8 depict user engagement perspectives on usability aspects on the toponym gazetteer courtesy of the Google Analytics tool from January 1, 2020, to September 26, 2021. Figure 3 shows the frequency of gazetteer engagement per country.

Many people accessed the gazetteer from Kenya, as seen in Figure 3 due to simulations done. Different users from other nations may have been drawn to the gazetteer through organic acquisition. Apart from website and search engine personalization, many users were drawn from Kenya due to its data being relevant to them. Also, this may be attributed to the users being Kenyans abroad or tourists searching for the data from outside the country; the open, nonlocalized state of the website allows open free usage.

Figure 4 shows that English is the most preferred language in accessing the gazetteer. The highest uptake of English is attributed to English being the original language of the computer and the internet. Most devices come with English as the default language before being translated to other languages. In this case, most users access the geographic data using the English language by default.

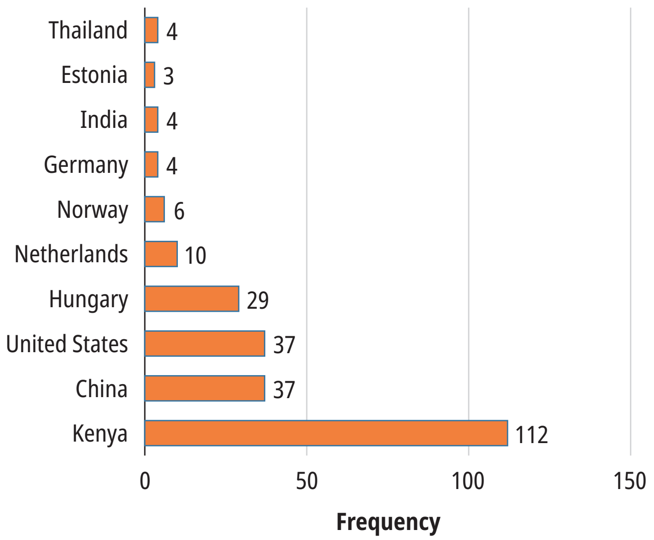

Figure 3. Frequency of access by country

Source: Google Analytics of Toponyms Gazetteer (https://mapearth.co.ke/).

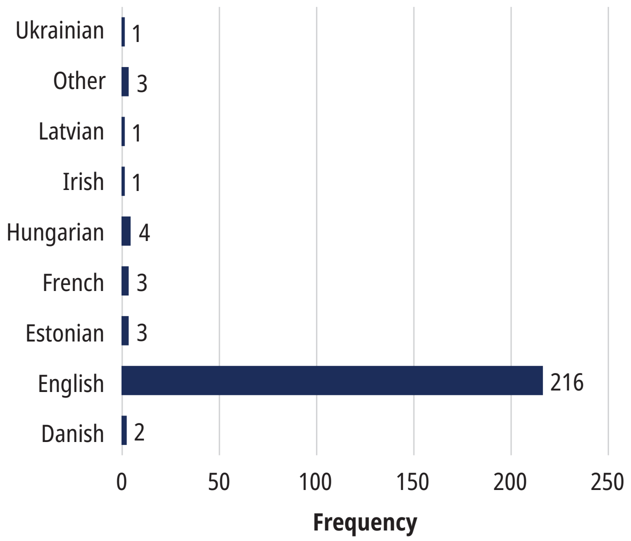

Figure 4. Language of engagement frequencies of a gazetteer

Source: Google Analytics of Toponyms Gazetteer (https://mapearth.co.ke/).

The highest acquisition was from new users followed by organic and referrals as indicated in Figure 5. Since most people search for new data, new users were attracted to use the web gazetteer service to access its open-source services.

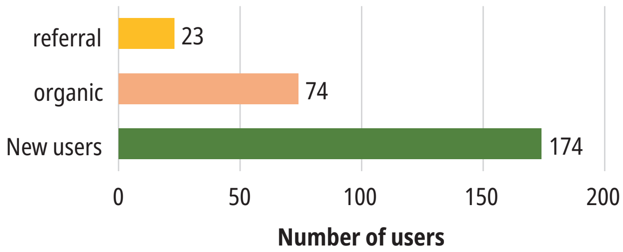

Figure 5. User acquisition for gazetteer

Source: Google Analytics of Toponyms Gazetteer (https://mapearth.co.ke/).

Users were able to access the gazetteer through tablets, mobile phone and desktop as indicated in Figure 6. Most users preferred to use PCs which have large screens and thus the capability to view the web gazetteers with clear zooms; this was followed closely by mobile phones whose advantage is portability. In this case, there was a trade-off between large area visualization and portability – large screens allow for the access of more details.

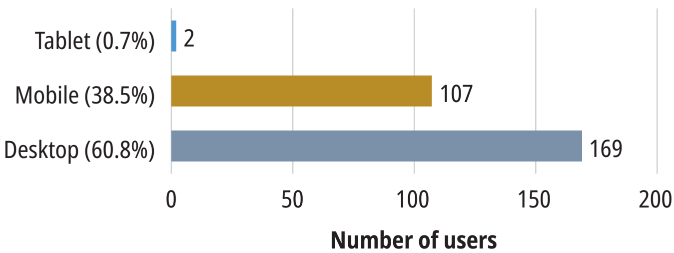

Figure 6. Platform of engagement

Source: Google Analytics of Toponyms Gazetteer (https://mapearth.co.ke/).

Figure 7 shows that Google Chrome browser is the most used browser followed by UI browser, Firefox, Safari, Edge, and Samsung in that order. The high usage of Chrome is attributed to the fact that in most devices Chrome comes preinstalled in Android OS, it is widely used in the viewing and access of Google products (such as Google Play) and the overall trust the general public have in it.

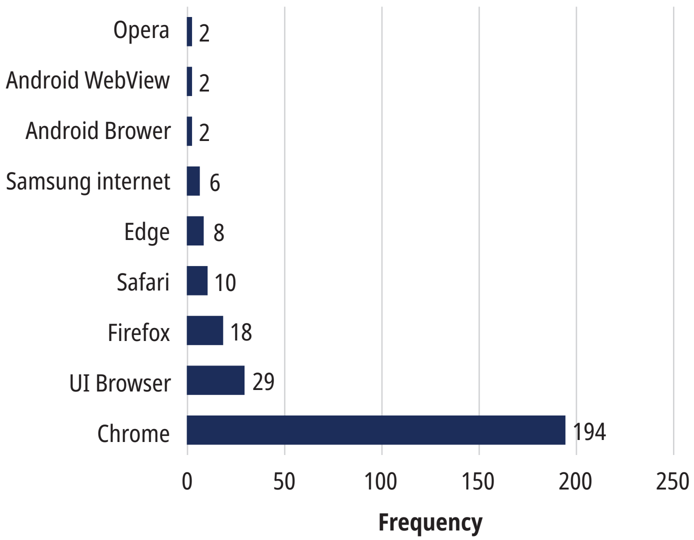

Figure 7. Type of browser used in accessing gazetteer

Source: Google Analytics of Toponyms Gazetteer (https://mapearth.co.ke/).

Figure 8 demonstrates that Google's Android OS is the most popularly used OS in mobile apps, followed by Windows and Linux. The high usage is attributed to devices with Android OS being mostly accessible to a wider audience due to their ease of usage, cheap acquisition and maintenance, and also wide access to free and open-source apps, unlike the other OS's.

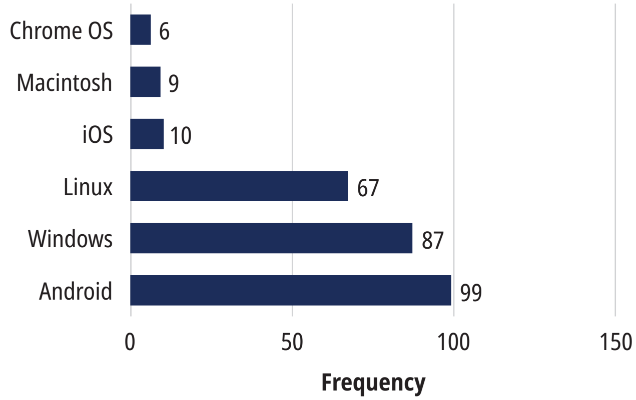

Figure 8. Operating system

Source: Google Analytics of Toponyms Gazetteer (https://mapearth.co.ke/).

5. Discussion

The best practices of a gazetteer service may be achieved by implementing open systems, which do not require much effort, skill, economic cost, and limitation on the access of resources. This is indicated by Google's free products built on Free and Open-Source Software (FOSS) distributions. Top of the list of best practices is the use of open standards such as those from Open Geospatial Consortium (OGC) which include the use of Web Feature Service (WFS) standards on gazetteers. Also, Geospatial Fusion Service (WFS-G) specifically implemented for GeoNames before making it an OGC best practices document (05-035R2) makes the service accessible to many. The WFS-G standard allows a client to search and retrieve georeferenced entities.

The use of toponymical ontologies helps to connect to ubiquitous data frameworks, update gazetteers on heritage and governance, and integrate open data via LOD, DBpedia, QUDT and Wiki sites. However, there are shortcomings of using ontologies such as linking several knowledge graphs, complexity in making comparisons of metadata (especially in the inclusion of graphic entities) and managing updates. Despite the availability of Geospatial Data

Abstraction Library (GDAL) standards for raster and vector data formats, there is no compliance enforced; this is because of the use of geographically connected fields but open and fragmented standards. GeoNames, for example, includes 663 characteristics in 9 classes (Maltese & Farazi, 2013) while OSM data has its geographic information for features packaged individually as tags, ordered as a list of ways defining lines, points and relations for all entity elements (Patriarca et al., 2019). It also includes land use and land cover data.

The development of a smarter, spatially linked ontology faces challenges, possibly due to the unbounded requirements and attributes of different databases from different sources. The ontologies are heterogeneous in terms of feature types and schemas, due to advanced technology deployment for ontology usage in open web gazetteers as shown by Tables 1 and 2. The advanced gazetteers have integrated interactive features such as Stonly.[11] In some cases, some have lean, simple and clean taxonomy of schema (Guha et al., 2016) and others have matured into sustainable projects. Sadly, others have disbanded, such as OpenCyc, which was active from 1984 until the end of 2017 (Chah, 2018).

6. Conclusion

The goal of this paper was to assess toponym gazetteer entities' usages worldwide and toponym gazetteer ontology performance. It was found out that there are 1,433 ontologies as of 7th November 2021 as per the DBpedia archive[12] compilation, where some are not updated. As a result, coordination between DBpedia and the ontology generators is essential. It is vital to opt for an all-inclusive ontology that is interconnected with diverse attribute entities after attribute analysis for selection and application (Mukhtar et al., 2013; Sikelis et al., 2021). However, most evaluation tools lack long-term support, which is a key issue for their sustainability. The difficulty is exacerbated by the deployment and use of the connected open big data cloud, a new concept with multiple

[11] https://stonly.com
[12] https://archivo.dbpedia.org/list

variations in incorporating different knowledge bases with one another. Mostly these knowledge bases utilize open compliant standards other than those sanctioned by OGC and International Organization for Standardization (ISO) Systems (Sikelis et al., 2021).

Expert knowledge on accessing the toponymic gazetteer does not differ semantically on shared products so long as the data is distributed in an open standard where anybody can access the resource. The user engagements study indicated the existing potential of ontologies usage in ubiquitous linking of toponymic data, drawing relations on toponyms using existing open API frameworks and possibilities for updating gazetteers seamlessly. The results are twofold since data gathered from usages can be used in propping governance of heritage and development through incorporation and integration of open data standards through LOD cloud, DBpedia and Wiki sites, which have been open knowledge bases since their commencement in the 1980s as corroborative projects. Furthermore, based on the language of access, some inequities exist on language which varies among countries with different backgrounds of languages and knowledge. The results of our study can be used to probe why users in different parts of the world access the web gazetteer through its interface.

The review limitation is that only a few global or regional gazetteers were evaluated while there are so many ontologies that can be evaluated for one to make an informed decision on the best approaches to take. Future analysis should include other reasoners to test their efficiency.

Abbreviations

AGI – Authoritative Geographic Information

API – Application Programming Interface

CD – Continuous Development

CI – Continuous Integration

CPU – Central Processing Unit

FOSS – Free and Open-Source Software

GDAL – Geospatial Data Abstraction Library

GNIS – Geographical Names Information System

JSON – Java Script Object Notation
LOD – Linked Open Data
ODS – Open Data Spreadsheet
OGC – Open Geospatial Consortium
OS – Operating System
OSM – OpenStreetMap
OWL – Ontology Web Language
QUDT – Quantities, Units, Dimensions and Types
RDF – Resource Description Frameworks
SSH – Secure Shell
URI – Uniform Resource Identifier
URL – Uniform Resource Locator
VGI – Volunteered Geographic Information
WFS – Web Feature Service
WFS-G – Geospatial Fusion Service
XML – Extensible Mark-up Language
YAGO – Yet Another Great Ontology

References

Abdelmoty, A. I., Smart, P., & Jones, C. B. (2007). Building Place Ontologies for the Semantic Web : Issues and Approaches. In R. Purves & C. B. Jones (Eds.), *Proceedings of the 4th ACM Workshop On Geographic Information Retrieval, GIR 2007, Lisbon, Portugal, November 9, 2007* (pp. 7–12). https://doi.org/10.1145/1316948.1316951

Acheson, E., Purves, R., & De Sabbata, S. (2017). A quantitative analysis of global gazetteers: Patterns of coverage for common feature types. *Computers, Environment and Urban Systems, 64*, 309–320. https://doi.org/10.1016/j.compenvurbsys.2017.03.007

Ballatore, A., Wilson, D. C., & Bertolotto, M. (2013). A survey of volunteered open geo-knowledge bases in the semantic web. In G. Pasi, G. Bordogna, & L. C. Jain (Eds.), *Quality Issues in the Management of Web Information* (pp. 93–120). Springer. https://doi.org/10.1007/978-3-642-37688-7_5

Berners-Lee, T. (2010). *5-Star Open Data*. 5-Star Linked Open Data. https://5stardata.info/en/

Chah, N. (2018). *OK Google, what is your ontology? Or: Exploring freebase classification to understand Google's knowledge graph*. https://doi.org/10.48550/arXiv.1805.03885

Chen, H., Vasardani, M., & Winter, S. (2017). *Geo-referencing Places From Everyday Natural Language Place Descriptions*. https://arxiv.org/abs/1710.03346

Frederico T., F., Max J., E., Peggy, A., & Gilberto, C. (2002). Using ontologies for integrated geographic information systems. *Transactions in GIS*, *6*(3), 231–257. https://doi.org/10.1111/1467-9671.00109

Goodchild, M. F., & Hill, L. L. (2008). Introduction to digital gazetteer research. *International Journal of Geographical Information Science*, *22*(10), 1039–1044. https://doi.org/10.1080/13658810701850497

Guarino, N. (1997). Understanding, building and using ontologies. *International Journal of Human-Computer Studies*, *46*(2–3), 293–310. https://doi.org/10.1006/ijhc.1996.0091

Guha, R. V., Brickley, D., & Macbeth, S. (2016). Schema.org: Evolution of structured data on the web. *Communications of the ACM*, *59*(2), 44–51. https://doi.org/10.1145/2844544

Hill, L. L. (2009). Gazetteers. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of Database Systems* (pp. 1217–1218). Boston: Springer. https://doi.org/10.1007/978-0-387-39940-9_174

Hu, H., Ge, Y., & Hou, D. (2014). Using web crawler technology for geo-events analysis: A case study of the Huangyan Island incident. *Sustainability*, *6*(4), 1896–1912. https://doi.org/10.3390/su6041896

Janowicz, K., & Keßler, C. (2008). The role of ontology in improving gazetteer interaction. *International Journal of Geographical Information Science*, *22*(10), 1129–1157. https://doi.org/10.1080/13658810701851461

Lopez-Pellicer, F. J., Silva, M. J., & Chaves, M. (2010). Linkable geographic ontologies. In R. Purves, P. Clough, & C. B. Jones (Eds.), *Proceedings of the 6th Workshop on Geographic Information Retrieval, GIR 2010, Zurich, Switzerland, February 18-19, 2010*. https://doi.org/10.1145/1722080.1722082

Machado, I. M. R., de Alencar, R. O., Campos, R. de O., & Davis, C. A. (2011). An ontological gazetteer and its application for place name disambiguation in text. *Journal of the Brazilian Computer Society*, *17*(4), 267–279. https://doi.org/10.1007/s13173-011-0044-4

Maltese, V., & Farazi, F. (2013). A semantic schema for GeoNames. (Technical Report # DISI-13-004). http://eprints.biblio.unitn.it/4088/1/techRep004.pdf

Mark, D. M., & Egenhofer, M. J. (1994). Modeling spatial relations between lines and regions: Combining formal mathematical models and human subjects testing. *Cartography and Geographic Information Systems*, *21*(4), 195–212. https://www.tandfonline.com/doi/abs/10.1559/152304094782540637

Matuszek, C., Cabrai, J., Witbrock, M., & DeOliveira, J. (2006). An introduction to the syntax and content of Cyc. In *AAAI Spring Symposium – Technical Report*, *SS-06-05* (pp. 44–49). https://aaai.org/papers/0007-an-introduction-to-the-syntax-and-content-of-cyc/

Mechouche, A., Abadie, N., Prouteau, E., & Mustière, S. (2013). Ontology-based discovering of geographic databases content. In A. Ruas (Ed.), *Advances in Cartography and GIScience: Vol. 1. Selection from ICC 2011 Paris* (pp. 311–330). Springer. https://doi.org/10.1007/978-3-642-19143-5_18

Miller, G. A. (1995). WordNet: An electronic lexical database. *Communications of the ACM*, *38*(11), 39–41. https://doi.org/10.1145/219717.219748

Mukhtar, N., Shahzad, S., Khan, M. A., & Nazir, S. (2013). Ontology for feature based selection of web development tools. In *Eighth International Conference on Digital Information Management, ICDIM 2013* (pp. 90–95). https://doi.org/10.1109/ICDIM.2013.6693980

Nagel, S. (2005). An ontology of German place names. *Corela*, *HS-2*. https://doi.org/10.4000/corela.1178

Nyangweso, D., & Gede, M. (2021). An open-source framework for publishing geographical names – A case study of Kenya. *Geodézia És Kartográfia*, *2*, 24–30. https://doi.org/10.30921/GK.73.2021.2.4

Patriarca, J., Fonte, C. C., Estima, J., Almeida de, J.-P., & Cardoso, A. (2019). Automatic conversion of OSM data into LULC maps: Comparing FOSS4G based approaches towards an enhanced performance. *Open Geospatial Data, Software and Standards*, *4*, Article 11. https://doi.org/10.1186/s40965-019-0070-2

Pellissier Tanon, T., Weikum, G., & Suchanek, F. (2020). YAGO 4: A reasonable knowledge base. In A. Harth, S. Kirrane, A.-C. N. Ngomo, P. Heiko, A. Rula, A. L. Gentile, M. Haase, & M. Cochez (Eds.), *The Semantic Web. ESWC 2020. Lecture Notes in Computer Science* (Vol. 12123, pp. 583–596). Cham: Springer. https://doi.org/10.1007/978-3-030-49461-2_34

Qian, S., Kang, M., & Wang, M. (2016). An analysis of spatial patterns of toponyms in Guangdong, China. *Journal of Cultural Geography*, *33*(2), 161–180. https://doi.org/10.1080/08873631.2016.1138795

Rada, R., Mili, H., Bicknell, E., & Blettner, M. (1989). Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man and Cybernetics*, *19*(1), 17–30. https://doi.org/10.1109/21.24528

Regalia, B., Janowicz, K., Mai, G., Varanka, D., & Usery, E. L. (2018). GNIS-LD: Serving and visualizing the geographic names information system gazetteer as linked data. In A. Gangemi et al. (Eds.), *The Semantic Web: ESWC 2018. Proceedings* (pp. 528–540). https://doi.org/10.1007/978-3-319-93417-4_34

Salayandia, L., Huang, Y., Gates, A. Q., & Roach, S. (2006). GeoNet: Use of grid technologies in geoinformatics for the transition zone between the Colorado Plateau and the Basin and Range province. In A. K. Sinha (Ed.), *Geoinformatics: Data to Knowlegde* (pp. 183–193). https://doi.org/10.1130/2006.2397(13)

Sikelis, K., Tsekouras, G. E., & Kotis, K. (2021). Ontology-based feature selection: A survey. *Future Internet*, *13*(6). https://doi.org/10.3390/fi13060158

Suchanek, F. M., Kasneci, G., & Weikum, G. (2007). Yago: A core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web, WWW 2007* (pp. 697–706). ACM. https://doi.org/10.1145/1242572.1242667

Talmy, L. (1983). How language structures space. In H. Pick & L. Acredolo (Eds.), *Spatial Orientation: Theory, Research and Application* (pp. 225–282). Boston: Springer. https://doi.org/10.1007/978-1-4615-9325-6_11

Watt, A., & Eng, N. (2014). *Database Design* (2nd ed.). Victoria, B.C.: BCcampus. https://opentextbc.ca/dbdesign01/

Zhu, R., Hu, Y., Janowicz, K., & McKenzie, G. (2016). Spatial signatures for geographic feature types: examining gazetteer ontologies using spatial statistics. *Transactions in GIS*, *20*(3), 333–355. https://doi.org/10.1111/tgis.12232