

Towards an Ontology for Toponyms in Nepalese Historical Documents

Sabine Tittel

Heidelberg Academy of Sciences and Humanities

Heidelberg

sabine.tittel@hadw-bw.de

Abstract

Nepalese historical legal documents contain a plethora of valuable information on the history of what is today Nepal. An empirical study based on such documents enables a deep understanding of religion and ritual, legal practice, rulership, and many other aspects of the society through time. The aim of the research project ‘Documents on the History of Religion and Law of Pre-modern Nepal’ is to make accessible a text corpus with 18th to 20th century documents both through cataloging and digital text editions, building a database called Documenta Nepalica. However, the lack of interoperability with other resources hampers its seamless integration into broader research contexts. To address this problem, we target the modeling of the Documenta Nepalica as Linked Data. This paper presents one module of this larger endeavour: It describes a proof of concept for an ontology for Nepalese toponyms that provides the means to classify toponyms attested in the documents and to model their entanglement with other toponyms, persons, events, and time. The ontology integrates and extends standard ontologies and increases interoperability through aligning the ontology individuals to the respective entries of geographic authority files such as GeoNames. Also, we establish a mapping of the individuals to DBpedia entities.

Keywords: Ontology, Nepal, Place Names, Toponymy, Linked Data, Text Edition

1. Introduction

Recent years have witnessed a growing estimation of modeling language data as resources of the Semantic Web. Linguistic resources in increasing numbers are converted into Linked Data (LD), a set of standard practices for representing and interlinking structured data on the web using Resource Description Framework (RDF). Ontologies of numerous domains provide the necessary structures to formalize the information in the resources and to embed the resources in a cross-discipline, cross-domain and cross-linguistic context. We here describe a proof of concept for the creation of an ontology suitable for the modeling of Nepalese place names and their entanglement with other place names, events, persons, and dates. The ontology, called NEPALPLACES, will be part of a bigger vision (LINKEDOPENNEPAL): to build a set of LD data resources with a text corpus as its focal point. The text corpus is part of the Documenta Nepalica¹ and comprises texts and documents on the history of religion and law of pre-modern Nepal. LINKEDOPENNEPAL will further comprise an ontological model also for person names, and a lexicographic module. NEPALPLACES is the outcome of a collaboration of domain experts of South Asian studies² and of ontology engineering.

The Nepalese language is the official language of

Nepal. In a country with 92 (to 124) different languages (and language varieties, resp.) [in 2001] (Diwasa et al., 2007, 10), it is the mother tongue of approx. 45% of the population of >29MM people [in 2021]³, serving also as a lingua franca (Hutt (1988, 23); van Driem (2001, 1125–1128; 1130f.; 1142)). It is also spoken in north-eastern India, Myanmar, and Bhutan (Riccardi, 2003, 539–541). Through different historical periods, Nepali has developed a large body of literary works (Hutt (1988, 71–76); van Driem (2001, 1136f.)), catalyzed also by the legal code from 1854 (the *Mulukī Ain*, see Khatiwoda et al. (2021)) written entirely in Nepali (Riccardi, 2003, 544). Nepali literature is clearly under-represented in the digital context, in particular with respect to historical language stages. The aim of NEPALPLACES (\subset LINKEDOPENNEPAL) is, thus, to not only increase the visibility and re-usability of the valuable historical Nepali documents but also to establish interoperability with other language resources. This is particularly relevant in the context of South Asian countries and societies that are historically connected to and share geographic, cultural, economic, philological and linguistic aspects with Nepalese society. The challenges of this task lie in extending existing ontological models, in ambiguities within the data to be modeled, linguistic hurdles, complex relations between toponyms and connected information as witnessed by the documents.

The paper is structured as follows: We describe the linguistic resource that provides the data in section 2 and introduce the paradigm of Linked Data, together with related work, in section 3. Section 4 shows the steps

¹Cf. <https://nepalica.hadw-bw.de/nepal/>. All web pages have been accessed 03-17-2022.

²We particularly thank S. Cubelic, M. Grujovska, and A. Zotter for many fruitful and intense discussions during the cooperation, and we also thank M. Bajracharya, R. Khatiwoda, A. Michaels, and Ch. Zotter for their valuable input.—Moreover, we would like to give our thanks to the reviewers for their thorough proofreading and commenting.

³Cf. <https://www.worlddata.info/languages/nepali.php>, <https://knoema.de/atlas/Nepal/Bev%C3%B6lkerung>.

from the linguistic resource towards a Semantic Web resource, with a motivation in section 4.1, the digital status quo in section 4.2, the development of the ontology in section 4.3, and discussions guiding the ontology engineering process as well as ongoing work in section 4.4. We close with a conclusion in section 5.

2. The Linguistic Resource ‘Documenta Nepalica’

The research project ‘Documents on the History of Religion and Law of Pre-modern Nepal’ (Heidelberg Academy of Sciences and Humanities, with research units in Heidelberg, Germany, and Patan, Nepal) makes accessible for the first time a corpus of historical texts from the early Śāha (1769-1846) and Rāṇā (1846-1951) periods. This rich textual material is held by the National Archives of Nepal⁴ and other archives and collections. It consists of temple documents and administrative and legal documents, and it essentially lays the ground of our knowledge about topics still to a large extent unexplored: the history of religious institutions in Nepal, of legal practice in South Asia, the developments entailed by the formation of the Himalayan state, such as the restructuring of social institutions, elite cultures, the legitimization and affirmation of rulership, and the expansion of Hindu rule.⁵

The project’s research results accessible through the Documenta Nepalica consist of a comprehensive catalogue of descriptive metadata of the documents (>65,000 entries) and of scholarly digital text editions (>450) including English translations, comments, notes, and facsimiles (published in open access and partly already with DOIs, in cooperation with Heidelberg University Library)⁶; for an example, see Fig. 1.

The creation of an accompanying, comprehensive bibliography and a glossary growing into a future dictionary is also underway. A newly added task is the identification of named entities within the corpus (person and place names with significant spelling variation) and their organization within a register. The digital research results are groundbreaking in their contribution to text-based empirical studies on the history of Nepal. However, they constitute a typical data silo (Berners-Lee, 2009), i.e., a valuable resource in the World Wide Web (WWW) but with limited access to the catalogue and text editions, not allowing queries beyond the implemented possibilities; the glossary and named entity register are currently not open to the public, the bibliography only partly. Thus, interoperability with related resources is not given. To address this problem, we adopt the paradigm of Linked Data.

⁴Cf. <http://narchives.gov.np/>.

⁵Cf. the project’s website <https://www.hadw-bw.de/Nepal>.

⁶Effective 03-18-2022, cf. <https://nepalica.hadw-bw.de/nepal/>.

3. Linked Data

Over the last years, the paradigm of LD (LOD respectively, with the ‘O’ symbolising open access) has developed into a widespread, powerful means to turn the heterogeneous, unstructured information of the WWW into machine-readable, semantically accessible data (Berners-Lee, 2009; Bizer et al., 2009). As the most solid grounding of the Semantic Web it provides a set of best practices for the interlinking of datasets, with *Resource Description Framework* (RDF, Cyganiak et al. (2014)) as a standard data model (Wood et al., 2014, 4–12). RDF represents data in the form of graphs, i.e., triples with a source node (a subject), and a target node (an object), connected through a directed edge (predicate) pointing from the former to the latter; subject, object and predicate are each identified through URIs that must be accessible via HTTP (the object can also be a literal described as a string).

There are many advantages to representing linguistic resources as LD, such as structural interoperability (through same format and same query language), conceptual interoperability (through shared vocabularies and ontologies), uniform access (through the use of standard Web protocols), and resource integration (linking resources) and federation (cross-resource access); cf. Chiarcos et al. (2013).

3.1. Related Work

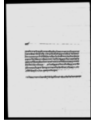
Given the potential of LOD to create interoperability crossing borders of disciplines, content, languages, original formats and places of publication of resources, the last decade has witnessed a significant growth in contributions to the LOD ecosystem. This is also reflected by the strong increase of the LOD cloud (Gandon et al., 2017, 2) with resources from geography, life sciences, (political) administration, social media, etc., showing semantically interconnected information across all fields.⁷ Also in linguistics, many proposals have been made to integrate language resources and their lexical, semantic, and morpho-syntactic aspects into the Semantic Web (Declerck et al., 2015; Aguado de Cea et al., 2016; Jarrar et al., 2019; Bellandi et al., 2018); for the recent state of the art see Bosque-Gil et al. (2018) and Cimiano et al. (2020).

Also for the modeling of text editions as LD, we can build on previous work. Approaches to transform text editions from XML/TEI (TEI Consortium, 2017) into RDF have been proposed, e.g., a mapping of TEI markup to CIDOC CRM (Crofts et al., 2003) in Eide (2014 2015, 23-38), a formalization of the TEI data model into an ontology (Ciotti and Tomasi, 2016 2017), and the integration of RDFa (Herman et al., 2015) into XML/TEI (Tittel et al., 2018), adapted by Cimiano et al. (2020, 253-262). All approaches depend on an interpretation of each individual TEI markup; however, efforts to define a binding mapping of TEI and OntoLex

⁷Cf. <http://lod-cloud.net>.

Abstract

Raṇadala Pāḍe writes this letter to the authority concerned in respect of offering an elephant named Raṇa Prakāśā from Citavana, Nepal.



Diplomatic edition

[1r]
1 अजि -----

2 उग्रान्त-भ्येस्पृक-तप्यामा-भक्रियाकोडुनोमता-हातिरहजुमाचहा
3 ईपदायाकोछदापिलहोला-हातिरन्धेउस्तोकामलान्याहोइन-येसपालादा
4 रोगाराप्रदाहुतुरुस्लेधेर्मिहित्ताद-हेदयेगलेपयो-हातिसारकोनाज्या
5 दामुसहाति-हातिथेदामासधेकाम्लाग्या-सवारिमाफिककाहातिहस्यक्र-या
6 हातिहो-आहाकोहातिसारमाअङगवाहादुरहातिपिनाडाइवसनुभयाज्या
7 कोहातिसारसेसरीजम्या-...[...]-वारिमाफिककाहातिहस्यपिफकिहजुमादापिल
8 हुँदहदाहुन-नोहुकुम-विउप्रभुररणकमलेभुकिमधिकमइतिसयद् ८१४सा
9 लमितिचेरुदिर-प्लोजरमुकाम-शुभम्

10 सदासेकस्-कोटिकोटिकुँससाथांगदवत्सेवासेवासहस्यमुभम्

Translation

[1r]
Arjī –
[Regarding the] following: This time a big rutish (*matā*)¹ elephant captured in an area of [redacted] [and given the name] [redacted] has been sent to you. [Hopefully], it will have reached you. The elephant [named] [redacted] is not always so useful. This time, thanks to the hard labour of the elephant stable manager (*dārogā*)², head of the elephant care team (*raut*)³ and elephant riders (*māhu*), Haidaraveg caught [Raṇa Prakāśā]. The popular chief elephant [redacted] from the elephant stable (*hātsāra*) in [redacted] is the one who is always useful for elephant hunts,⁴ and catches elephants best suited for your outings. The elephant stable here would [therefore] be much better, if you could also send the elephant Aḍaṅga Bahādura. Once elephants suited for outings are caught, they would be continuously turned over to you. [We will do] as you order. What more [to say] to our learned lord whose feet are lotuses!
Monday, the 15th of the bright fortnight of Caitra in the [Vikrama] era 1894 from the [redacted] residence. [Let it be] auspicious.
Tens of millions of eightfold salutations⁵ from [your] always [faithful] servant [redacted], at [your] service. Three times reverence. [Let it be] thousandfold auspicious.

Commentary

Most documents on elephants deal with elephants used for hunting and riding for kings and nobility. They are regarded as a prized symbol of status and, due to their association with Gaṇeśa, of power combined with auspicious qualities. This became last evident in 1975 when King [redacted] and Queen [redacted] undertook their coronation procession on the back of an elephant. The Sāhas and Rāṇās were famous for organizing hunts which could last several weeks and involve hundreds of elephants (Locke 2006: 11). [redacted]’s fondness for elephants and his courage in dealing with them is well-known.
Elephants were important for the economy and lucrative trading commodity (Regmi 1984: 198-199). They were also used in war and in controlling the borders, as is clear from a *śālamohare* issued by King [redacted] dated Mārga sudi 1, VS 1867, to all officials where elephant stables had been established and in which the *dārogās*, *rautas*, *māhutus* etc. are told: “In case you vacate a single inch of the territory under our occupation, you shall be held to have committed a serious crime” (Regmi 1972: 46 referring to RRC, vol. 38, p. 645).

Notes

1. An elephant who is erotic for female elephants and whose ear-glands are flowing. [?]
2. Cp. Edwards 1975: 109; Krauskopff and Meyer 2000: 183, Locke 2006: 149 [?]
3. Locke 2006: 148f.: “In the modern era, *raut* is responsible for managing the team of elephant driving staff (the *māhuts*, *pačhuwas*, and *phanets*); also a Tharu surname: Krauskopff and Meyer 2000: 185. [?]
4. Hunt by stockade method or a “fenced enclosure into which wild elephants were herded before being subjected to training” (Locke 2006): 26. [?]

Figure 1: A text edition as part of the Documenta Nepalica (annotated named entities are highlighted in green).

(Cimiano et al., 2016) are underway.⁸ Textual corpora have received less attention (Cimiano et al. (2020, 43), overview ib. 61-87; 89-122).

Some work has been conducted in the field of Nepalese studies with respect to LD and ontologies. Pokharel et al. (2014) discuss the use of ontologies for improving the effectiveness of farming in Nepal facilitating resources with statistic and administrative data, and weather, soil and crop growing days information. The research project “Opportunity Recognition Model of Nepalese Entrepreneurs” seems to be grounded on an ontology, as well.⁹ CRAI¹⁰ provides a geographical thesaurus with information downloadable as RDF, with a microthesaurus also for Nepal; however, only four Nepalese place names are recorded. Open Knowledge Nepal¹¹ is a non-profit organization making Nepal data openly accessible, providing data-driven services and education of citizens, government, etc.; even though RDF is said to be one of the data formats available (see

the FAQs), RDF datasets are hard to find; the included geographical dataset (i) is not available as RDF, and (ii) focuses on modern administrative boundaries.

4. From the Documenta Nepalica to a Semantic Web Resource

The overall goal is to model the Documenta Nepalica as LINKEDOPENNEPAL, a set of stand-alone features ‘text editions’, ‘person names’, ‘places names’, and ‘glossary’, yet aligned to a multifaceted Semantic Web resource.

Adapting the approach described in Tittel et al. (2018), we give a minimal data sample with RDF data from the edition of document ID K.0440.0007¹² that includes a toponym (we use the English translation for the code example, the Devanagari transcription of the Nepalese original is shown in Fig. 2): *The guthi for the cakrapūjā [...] 4 ropantīs near Deupāṭham*.¹³

⁸Cf. <https://github.com/elexis-eu/tei2ontolex>.

⁹Cf. <https://tinyurl.com/24u2vp9d>.

¹⁰University of Barcelona, *Centre de Recursos per a l’Aprentatge i la Investigació*, cf. <https://vocabulary.crai.ub.edu/en/thub/concept/thub:981058505354306706>.

¹¹Cf. <https://oknp.org/>.

¹²A royal donation, cf. <https://nepalica.hadw-bw.de/nepal/editions/show/8637>.

¹³For sake of brevity, namespaces are assumed defined the usual way. For the yet unpublished resource we use the common <<http://example.org>>, see also in the following code examples. We also use the prefixes ‘occ’ (occurrence) for <<http://example.org/K.0440.0007.xml/#>>, ‘gloss’ (glossary entry) for <<https://nepalica.hadw-bw.de/nepal/words/viewitem/>>, ‘place’

- 52 वालाचतुर्दसिकादीन कावमा
 53 हाचक्रपुजाकोगुठी
 ...
 56 रोपनी४देउपाटंछेउको

Figure 2: Text edition of K_0440_0007.

```

1 The
2 <seg about="occ:210"
3   property="rdfs:seeAlso" resource="gloss:931">
4   <w property="rdfs:label" lemma="guṭhi"
5     type="n.">guṭhi</w>
6   <gloss property="skos:definition">endowed lands or
7     other sources of revenue for financing religious
8     and charitable functions</gloss>
9   </seg>
10 for the
11 <seg about="occ:213"
12   property="rdfs:seeAlso" resource="gloss:2002">
13   <w property="rdfs:label" lemma="cakrapūjā"
14     type="n.">cakrapūjā</w>
15   <gloss property="skos:definition">worship in a
16     circle</gloss>
17   </seg> [...] 4
18 <seg about="occ:235"
19   property="rdfs:seeAlso" resource="gloss:2051">
20   <w property="rdfs:label" lemma="ropanī"
21     type="n.">ropanīs</w>
22   <gloss property="skos:definition">unit of land
23     measurement in the hill region, including the
24     Kathmandu Valley, comprising four muris</gloss>
25   </seg>
26 near
27 <seg about="occ:176" property="rdfs:seeAlso"
28   resource="place:28">
29   <w property="rdfs:label" lemma="Deopatan">
30   Deupāṭhaṃ</w>
31 </seg>

```

Listing 1: Data sample from a text edition of the Documenta Nepalica with XML+RDFa.

For RDF to be Linked Data it must adhere to principles defined by Berners-Lee (2009), one of which reads: “When someone looks up a URI, provide useful information, using the standards” (called a dereferenceable URI). The reference to the Deopatan entry (normalized spelling) of the named entities register in l. 28–31 does not meet this condition.¹⁴ To enhance the information provided and to reach compliance with other resources, a link to the respective entry in a gazetteer or an authority file in the domain of geographic references can be added. These are, e.g., the *Getty Thesaurus of Geographical Names* (TGN), GeoNames, the *Virtual International Authority File* (VIAF), and the *Gemeinsame Normdatei* (GND).¹⁵ The city of Kathmandu, for example, can be connected to GeoNames

for `<.../nepal/ontologies/viewitem/>` (Nota bene: Both resources will be made open access in the course of 2022).

¹⁴Note that this is also true for the references of *guṭhi*, *cakrapūjā*, and *ropanī* in l. 2; 12; 19 to the entries in the glossary of the Documenta Nepalica.

¹⁵Cf. <https://www.getty.edu/research/tools/vocabularies/tgn/>, <https://www.geonames.org/>, <http://viaf.org/>,

ID 1283240, GND 4030036–5, TGN 1083294, and VIAF 158284130.

Also, a mapping to each corresponding DBpedia entity is established, e.g., for the Trishuli River, http://dbpedia.org/resource/Trishuli_River, see List. 3, l. 19.

However, many of the geographic names found in the Documenta Nepalica (cities, mountains, rivers, etc.) are not registered by these authority files. The marginalization of less ‘developed’ countries in global gazetteers such as GeoNames and TGN with respect to coverage, balance, and completeness is a known issue (Acheson et al., 2017). Furthermore, field names for, e.g., a grove, a pasture, a lot, a place related to a religious site—all playing a potentially important role for events such as rituals and lawsuits—are, to the best of our knowledge, not represented. One example is the Pashupati Aryaghat, a socially important ghat on the banks of the Bagmati River serving as a cremation ground for the Nepalese nobility (Michaels, 2008, 5f.).¹⁶ Also, these authority files focus on present day toponymy and, hence, their suitability is limited for the mapping of diachronic changes and historic place names, such as aforementioned Deopatan, a historic town in the Kathmandu Valley and now a city quarter of Kathmandu.

And the problem is more complex: Even with an existing reference, the information returned when navigating to the entry of an authority file is not in an LD standard, thus, it is not ‘useful’ in the sense of LD. We tackle this problem with an ontology for Nepalese toponyms.

4.1. Why an Ontology for Toponyms

The texts of the Documenta Nepalica attest complex connections of norms, ideas, and rules to places, practices, persons, castes, and the material world. This makes them “such exciting material” (Cubelic et al., 2018, 1) and a key component for understanding the history of Nepal. However, they have not yet been sufficiently studied, neither as a self-sustained textual category, nor as source material for a historiography of South Asia, nor in relation to other texts, such as inscriptions, shastric texts, chronicles, belles lettres, etc. Their content is varied: festivals and rituals of religious communities, emerging scribal and administrative elites, court proceedings and litigation, the development of bureaucratic policies, offices, social roles, military positions, and state duties, cp. ib. 11–13. They also give insight into the social and cultural circumstances and forms of slavery in Nepal (and South Asia) differing significantly from better-studied African and North American forms, being more familiar and strongly related to land and landownership

<http://www.dnb.de/gnd>.

¹⁶Attested in numerous documents, e.g., in doc. RRC.0062_0180, <https://nepalica.hadw-bw.de/nepal/editions/show/47509>).

(Bajracharya and Michaels, 2022, 1f.). All aspects are entangled with places, and all require a careful regional contextualisation to enable a deeper understanding of Nepalese socio-historical development, especially considering trans-regional migration patterns (e.g., of scribal groups) and fluctuating borders with emerging or perishing kingdoms, cp. Khatiwoda et al. (2021, 47–56).

Therefore, as a major step towards a LOD transformation of data from the Documenta Nepalica, we develop an ontology that will provide the necessary classes and properties for a modeling of this entanglement. We will populate the ontology with individuals taken from the Documenta Nepalica, turning it into a means to create a bird’s eye view on the interlacing of Nepalese places with events and personalia through time.

4.2. Digital Status Quo

While the format of the text editions is in XML conform with the standard TEI P5 (published online as HTML) and can thus be utilized for a future automated transformation into RDF statements, the entries of the named entities register are formatted using HTML:

```

1 <p>
2 capital of Nepal, declared capital by
3 <a href="/nepal/ontologies/viewitem/178">
4 Pṛthvinārāyaṇa Śāha</a> on 21 March 1770;</p>
5 <p>
6 a district in Central Nepal, part of
7 <a href="/nepal/ontologies/viewitem/651">
8 Bagmati Province</a>
9 </p>

```

Listing 2: HTML entry for Kathmandu.

The code example shows three aspects that are perfectly suited for the current data use but that are shortcomings with respect to automated data processing: (i) all entries of the register are annotated as undifferentiated hyperlinks: in the example, the person name Pṛthvinārāyaṇa Śāha, a king, and the toponym Bagmati Province; (ii) the information relevant for the classification of the entry is an unstructured string, in this case, the fact that Kathmandu is both the capital of Nepal and a district; (iii) the founding date of Nepal’s capital is not annotated. For the time being, we resolve this by extracting the information manually.

4.3. Ontology Development

We construct a conceptual model through the Protégé v5.2.0 desktop version¹⁷ which we formalize in OWL (Bechhofer et al., 2004), integrating existing standards for cross-resource compatibility. As for metadata, this is straightforward: We use DublinCore properties (`creator`, `title`, `issued`, `description`, `rights`)¹⁸ and VANN (`vann:preferredNamespacePrefix`, `vann:preferredNamespaceUri`).¹⁹; the license is defined as Creative Commons Zero. We also use

¹⁷Cf. <http://protege.stanford.edu>.

¹⁸Cf. <http://purl.org/dc/terms>.

¹⁹Cf. <https://vocab.org/vann/>.

RDF, RDFS (Brickley and Guha, 2014), and the datatype `xsd:dateTime` (Fallside and Walmsley, 2004). However, to represent the information in focus, this does not suffice. Two aspects are crucial: the establishment of the classes and properties that will enable (i) a classification of the individuals of the ontology (e.g.: city of Kathmandu) and (ii) a modeling of their entanglement with other individuals, i.e., toponyms, events, dates, agents, etc. (e.g.: Deopatan is a historic town in the Kathmandu valley, now part of the city of Kathmandu). For the first point we are fortunate to be able to adopt the ontology established by DBpedia, starting with `Place` as the hierarchically highest class and including only classes relevant for our data.²⁰

The second aspect is more time-consuming. It is work in progress: The incremental population with entries from Documenta Nepalica as ontology individuals leads to repeated revisions of the created properties (and, to a lesser extent, of classes). At this point, the properties are as generic as possible, with specifications focusing on rivers (with object properties `hasTributary`, `isTributaryOf`, `flowsIntoRiverAt`). The property `connectsTo` with `Place` as its domain and `Person` as its range facilitates the connection to the person name ontology `NEPALPEOPLE`, see *infra*. For alternative designations of a toponym we use the `GeoNames` property `alternateName`.²¹

We give an example for the river related properties with the modeling of the Trīsūlī river, serialized with Turtle Syntax (Prud’hommeaux and Carothers, 2014):

```

1 ### https://example.org/nepalplaces#
2 Trishuli_River
3 :Trishuli_River rdfs:type owl:NamedIndividual ,
4 :River ;
5 :flowsIntoRiverAt :Devghat ;
6 :hasTributary :Marsyangdi_River ;
7 :isLocatedIn :Nepal ;
8 :isTributaryOf :Narayani_River ;
9 rdfs:label "Trīsūlī"@ne ;
10 gn:alternateName "Trīsūlagangā"@ne ,
11 "Trīsūlī nadi"@ne , "Trīsūla Gaṅgā"@ne ;
12 rdfs:comment "One of the major tributaries
13 of the Narayani River basin in central
14 Nepal; originates in Tibet as a stream
15 and enters Nepal at Gyirong Town; joins
16 the Narayani River at Devghat."@en ;
17 rdfs:seeAlso <https://nepalica.hadw-bw.de/
18 nepal/ontologies/viewitem/213> ;
19 owl:sameAs dbr:Trishuli_River .

```

Listing 3: Individual Trīsūlī river.

Currently, 57 individuals, 76 classes (used 39 times), seven object properties (used 47 times), and 15 annotation properties are registered in `NEPALPLACES`.

4.4. Ongoing Work

The ontology development is a joint activity with domain experts from the research project’s team resulting

²⁰Cf. <http://mappings.dbpedia.org/server/ontology/classes/#Place>.

²¹<https://www.geonames.org/ontology/documentation.html#alternateName>.

in and profiting from numerous fruitful discussions, exemplified in the following.

4.4.1. Identification of Shortcomings of the Class Structure

NEPALPLACES must provide the means to map Nepalese toponyms to a class. As mentioned above, we can draw on the DBpedia ontology. However, the DBpedia ontology proved to be (i) not fine-grained enough and (ii) not accurate for the modeling of Nepalese reality and—for that matter—of the reality of the Indian subcontinent. Thus, we extend it, integrating new subclasses into the hierarchy. This is a process driven by several aspects:

Population with Individuals. The population of NEPALPLACES with data from the named entities register from Documenta Nepalica uncovers shortcomings of the class structure in a straightforward way. E.g., the aforementioned Pashupati Aryaghat is a ghat and needs to be classified as such in the ontology. Since this is not possible with the DBpedia ontology we introduce Ghat as a sub-class of `ArchitecturalStructure` (on the level of `Building`), annotated with the comment “A set of steps leading down to a body of water with a platform for bathing typically situated at their base”, cf. Fig. 3. We said already that Pashupati Aryaghat is also used as a cremation ground with high social significance and also needs to be mapped to this concept. Thus, we also introduce `CremationGround` to the ontology, where the pre-existing `Cemetery` “burial ground” is not accurate. So far, with currently 57 individuals, twelve classes have been added [effective 03-12-2022].

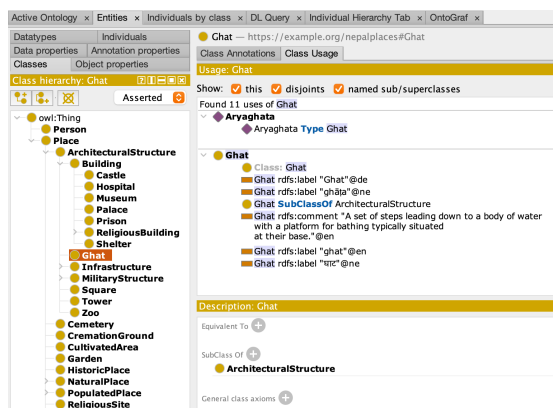


Figure 3: Newly added class Ghat (Protégé).

Concepts behind Classes. For the classification of each individual, we closely look at a pre-existing class that seems suitable at first sight. E.g., for the modeling of the toponym Bhadrakali, an open air shrine of the goddess Bhadrakālī east of Tundikhel in Kathmandu, we turn to the class `Shrine`, in the DBpedia ontology labeled as Engl. “shrine” and placed within the hierarchy `< ReligiousBuilding < Building`

`< ArchitecturalStructure`.²² A closer look at the concept behind `Shrine` reveals a problem: Engl. *shrine* [`< lt. SCRĪNIUM` “case or chest for books or papers”, (von Wartburg, since 1922, 11,337b)] is attested since ca. 1000 (Ælfric of Eynsham) and seems strongly connected to the Christian Occident, typically suggesting a container with physical remains of a religiously venerated person, cf. the senses registered by OED `SHRINE` n. with sense n°1 “A box, coffer [...]”, n°2 “The box [...] in which the relics of a saint are preserved [...]”, etc., and sense n°5a defines a more generic concept that, however, still refers to an architectural structure: “A place where worship is offered or devotions are paid to a saint or deity; a temple, church”.²³ It becomes clear that the class `Shrine` is not adequate for a mapping of Nepalese shrines that are often open-air places of worship. Thus, we include a generic concept represented by `ReligiousSite` that is not related to `ArchitecturalStructure`, accompanied by the comment “A natural site where religious rituals and activities, e.g., prayer and sacrifice, are carried out.”.

Class and Property Labeling. When defining the classes and properties, our aim is to include labels in English, German, and Nepali (alongside English comments). There, we experience a noticeable impact on the ontology engineering by the translation process of the labeling, unveiling shortcomings within the class structure adopted from DBpedia with respect to Nepalese reality. E.g., there is no adequate Nepali term for a monastery in a generic sense: the Nepalese language differentiates *bahāla / vihāra*, a Buddhist monastery, from *maṭha*, a monastery of Hindu ascetics, and from *girjāghara*, a Christian monastery.²⁴ This results in adding three sub-classes to the pre-existing `Monastery`: `MonasteryChristian`, `MonasteryBuddhist` and `MonasteryHindu`, cf. the individual `Bhinchē Bāhāla` as an example:

```

1 :Bhinc̣he_Bahala rdf:type owl:NamedIndividual ,
2                               :MonasteryBuddhist ;
3   :isLocatedIn
4     rdf:slabel "Patan ;
5     "Bhinchē Bāhāla"@ne ,
6     "Bhīṃkṣe Bāhāla"@ne ;
7   rdfs:comment "Buddhist monastery in Patan
8   and the city quarter surrounding it."@en ;
9   rdfs:seeAlso <https://danam.cats.uni-
10  heidelberg.de/report/711b19a4-ec10-
11  11e9-b125-0242ac130002> ,
12  <https://nepalica.hadw-bw.de/nepal/
  ontologies/viewitem/821> .

```

Listing 4: The Buddhist monastery Bhinchē Bāhāla.

Note the linking (through `rdfs:seeAlso`, l. 8-10) to the respective entry in the *Nepal Heritage Documentation Project* (NHDP; Heidelberg Centre for Transcultural Studies, Heidelberg University).

²²Cf. <http://mappings.dbpedia.org/server/ontology/classes//Shrine>.

²³*Oxford English Dictionary* – OED Online. March 2022. Oxford University Press. www.oed.com.

²⁴This term being rarely used and replaced by Nep. *carca* designating the church.

The examples described reveal a Western-centric cultural bias within the encyclopedic resource of DBpedia. This results in an information imbalance in the direction of the reality of non-western cultures such as the Nepalese. This flaw can be counteracted, e.g., by ontological work such as NEPALPLACES, and we hope that our efforts foster more endeavors in this regard.

4.4.2. Ambiguity within the Entries of the Documenta Nepalica

The entries of the named entities register show ambiguities that need to be resolved while integrating them into the ontology. See, e.g., Fig. 4, where the entry misses the distinction between the historical district of Deukhurī and the valley of Deukhurī. The entry for the place name of Dang is yet more complex: “a valley located in the Inner Tarai in midwest Nepal (north of Deukhurī valley) that was prior to the conquest by the Gorkha troops in 1786 divided among small principalities belonging to the Bāise Rājya: Dang and Chilli in the valley itself, and Salyan and Phalabang in the hills [...]; a district in Western Nepal, part of the Lumbini Province; until 1961 together with Deukhurī one of the 32 districts formed in the Rana period”, <https://nepalica.hadw-bw.de/nepal/ontologies/viewitem/389>: The entry describes both the Dang valley and the Dang district that, in turn, must be differentiated into a modern administrative unit and a historic administrative unit of the Rana period. Thus, for the content of these entries, different instances must be introduced to the ontology (e.g., `Deukhuri.District` and `Deukhuri.Valley` for the entry `Deukhurī`), explicitly defined as different individuals. This has so far been done manually. For future automatic processing, ways of an automatized information identification and entity recognition must be evaluated.

4.4.3. Geographical Names Entangled with Offices and People

The Documenta Nepalica include a great amount of information on administrative, religious and social offices and roles together with persons holding the office or role during a particular time and at a particular place. This empirical information is very valuable for a nuanced interpretation of responsibilities, dependencies, interacting, and structural developments within Nepalese society from a diachronic and a spatial perspective. To model the manifold relations of a place presents a challenge to the development of the ontology. The following gives an example of such relations: The Degutalejyū Temple located at the Hanumāndhokā palace in Kathmandu is the place of activity of a particular *nagarci*, a drummer who is a musician playing an important social and religious-ritual role. This role is connected to land endowments for financing the religious and charitable functions of the *nagarci*, as witnessed in different documents. One such document is a royal deed from King Gīrvāṇayuddha in Vikram Sam-

vat 1864 (= CE 1807²⁵). The deed grants plots of land (with rice fields, gardens, lots, etc.) to a person called Bandhuvā Damāi (Badhuvā Nagārci, resp.) for conducting rituals (providing the necessary material, e.g., buffalo, goats, cloth, etc.), for the upkeep of two large religious banners and three long trumpets offered to Degutalejyū, and, also, for military services.²⁶

Since we want to keep NEPALPLACES focused on toponymy and not populate it with aspects related to anthroponymy and prosopography, we establish a connection of NEPALPLACES to a sister ontology called NEPALPEOPLE. This sister ontology models persons within the Documenta Nepalica, including their proper names, life events and genealogical aspects, professions, societal roles, and more. In NEPALPEOPLE, Bandhuvā Damāi and all other persons with the same position (of *nagarci*) are modeled as individuals connected (via a specific NEPALPEOPLE property `hasPosition`) to a position defined through the class `Nagarci`; this position is, in turn, enriched with temporal information and, through a property `hasPlace`, with the individual ‘Degutalejyū Temple in Kathmandu’ of NEPALPLACES.²⁷

4.5. Integration of chronological aspects

The ontology provides classes representing historical concepts: `HistoricalPlace`, `HistoricalAreaOfAuthority`, `HistoricalSettlement`, and `HistoricalDistrict`, the latter classifying, e.g., former districts of the Rana period until 1961 such as said Deukhurī and Dang districts. However, these classes do not inform us about a precise date and, thus, the dimension of time needs to be integrated: When was, e.g., a district created, and when did the existence come to an end or merge into another administrative area? E.g., Tanahun is a modern district in Western Nepal and part of the Gandaki Province but it has also been a kingdom, i.e., one of the Caubīsī Rājya (twenty-four sovereign and intermittently allied petty kingdoms in South Asia ruled by the Khas people and unified between 1744 [by Pṛthvīnārāyaṇa Śāha, king of Gorkha Kingdom, cp. List. 2] and 1816 to what is now present-day Nepal (van Driem, 2001, 1107). The Time Ontology in OWL (Cox and Little, 2020) provides a wide range of classes, object properties and data types to model time instances and time spans that are of particular interest for Nepalese temporal entities entangled with toponyms: Alongside the Gregorian calendar as standard time, one can define alternative calendars through the property `time:hasTRS` with the object being, e.g., a DBpedia entry. The following example

²⁵The Nepalese documents show different calendars: Vikram Samvat (VS), Shaka Era (ŚS), and Nepal Sambat (NS); when not further specified, we use CE (Commen Era).

²⁶Document DNA_0013_0031, <https://nepalica.hadw-bw.de/nepal/editions/show/794>.

²⁷Details of NEPALPEOPLE are not subject of this paper.

Details for Deukhuri

Edit this item

ID	392			
Name	Deukhuri			
Type	placeName			
Notes	also Deukhuri, Deuṣuri; a valley in midwest Nepal, south of Dang; until 1961 together with Dang one of the 32 districts formed in the Rana period; wiki:Deukhuri.			
Surname				
	Name	XML file	Year	reliable?
	Deukhuri	K_0469_0045	1907 (VS 1964)	yes
	देउपुरी	K_0469_0045	1907 (VS 1964)	yes

Figure 4: Deukhuri in the named entities register of the Documenta Nepalica.

shows how to model the CE date (lines 1–3) and the VS date (lines 5–9) of the same event, i.e., the building of a shelter (nep. *pāṭī*) in a specific place, that is, in Kayatā Hiti (Kathmandu):

```

1 :Building_Kayatahiti_Pati_Gregorian
2   rdf:type time:DateTimeDescription ;
3   time:year "1796"^^xsd:gYear .
4
5 :Building_Kayatahiti_Pati_VS
6   rdf:type time:GeneralDateTimeDescription ;
7   time:hasTRS
8     <https://dbpedia.org/page/Vikram_Samvat> ;
9   time:year "1852"^^time:generalYear .

```

Listing 5: Inclusion of VS calendar with Time Ontology in OWL.

Including the Gregorian calendar parallel to the historical Nepalese calendar (VS in the given example) is important for aligning the dates with standard CE time and, thus, creating a common ground with temporal information of other resources. However, to access the data one still needs to focus on an explicit date. Therefore, we want to evaluate how to define historical periods based on the evidence in the Documenta Nepalica (backed by related research) and how to make them interoperable with PeriodO. PeriodO is a gazetteer of historical periods, initially designed for European antiquity (Rabinowitz et al., 2016).²⁸ Its aim is to make concepts of periods explicit through a spatio-temporal definition based on an authoritative source. This adds new, finegrained concepts to globally agreed-upon concepts. The ‘19th century’ is one such concept. Much like the temporal borders of the ‘Renaissance’ differ depending on space, the 19th century also proves to be a dynamic notion: Based on its aspects of ‘modernity’ brought “through massive scientific and technological changes, industrialisation, overseas exploration, nationalism, new forms of administration and new media” (Cubelic et al., 2018, 1), the term of the ‘Nepalese long 19th century’ has become an authoritative notion of a period for Nepal, cf. Osterhammel (2011, 87–88), Michaels (2018), Zotter (2018). Hence, it could be integrated into PeriodO. Through such periods, historical geographical concepts such as former kingdoms

²⁸Cf. <https://perio.do/en/>.

and areas of rulership can be made comparable and interoperable with other places, be they chronologically congruent, previous, subsequent, or overlapping.

In this context, it needs to be discussed (i) when to consider a place as historical and (ii) how to classify it. Would the evidence of a continuous, legal administrative succession decide on the classification as an area of authority (cf. the hierarchy in Fig. 5)? For instance, is Punjab under Maharaja Ranjit Singh, the ‘Lion of Punjab’, a *HistoricalAreaOfAuthority*, a *Country* to be classified as ‘historical’, or even a *HistoricalPlace*?

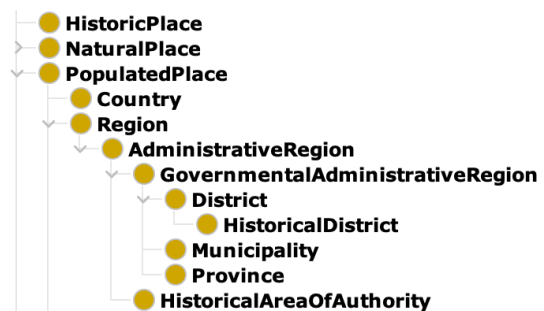


Figure 5: Historical concepts of NEPALPLACES.

5. Conclusion

NEPALPLACES presents the first step towards the creation of LINKEDOPENNEPAL. It shows very promising results but still is a work in progress: By incrementally populating the ontology with all the toponyms registered in the Documenta Nepalica text corpus, the ability of NEPALPLACES to facilitate modeling of all relevant information will be put to the test. This is a process that will most likely necessitate repeatedly revisiting its structure. To populate NEPALPLACES in an exhaustive way is, thus, a task that cannot easily be integrated into the research project’s time frame and work flow. Therefore, we consider executing this task either in the form of a satellite or, more likely, a follow-up project.

6. Bibliographical References

- Acheson, E., De Sabbata, S., and Purves, R. (2017). A quantitative analysis of global gazetteers: Patterns of coverage for common feature types. *Computers, Environment and Urban Systems*, 64:309–320.
- Aguado de Cea, G., Montiel-Ponsada, E., Kernerman, I., and Ordan, N. (2016). From Dictionaries to Cross-lingual Lexical Resources. *Kernerman DICTIONARY News*, 24:25–31.
- Bajracharya, M. and Michaels, A. (2022). Introduction. In Manik Bajracharya, editor, *Slavery and Un-free Labour in Nepal. Documents from the 18th to Early 20th Century*, pages 6–36. Heidelberg University Publishing, Heidelberg.
- Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D., Patel-Schneider, P., and Stein, L. (2004). OWL Web Ontology Language. Reference. W3C Recommendation 10 February 2004. URL: <https://www.w3.org/TR/2004/REC-owl-ref-20040210/> [accessed: 03-10-2022].
- Bellandi, A., Giovannetti, E., and Weingart, A. (2018). Multilingual and Multiword Phenomena in a lemon Old Occitan Medico-Botanical Lexicon. *Information*, 9 (3), 52.
- Berners-Lee, T. (2009). Linked Data. URL: <https://www.w3.org/DesignIssues/LinkedData.html> [accessed: 03-15-2022].
- Bizer, C., Heath, T., and Berners-Lee, T. (2009). Linked Data – The Story So Far. *International Journal on Semantic Web and Information Systems*, 5:1–22.
- Bosque-Gil, J., Gracia, J., Montiel-Ponsoda, E., and Gómez-Pérez, A. (2018). Models to represent linguistic linked data. *Natural Language Engineering*, 24(6):811–859.
- Brickley, D. and Guha, R. (2014). XML Schema Part 0: Primer Second Edition. URL: <https://www.w3.org/TR/rdf-schema/> [accessed: 03-14-2022].
- Chiarcos, C., McCrae, J., Cimiano, P., and Fellbaum, C. (2013). Towards Open Data for Linguistics: Lexical Linked Data. In Alessandro Oltramari, et al., editors, *New Trends of Research in Ontologies and Lexical Resources: Ideas, Projects, Systems*, pages 7–25. Springer, Berlin, Heidelberg.
- Cimiano, P., McCrae, J. P., and Buitelaar, P. (2016). Lexicon Model for Ontologies: Community Report. Final Community Group Report, 10 May 2016. URL: <https://www.w3.org/2016/05/ontolex/> [accessed: 03-15-2022].
- Cimiano, P., Chiarcos, C., McCrae, J., and Gracia, J. (2020). *Linguistic Linked Data: Representation, Generation and Applications*. Springer, Cham.
- Ciotti, F. and Tomasi, F. (2016–2017). Formal Ontologies, Linked Data, and TEI Semantics. *Journal of the Text Encoding Initiative* [Online], Issue 9.
- Cox, S. and Little, C. (2020). Time Ontology in OWL. W3C Candidate Recommendation 26 March 2020. URL: <https://www.w3.org/TR/owl-time/> [accessed: 03-10-2022].
- Crofts, N., Doerr, M., and Gill, T. (2003). The CIDOC Conceptual Reference Model. A Standard for Communicating Cultural Contents. *Cultivate Interactive*, 9.
- Cubelic, S., Michaels, A., and Zotter, A. (2018). Studying Documents of South Asia: An Introduction. In Simon Cubelic, et al., editors, *Studies in Historical Documents from Nepal and India*, pages 1–33. Heidelberg University Publishing, Heidelberg.
- Cygnaniak, R., Wood, D., and Lanthaler, M. (2014). RDF 1.1. concepts and abstract syntax: W3C recommendation 25 February 2014. URL: <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/> [accessed: 03-14-2022].
- Declerck, T., Wandl-Vogt, E., and Mörth, K. (2015). Towards a Pan European Lexicography by Means of Linked (Open) Data. In Iztok Kosem, et al., editors, *Electronic Lexicography in the 21st Century: Linking Lexical Data in the Digital Age. Proceedings of the eLex 2015 Conference, 11-13 August 2015, Herstonceux Castle, United Kingdom*, pages 342–355. Ljubljana/Brighton. Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd.
- Diwasa, T., Bandhu, C., and Nepal, B. (2007). *The Intangible Cultural Heritage of Nepal: Future Directions*. UNESCO Kathmandu Series of Monographs and Working Papers No 14, Kathmandu.
- Eide, Ø. (2014–2015). Ontologies, Data Modeling, and TEI. *Journal of the Text Encoding Initiative*, Issue 8.
- Fallside, D. and Walmsley, P. (2004). RDF Schema 1.1. W3C Recommendation 25 February 2014. URL: <https://www.w3.org/TR/xmlschema-0/> [accessed: 03-14-2022].
- Gandon, F., Sabou, M., and Sack, H. (2017). Weaving a Web of Linked Resources. *Semantic Web Journal*, 6:1–6.
- Herman, I., Aside, B., McCarron, S., and Birbeck, M. (2015). RDFa Core 1.1 – Third Edition. URL: <https://www.w3.org/TR/rdfa-core> [accessed: 03-14-2022].
- Hutt, M. (1988). *Nepali. A National Language and its Literature*. Sterling Publishers, New Delhi.
- Jarrar, M., Amayreh, H., and McCrae, J. (2019). Representing Arabic Lexicons in Lemon – a Preliminary Study. In Thierry Declerck et al., editors, *2nd Conference on Language, Data and Knowledge (LDK 2019). Posters Track*. Birzeit University.
- Khatiwoda, R., Cubelic, S., and Michaels, A. (2021). *The Mulukī Ain of 1854. Nepal’s First Legal Code*. Heidelberg University Publishing, Heidelberg.
- Michaels, A. (2008). *Siva in Trouble. Festivals and Rituals at the Pasupatinatha Temple of Deopatan*. OUP, New York.

- Michaels, A. (2018). *Kultur und Geschichte Nepals*. Alfred Kröner Verlag, Stuttgart.
- Osterhammel, J. (2011). *Die Verwandlung der Welt: Eine Geschichte des 19. Jahrhunderts*. C.H. Beck, München.
- Pokharel, S., Sherif, M., and Lehmann, J. (2014). Ontology Based Data Access and Integration for Improving the Effectiveness of Farming in Nepal. In *Proceedings of the International Conference on Web Intelligence and Intelligent Agent Technologies (WI-IAT '14)*, pages 319–326.
- Prud'hommeaux, E. and Carothers, G. (2014). RDF 1.1 Turtle: Terse RDF Triple Language. W3C Recommendation, 25 February 2014. URL: <http://www.w3.org/TR/turtle/> [accessed: 03-10-2022].
- Rabinowitz, A., Shaw, R., Buchanan, S., Golden, P., and Kansa, E. (2016). Making Sense of the Ways We Make Sense of the Past: The Periodo Project. *Bulletin of the Institute of Classical Studies*, 59(2):42–55, 12.
- Riccardi, T. (2003). Nepali. In George Cardona et al., editors, *The Indo-Aryan Languages*, pages 538–580, New York. Routledge.
- TEI Consortium. (2017). TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 3.2.0. Last updated on 10th July 2017. URL: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/> [accessed: 03-16-2022].
- Tittel, S., Bermúdez-Sabel, H., and Chiarcos, C. (2018). Using RDFa to Link Text and Dictionary Data for Medieval French. In John P. McCrae, et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). 6th Workshop on Linked Data in Linguistics (LDL-2018), 12 May 2018, Miyazaki, Japan*, pages 30–38, Paris. ELRA.
- van Driem, G. (2001). *Languages of the Himalayas*. Brill, Leiden/Boston/Köln.
- von Wartburg, W. (since 1922). *Französisches Etymologisches Wörterbuch. Eine darstellung des galloromanischen sprachschatzes – FEW*. ATILF. [continued by O. Jänicke, C.T. Gossen, J.-P. Chambon, J.-P. Chauveau, and Yan Greub].
- Wood, D., Zaidman, M., Ruth, L., and Hausenblas, M. (2014). *Linked data: structured data on the web*. Manning Publications Co., New York.
- Zotter, A. (2018). Papier ist geduldig. Digital edierte Dokumente geben Aufschluss zur Nationalstaatsbildung in Nepals “langem 19. Jahrhundert” (1768–1951). In Union der deutschen Akademien der Wissenschaften, editor, *Die Wissenschaftsakademien – Wissensspeicher für die Zukunft Forschungsprojekte im Akademienprogramm*, Berlin.