

Narrative Geospatial Knowledge in Ethnographies: Representation and Reasoning^{*}

Chin-Lung Chang^{1,3}, Yi-Hong Chang², Tyng-Ruey Chuang¹,
Dong-Po Deng^{1,4}, and Andrea Wei-Ching Huang¹

¹ Institute of Information Science

² Institute of Ethnology

Academia Sinica

Taipei, Taiwan

³ Department of Computer Science and Information Engineering

National Taiwan University of Science and Technology

Taipei, Taiwan

⁴ ITC — International Institute for Geo-Information Science and Earth Observation
Enschede, The Netherlands

Abstract. Narrative descriptions about populated places are very common in ethnographies. In old articles and books on the migration history of Taiwan aborigines, for example, narrative sentences are the norms for describing the locations of aboriginal settlements. These narratives constitute a form of geospatial knowledge, and there is a need to develop knowledge representation and reasoning techniques to help analyze literatures, and to aid field works. In this paper, we outline the design of a formal vocabulary to represent and reason about geospatial narratives about populated places, keeping as close as possible to the phrases used in ethnographies. The vocabulary is implemented as OWL concepts and properties, and the rules for geospatial reasoning are expressed in SWRL.

1 Narrative Geospatial Knowledge

In research and study about People and Place, it is necessary to acquire and analyze geospatial information about populated places. Such geospatial information — the location of a settlement relative to certain geographic features and other settlements, for example — often is described in a natural language, and the descriptions themselves are expressed in an everyday vocabulary that is intuitive to human but difficult to process automatically. As an example, let us look at the following sentence, which is taken and translated from an ethnohistorical article about the Atayal aborigine in Taiwan:

Kanaongan Settlement ... is located at the right bank of Dacingshuei River, where the river meets the ocean. [17, p. 137]

^{*} This research is supported in part by National Science Council of Taiwan under grant NSC 98-2410-H-001-075-MY2 (“Ontology-based Collaborative Production of Geospatial Information”).

The location of Kanaongan Settlement is described by two statements about the place’s geospatial relations to others: 1) it is at the right bank of the Dacingshuei river, and 2) it is near the Dacingshuei estuary. Although we are not given the coordinates of the Kanaongan settlement, we have a good idea of the settlement’s position, even more so if we consider the facts that the Dacingshuei river is eastbound, and it flows into the Pacific Ocean. That is, we can conclude that the settlement is near by the southern river bank of the Dacingshuei estuary (as the right bank of an eastbound river is the southern bank of the river).⁵ Suppose that we learn from the same article or other sources that settlement *X* and Kanaongan Settlement are on the opposite sides of the same river, and they face each other. We will be able to conclude that settlement *X* is near by the northern river bank of the Dacingshuei estuary. Note that the conclusion is drawn without the need to know the location of the Dacingshuei estuary.

Narrative descriptions about geospatial relations among Places — collected from ethnohistorical records or field interviews — constitute a kind of geospatial knowledge that is rich in domain semantics, difficult to acquire by other means, and defying easy assimilation and analysis in and by conventional geographic information systems (GIS). To systematically process and analyze large collections of these geospatial narratives, therefore, it calls for new methods and new techniques. If successful such developments will help shed new lights on several areas such as knowledge representation and reasoning, ontology and semantic web, humanistic GIS, *etc.*

We list below several characteristics of narrative geospatial descriptions. These characteristics also magnify themselves as the main technical issues in processing narrative geospatial knowledge.

- The narratives are expressed in everyday vocabularies that are rich (even diverse) in their linguistic and cultural interpretations (the “left bank” of a river, the “foot” of a mountain, *etc.*).
- Directional and relative terms are used to describe the location of a place, as well as its spatial relations with others. The terms are approximative (“opposite to”, “about 100 kilometers away”, “near by”, *etc.*).
- Places are identified by (common) names, not by coordinates. Their positions and footprints are left unspecified. If specified, they are imprecise or vague by definition. (For example, what is the definite spatial extent of a mountain?)
- The descriptions about a place can be conflicting, incomplete, or missing. Often such places are of an ethnohistorical nature and cannot be identified by nowadays technologies. Nevertheless they can be used as the primary references for other places. (Where is Shangri-La?)

⁵ However, the map in the same article puts Kanaongan Settlement at the northern bank of the Dacingshuei estuary [17, p. 135]. (Also see Figure 6 in Appendix A.) This case illustrates that geospatial reasoning can be used to detect inconsistency in geospatial statements — in the article either the sentence or the map is wrong. Or it could be that our understanding of the term “right bank” is different from that of the author.

2 A Vocabulary for Qualitative Geospatial Expressions

The main purpose of qualitative spatial representation and reasoning is to make explicit common-sense knowledge, so that given appropriate reasoning techniques, a computer could make prediction, diagnose and explain the behavior of physical system in qualitative manner without resolving to an often intractable or perhaps unavailable quantitative model [6]. Therefore, qualitative spatial representation and reasoning not only acts as a model to clarify formal semantics of qualitative spatial objects and relations from narrative descriptions, but is also used to find new information from what is already known. There are numerous studies on this subject, on various aspects of spatial relationships such as topology, orientation, distance, size, and shape [4, 5, 9]. However, given the nature of geospatial information present in narrative descriptions, currently we only concern about the representation of certain topological, directional, and orientational information when specifying qualitative relations among geospatial entities.

As a study on narrative geospatial knowledge, we have looked into ethnographies for descriptions about the locations of Taiwan aboriginal settlements, so as to use them as the sources of actual vocabularies for geospatial expressions. Take the following sentences about Sikilian Settlement as examples.⁶

Sikilian Settlement is located at the left bank of midstream Liwu River, and is about 1.6 kilometers to the northwest of the junction of its branch Wahei-er River. It is opposite to the small terrace slightly east of and below the Syuejiachang Station on Central Cross-Island Road. That is, it is at the mountain belly north of Mantou Mountain. [16, p. 178]

One immediately notices geospatial phrases that may subject to different interpretations (“left bank” and “mountain belly”). The use of size and direction modifiers are also problematic (“the *small* terrace *slightly* east of”). There are pronouns and missing nouns to resolve too (“the junction of *its* branch Wahei-er River”). Nevertheless, from these sources we have identified a small set of phrases frequently used for geospatial references:

- phrases for directional references such “is x kilometers to the northwest of” and “is north of”;
- phrases for orientational references such as “left bank” and “opposite to”;
- phrases for references to a part of an geospatial object, such as the “midstream” of a river, and the “belly” of a mountain;
- phrases for different types of natural features (“river” and “junction”) and artificial landmarks (“road” and “station”).

We list in Appendix A a set of nine sample paragraphs drawn from the same ethnography [16]. They are all about the locations of (historical) aboriginal settlements. Based on the geospatial narratives used in these paragraphs, we develop

⁶ These are direct quotations, but of our translations. The original texts are written in traditional Chinese. We take care in making accurate translations of the narratives, in particular about how geospatial references are used in the original texts.

a vocabulary for geospatial references, keeping as close to the phrases that are actually used in these paragraphs. Our goal is that, by using this vocabulary, we can represent and reason in a formal way the geospatial narratives in these ethnographies.

Note that we are not aiming at new techniques which will automatically process natural language texts for the extraction of geospatial knowledge. Rather, we seek systematic methods to express geospatial knowledge in ethnographic narratives so that such knowledge can be aggregated and analyzed, and becomes more useful and reusable to human and machines. In this paper, we only deal with qualitative geospatial knowledge; we consider direction but not distance, for example. We have also avoided relying on numerical calculation in the representation and reasoning of geospatial narratives.

3 Framing A Geospatial Knowledge Domain

Before setting out to acquire and process narrative geospatial knowledge, we need to decide on the scope and the level of details of narrative geospatial knowledge when it is represented in a system. The decision inevitably depends on the formalism we are adapting, and is constrained by the restrictions of the formalism. For this case study, we have settled on OWL DL (the Description Logic subset of the OWL Web Ontology Language) [1] as the representation formalism for the narrative knowledge, and use SWRL (a Semantic Web rule language combining OWL and RuleML) [2] for rule representation and inference.

Compared to the decision on a representation mechanism (OWL DL in our case), the process of adapting and/or devising an appropriate ontology, and that of mapping the assumptions and vocabularies in a narrative knowledge domain to those in the formal ontology, is much harder. These decisions frame the domain of the narrative knowledge, hence, effect the kinds of facts to be admitted and to be reasoned with in the system. Let us take the term “mountain” as an example. In our study of geospatial narratives, “mountain” almost always refers to the area of a mountain. The terms “top”, “side” and “foot” are often used in combination with term “mountain” to refer to the top, side, and foot sub-areas of a mountain area, respectively. These areas and sub-areas may be inhabited by people. In this section, we will use this example to elaborate some of the considerations when framing the domain of geospatial areas, and use these areas as references to the places of settlements.

One can choose to use Region Connection Calculus (RCC) [5] to model the various mountain sub-areas and their relations. An area is mapped to a region in the sense of RCC, and the relation between two areas is but one of the RCC-8 (or rather, RCC-5) relations. That is, each of the mountain top, mountain side, and mountain foot regions *is a proper part of* (PP) the mountain region, see Figure 1. Figure 1, however, expresses more. It illustrates that the mountain region is partitioned into three sub-regions of which the mountain side region surrounds the mountain top region, and the mountain foot region surrounds the mountain side region. Observing this, it could be useful to first define three regions m, s, t

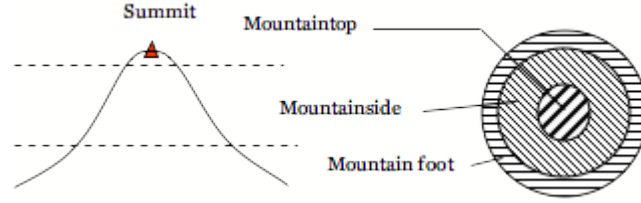


Fig. 1. Topological relations among the top, side, and foot areas of a mountain.

of which m is the region for the entire mountain area, and $PP(s, m)$ and $PP(t, s)$. One can then describe the mountain top area as t , the mountain side area as the area of $s - t$, and the mountain foot area as $m - s$. However, in our case study geospatial narratives do not impose or require topological precision of this kind. The use of a region difference operator on top of RCC-5, we think, will also impose further demands to a system based on OWL DL and SWRL.

One can choose instead to model the top, side, and foot sub-areas of a mountain as three distinct parts of a mountain area, but without explicit mentioning their topological relations. That is, a mountain area will have three separate object properties, each keeping the top, the side, and the foot sub-area, respectively. The classes for the three kinds of sub-areas, as well as that of the entire mountain areas, moreover, are kept distinct from one other, so there will be no confusion of mistaking one kind for another kind. In description logic, they are disjoint concepts. We have adapted this approach to modeling inhabited areas that are identified by common geographic names. That is, in our narrative geospatial domain, the part-whole relationship of a place is elaborated more than its topological relationship.

For the area (and each of their sub-areas) identified by a common geographic name (e.g., “the foot of mountain X ”), we require that it also possesses a “position” property indicating its geophysical location. For example, for the top, the side, and the foot areas of a mountain (and the mountain itself), they all have the summit as their position. By their unique position, the sub-areas can be related together, and associated to the mountain to which they belong. Here, a position serves two purposes — as the coordinates to be used when the geographic feature is visualized on a map, and as an (imprecise) identifier from which parts of the feature can be related together and associated with the whole. Currently we do not require an area to have a “footprint” (or, boundary) property, as such property is not easily quantifiable, even in non-narrative geospatial domains.

By modeling areas referred to by common geographic names as above, we establish a framework of geospatial references for settlements, whose positions can be vague or lost. By translating narrative descriptions of settlement locations to directional, orientational, and topographical relations to existing (and persistent) geographic features and their positions, in our model, we therefore establish the relative positions of the settlements.

4 Representation and Reasoning

We first give a short introduction to an alternative syntax we devise for OWL DL, then proceed to describe our current implementation for the representation and reasoning of geospatial narratives in ethnographies. Figures 2 and 3 list some of the basic concepts and common properties we use in the implementation. The syntax we use is very close to those used in Description Logic (or OWL DL) but is more concise.

Named concepts are introduced by the keyword **Class**. An anonymous concept can be constructed from a property with certain restrictions on its range, *e.g.*, “originateFrom min 1” and “meet some Upstream”. In Figure 2, for example, a River is defined as a Place that originates from at least one Place, ends at one Place, has one Upstream, has one Midstream, and has one Downstream. (From now on, we will simply use the term ‘a river’ to refer to an individual/object in the concept/class River.) Upstream, Midstream, and Downstream, furthermore, are all subclasses of Place. An Upstream has at least one position (of class Point), a Downstream has one position (of class Point), and a Midstream must meet some Upstream and some Downstream. The intuition behind such definitions is that the place from which a river originates are associated to the sources of its upstream, and the place from which a river ends is associated to the sink of its downstream. Moreover, the up/middle/down streams of a river are pair-wise connected by the ‘meet’ relations as required in the definition of Midstream.

Figure 3 shows common properties about places; these are relations among the classes we just define. The keyword **Property** declares a relation between two classes; it specifies the name of the relation, and the class names of its domain and its range. The square bracket immediately follows the declaration can specify the property to be “functional”, “inverseFunctional”, “transitive”, and/or “symmetric” if that is the case for the property. An inverse property f , if exists, is indicated at the end with the notation f^{-1} . In Figure 3, for example, it declares that “southOf” is a transitive property of places with “northOf” as its inverse property.

By using SWRL, we can define additional properties by composing them from existing properties. See Figure 4 for some basic rules for geospatial reasoning. This way, we are free from asserting in OWL DL those facts that can be deduced from SWRL rules, *e.g.*, the fact that a is northeast of b given the facts that a is east of some individual x , and x is north of b . The rules in Figure 4 basically fall into two categories. The first category is about the transfer of property from the part to the whole. For example, if x is located at a mountaintop t which is belonging to a mountain m , then x is located at mountain m as well. The second category is about common-sense geospatial reasoning such as the northeast rule we just mentioned.

In our case, the geospatial narratives used in ethnographies are very concise, but to define the corresponding common-sense geospatial reasoning rules can be quite involving. For example, given the following geospatial narratives:

Class	People.	
Class	Place.	
Class	Settlement	\sqsubseteq Place \sqcap (populatedBy min 1).
Class	Geo.	
Class	Point	\equiv Geo \sqcap (hasHeight exactly 1) \sqcap (hasLatitude exactly 1) \sqcap (hasLongitude exactly 1).
Class	Landmark	\equiv Place \sqcap (hasPosition exactly 1).
Class	Bridge	\sqsubseteq Landmark.
Class	Station	\sqsubseteq Landmark.
Class	Mountain	\equiv Place \sqcap (hasPosition exactly 1) \sqcap (hasTop exactly 1) \sqcap (hasSide exactly 1) \sqcap (hasFoot exactly 1).
Class	Mountaintop	\sqsubseteq Place \sqcap (hasPosition exactly 1).
Class	Mountainside	\sqsubseteq Place \sqcap (hasPosition exactly 1).
Class	Mountainfoot	\sqsubseteq Place \sqcap (hasPosition exactly 1).
Disjoint		Mountaintop, Mountainside, Mountainfoot.
Class	River	\equiv Place \sqcap (originateFrom min 1) \sqcap (endAt exactly 1) \sqcap (hasUpstream exactly 1) \sqcap (hasMidstream exactly 1) \sqcap (hasDownstream exactly 1).
Class	Upstream	\sqsubseteq Place \sqcap (hasPosition min 1).
Class	Downstream	\sqsubseteq Place \sqcap (hasPosition exactly 1).
Class	Midstream	\sqsubseteq Place \sqcap (meet some Upstream) \sqcap (meet some Downstream).
Disjoint		Upstream, Midstream, Downstream.
Class	RiverCross	\sqsubseteq Place \sqcap (hasPosition exactly 1) \sqcap (hasInflow min 2)
Class	PlaceGroup	\sqsubseteq Place.
Class	Pair	\sqsubseteq PlaceGroup \sqcap consistOf exactly 2.

Fig. 2. Concepts for places and settlement in ethnographies.

Sakahen Settlement is opposite to Basawan Settlement; they are separated by Mugua River. Basawan is located at the right bank of upstream Mugua River.

we can infer that Sakahen is located at the left bank of the Mugua river. For this, we need to formulate and use the following rule:

$$\text{Pair}(p) \wedge \text{consistOf}(p, x) \wedge \text{consistOf}(p, y) \wedge \text{differentFrom}(x, y) \wedge \\ \text{River}(r) \wedge \text{inBetween}(r, p) \wedge \text{rightBankOf}(x, r) \rightarrow \text{leftBankOf}(y, r).$$

Before the above rule can be applied, we will first need to pack the two settlements into a pair, as well as to assert that the Mugua river is in between the pair, in addition to the usual assertions that Basawan is a settlement, Mugua is a river, *etc.*

We also need rules to enforce integrity constraints governing the facts admitted to a geospatial knowledge base. Earlier we mentioned the intuition behind

Property populatedBy(Settlement, People).		
Property hasPosition(Place, Point).		
Property locatedAt(Place, Place).		
Property hasTop(Place, Mountaintop).		
Property hasSide(Place, Mountainside).		
Property hasFoot(Place, Mountainfoot).		
Property originateFrom(River, Place).		
Property endAt(River, Place).		
Property hasUpstream(Place, Upstream).		
Property hasMidstream(Place, Midstream).		
Property hasDownstream(Place, Downstream).		
Property meet(Place, Place)	[symmetric].	
Property bankOf(Place, River).		
Property leftBankOf(Place, River)	\sqsubseteq	bankOf.
Property rightBankOf(Place, River)	\sqsubseteq	bankOf.
Property hasInflow(RiverCross, River).		
Property consistOf(PlaceGroup, Place).		
Property inBetween(Place, Pair)		separatedBy ⁻¹ .
Property separatedBy(Pair, Place)		inBetween ⁻¹ .
Property southOf(Place, Place)	[transitive]	northOf ⁻¹ .
Property northOf(Place, Place)	[transitive]	southOf ⁻¹ .
Property westOf(Place, Place)	[transitive]	eastOf ⁻¹ .
Property eastOf(Place, Place)	[transitive]	westOf ⁻¹ .
Property southWestOf(Place, Place)	[transitive]	northEastOf ⁻¹ .
Property northEastOf(Place, Place)	[transitive]	southWestOf ⁻¹ .
Property southEastOf(Place, Place)	[transitive]	northWestOf ⁻¹ .
Property northWestOf(Place, Place)	[transitive]	southEastOf ⁻¹ .

Fig. 3. Common properties about places.

$$\begin{aligned}
& \text{locatedAt}(s, \textit{top}) \wedge \text{hasTop}(m, \textit{top}) \rightarrow \text{locatedAt}(s, m). \\
& \text{locatedAt}(s, \textit{side}) \wedge \text{hasSide}(m, \textit{side}) \rightarrow \text{locatedAt}(s, m). \\
& \text{locatedAt}(s, \textit{foot}) \wedge \text{hasFoot}(m, \textit{foot}) \rightarrow \text{locatedAt}(s, m). \\
\\
& \text{eastOf}(s, \textit{top}) \wedge \text{hasTop}(m, \textit{top}) \rightarrow \text{eastOf}(s, m). \\
& \text{eastOf}(s, \textit{side}) \wedge \text{hasSide}(m, \textit{side}) \rightarrow \text{eastOf}(s, m). \\
& \text{eastOf}(s, \textit{foot}) \wedge \text{hasFoot}(m, \textit{foot}) \rightarrow \text{eastOf}(s, m). \\
& \text{westOf}(s, \textit{top}) \wedge \text{hasTop}(m, \textit{top}) \rightarrow \text{westOf}(s, m). \\
& \text{westOf}(s, \textit{side}) \wedge \text{hasSide}(m, \textit{side}) \rightarrow \text{westOf}(s, m). \\
& \text{westOf}(s, \textit{foot}) \wedge \text{hasFoot}(m, \textit{foot}) \rightarrow \text{westOf}(s, m). \\
& \text{southOf}(s, \textit{top}) \wedge \text{hasTop}(m, \textit{top}) \rightarrow \text{southOf}(s, m). \\
& \text{southOf}(s, \textit{side}) \wedge \text{hasSide}(m, \textit{side}) \rightarrow \text{southOf}(s, m). \\
& \text{southOf}(s, \textit{foot}) \wedge \text{hasFoot}(m, \textit{foot}) \rightarrow \text{southOf}(s, m). \\
& \text{northOf}(s, \textit{top}) \wedge \text{hasTop}(m, \textit{top}) \rightarrow \text{northOf}(s, m). \\
& \text{northOf}(s, \textit{side}) \wedge \text{hasSide}(m, \textit{side}) \rightarrow \text{northOf}(s, m). \\
& \text{northOf}(s, \textit{foot}) \wedge \text{hasFoot}(m, \textit{foot}) \rightarrow \text{northOf}(s, m). \\
\\
& \text{locatedAt}(s, \textit{down}) \wedge \text{hasDownstream}(r, \textit{down}) \rightarrow \text{locatedAt}(s, r). \\
& \text{locatedAt}(s, \textit{mid}) \wedge \text{hasMidstream}(r, \textit{mid}) \rightarrow \text{locatedAt}(s, r). \\
& \text{locatedAt}(s, \textit{up}) \wedge \text{hasUpstream}(r, \textit{up}) \rightarrow \text{locatedAt}(s, r). \\
\\
& \text{eastOf}(a, x) \wedge \text{northOf}(x, b) \rightarrow \text{northEastOf}(a, b). \\
& \text{eastOf}(a, x) \wedge \text{southOf}(x, b) \rightarrow \text{southEastOf}(a, b). \\
& \text{westOf}(a, x) \wedge \text{northOf}(x, b) \rightarrow \text{northWestOf}(a, b). \\
& \text{westOf}(a, x) \wedge \text{southOf}(x, b) \rightarrow \text{southWestOf}(a, b). \\
& \text{northOf}(a, x) \wedge \text{eastOf}(x, b) \rightarrow \text{northEastOf}(a, b). \\
& \text{northOf}(a, x) \wedge \text{westOf}(x, b) \rightarrow \text{northWestOf}(a, b). \\
& \text{southOf}(a, x) \wedge \text{eastOf}(x, b) \rightarrow \text{southEastOf}(a, b). \\
& \text{southOf}(a, x) \wedge \text{westOf}(x, b) \rightarrow \text{southWestOf}(a, b).
\end{aligned}$$

Fig. 4. Basic rules for geospatial reasoning about places.

the definitions of River, Upstream, Midstream, and Downstream. This kind of intuition actually is formalized as a SWRL rule over a river r and its constituting parts up , mid , and $down$ as follows (assuming river r originates from two sources m and n):

$$\begin{aligned} & \text{River}(r) \wedge \text{hasUpstream}(r, up) \wedge \text{hasMidstream}(r, mid) \wedge \text{hasDownstream}(r, down) \wedge \\ & \quad \text{originateFrom}(r, m) \wedge \text{originateFrom}(r, n) \wedge \text{endAt}(r, e) \wedge \\ & \quad \text{hasPosition}(m, p) \wedge \text{hasPosition}(n, q) \wedge \text{hasPosition}(up, p) \wedge \text{hasPosition}(up, q) \wedge \\ & \quad \text{hasPosition}(e, z) \wedge \text{hasPosition}(down, z) \wedge \text{meet}(mid, up) \wedge \text{meet}(mid, down) \\ & \rightarrow \text{Checked}(r). \end{aligned}$$

Similarly, for a mountain m , the part-whole constraints between the mountain top/side/foot and m itself are checked by the following SWRL rule:

$$\begin{aligned} & \text{Mountain}(m) \wedge \text{hasTop}(m, t) \wedge \text{hasFoot}(m, f) \wedge \text{hasSide}(m, s) \wedge \\ & \quad \text{hasPosition}(m, p) \wedge \text{hasPosition}(t, p) \wedge \text{hasPosition}(s, p) \wedge \text{hasPosition}(f, p) \\ & \rightarrow \text{Checked}(m). \end{aligned}$$

Note that in the above we use the concept Point to anchor the concept River: Individual points are used to relate a river to the river’s constituting parts. So is the case for Mountain. We can relate rivers to mountains in a similar way, by connecting the place from which a river originates to certain mountains. This allows us to say that a river originates from these mountains.

Finally, we mention that some implicit facts can be derived directly by an OWL DL reasoner without the need of SWRL rules. For example, from the narratives:

Kumuge Settlement is southeast of the river-cross of Mugua River and Cingshuei River. Kumuge is northwest of Tongment Settlement.

one can infer that Tongment is also southeast of the Mugua and Cingshuei river-cross. This is because the inverse property of northWestOf is southEastOf, and southEastOf is a transitive property. In addition, certain queries to a knowledge base can be formulated as class definitions (*i. e.*, new concepts with additional restrictions). By reasoning about whether the class is empty, an OWL DL reasoner will be able to deliver the results for us. For example, the query “what settlements are west of Sipan and are near by some rivers” can be answered by reasoning about the individuals in the following class:

$$\begin{aligned} & \text{Class WestOfSipanNearByRivers} \equiv \\ & \quad \text{Settlement} \sqcap (\text{westOf has Sipan}) \sqcap (\text{bankOf some River}). \end{aligned}$$

We now return to the illustrative example we use in Section 2:

Sikilian Settlement is located at the left bank of midstream Liwu River, and is about 1.6 kilometers to the northwest of the junction of its branch Wahei-er River. It is opposite to the small terrace slightly east of and below the Syuejiachang Station on Central Cross-Island Road. That is, it is at the mountain belly north of Mantou Mountain.

Geospatial knowledge in the above narrative is now formalized by the following assertions in our implementation:

```

Sikilian : Settlement.
Liwu, Wahei-er, _U : River.
  MidLiwu : Midstream.
Liwu_Wahei-er : RiverCross.
  Mantou : Mountain.
  MantouSide : Mountainside.
Syuejiachang : Station.
  _T : Place.
  _P : Pair.

hasMidstream (Liwu, MidLiwu).
hasInflow (Liwu_Wahei-er, Liwu).
hasInflow (Liwu_Wahei-er, Wahei-er).
hasSide (Mantou, MantouSide).

leftBankOf (Sikilian, Liwu).
locatedAt (Sikilian, MidLiwu).
northWestOf (Sikilian, Liwu_Wahei-er).
northOf (Sikilian, MantouSide).
eastOf (_T, Syuejiachang).
consistOf (_P, Sikilian).
consistOf (_P, _T).
inBetween (_U, _P).
sameAs* (_U, Wahei-er).

```

Note that in the above representation, we have ignored the distance (“... about 1.6 kilometers to ...”), some geographic features (“small terrace” and “Central Cross-Island Road”), and certain details about direction and orientation (“*slightly* east of” and “below”). In order to represent the narrative “it is opposite to the small terrace slightly east of and below the Syuejiachang Station”, we use three auxiliary individuals: `_T` for the place of the small terrace, `_P` for the pair consisting of the place `_T` and the settlement `Sikilian`, and `_U` for the river that separates the pair `_P`. The last assertion states that river `_U` is the same as `Wahei-er` River. This fact is not in, nor can be inferred from, the given narrative (hence marked by a `*`). However, we can learn that `Wahei-er` River separates `Mantou` Mountain (where `Sikilian` Settlement is located) and `Syuejiachang` Station (where the small terrace is near by) by looking at a topographical map of this region. (See Figure 5 in Appendix A.) Therefore we assert `_U` and `Wahei-er` River are the same.

5 Related Works and Discussions

We now compare our approach to narrative geospatial knowledge to other works. First, the traditional GIS approach is to provide the humanity research communities with information systems to visualize spatial data, and to explore, analyze,

and visualize domain-specific knowledge in association with spatial data (often presented as maps). When dealing with domain-specific knowledge of a narrative nature, however, a GIS would need to quantify the narratives before they can be represented and analyzed. This can be a major hurdle.

There are works emphasizing on the summary and analysis of textual information, often in combination with gazetteers and other geospatial metadata, in the production of a geospatial web for the purpose of humanistic research. These works range from using RDBMS (Relational Database Management System) with RDF (Resource Description Framework), to constructing suitable spatial-temporal-object data model [15], to combining data mining techniques with location ontologies for the extraction of mutual relations among places in the cultural heritage domain [12], and to geovisual analytic approach to producing geo-historical context from implicit sources [19]. Approaches based on textual and metadata analyses for the extraction and representation of geographic information, nevertheless, may not rely upon or even require the use of well-developed domain ontologies. On the contrary, our work emphasizes the use of formal expressions for knowledge representation and reasoning, as well as the development of a domain ontology even though the data we are processing is of a narrative nature.

The approach presented here can also be compared to works that developed conceptual primitives and frameworks for the modeling of spatial-temporal activities expressed in natural languages [10, 13, 18]. However, here our narrative geospatial domain is of a much larger scale. In our case, settlements and their movements are expressed in geographic terms, while expressions on spatial-temporal activities tend to be framed in languages about personal or inter-personal space. Also related to our work are standard vocabularies such as GML [11] and SpaceML [7] for expressing qualitative and/or quantitative relationships among geospatial objects. These vocabularies, however, are for the expression of standard geometric, topological, directional, or even temporal relations which are of a technical nature. Our vocabularies are closer to the actual phrases used in narratives which are of a cultural nature.

Finally but not the least, there is a long tradition of conceptual modeling and ontological design for geospatial information, especially when associated with common-sense knowledge and the use of rule-based inference scheme [3, 8, 14]. Our works can be considered as a practical study in this direction, only that we use a standard (Web) ontology language and the associated rule language. Our ontology design is just at the beginning, and it has been developed by only a few in a top-down manner. It currently only handle certain geospatial relations. Issues of collaborative ontology development, geospatial object identity, entity naming and multilingual information processing, maintenance of consistency in the knowledge base, multiple-sourced knowledge acquisition, *etc.* are yet to be dealt with. These are critical issues to look into when building large-scale geospatial knowledge base.

References

1. OWL Web Ontology Language Semantics and Abstract Syntax. <<http://www.w3.org/TR/2004/REC-owl-semantic-20040210>>, 2004.
2. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. <<http://www.w3.org/Submission/SWRL>>, 2004.
3. Guoray Cai. Contextualization of geospatial database semantics for Human–GIS interaction. *Geoinformatica*, 11(2):217–237, June 2007.
4. Eliseo Clementini, Paolino Di Felice, and Daniel Hernández. Qualitative representation of positional information. *Artificial Intelligence*, 95(2):317–356, 1997.
5. Anthony G. Cohn, Brandon Bennett, John Gooday, and Nicholas Mark Gotts. Qualitative spatial representation and reasoning with the region connection calculus. *GeoInformatica*, 1(3):275–316, 1997.
6. Anthony G. Cohn and Shyamanta M. Hazarika. Qualitative spatial representation and reasoning: An overview. *Fundamenta Informaticae*, 46(1–2):1–29, 2001.
7. Matteo Cristani and Anthony G. Cohn. SpaceML: A mark-up language for spatial knowledge. *Journal of Visual Languages and Computing*, 13(1):97–116, 2002.
8. Max J. Egenhofer and David M. Mark. Naive geography. In *Spatial Information Theory: A Theoretical Basis for GIS*, pages 1–15, 1995.
9. Christian Freksa. Using orientation information for qualitative spatial reasoning. In *Spatio-Temporal Reasoning*, pages 162–178, 1992.
10. Kathleen Stewart Hornsby and Naicong Li. Conceptual framework for modeling dynamic paths from natural language expressions. *Transactions in GIS*, 13(s1):27–45, 2009.
11. Open Geospatial Consortium Inc. OpenGIS Geography Markup Language (GML) Encoding Standard, 2007.
12. Tomi Kauppinen, Kimmo Puputti, Panu Paakkarinen, Heini Kuittinen, Jari Väätäinen, and Eero Hyvönen. Learning and visualizing cultural heritage connections between places on the semantic web. In *Proceedings of the Workshop on Inductive Reasoning and Machine Learning on the Semantic Web*, 2009.
13. Alexander Klippel, Alan MacEachren, Prasenjit Mitra, Ian Turton, Anuj Jaiswal, Kean Soon, and Xiao Zhang. Wayfinding Choremes 2.0 — Conceptual primitives as a basis for translating natural into formal language. In Nico Van de Weghe, Roland Billen, Bart Kuijpers, and Peter Bogaert, editors, *International Workshop on Moving Objects From Natural to Formal Language*, Park City, Utah, USA, 2008.
14. Stephan Mäs. Reasoning on spatial relations between entity classes. In Thomas J. Cova, Harvey J. Miller, Kate Beard, Andrew U. Frank, and Michael F. Goodchild, editors, *GIScience*, volume 5266 of *Lecture Notes in Computer Science*, pages 234–248. Springer, 2008.
15. Amit Sheth and Matthew Perry. Traveling the Semantic Web through Space, Time, and Theme. *IEEE Internet Computing*, pages 81–86, 2008.
16. Shou-cheng Liao. The migration and distribution of the East–Sedeq Atayal (I). *Bulletin of the Institute of Ethnology, Academia Sinica*, (44):61–206, Fall 1977.
17. Shou-cheng Liao. The migration and distribution of the East–Sedeq Atayal (II). *Bulletin of the Institute of Ethnology, Academia Sinica*, (45):81–212, Spring 1978.
18. Thora Tenbrink. *Space, Time, and the Use of Language: An Investigation of Relationships*. Mouton de Gruyter, 2007.
19. Brian Tomaszewski. Producing geo-historical context from implicit sources: A geovisual analytics approach. *The Cartographic Journal*, 45(3):165–181, 2008.

A Sources and Maps

In this Appendix, we list the original paragraphs which we have used in this study as the sources of sample narratives on the locations of Taiwan aboriginal settlements. These paragraphs are taken from the comprehensive survey on the migration and distribution of the East-Sedeq Atayal conducted by Shou-cheng Liao (Masaw Mowna) in the 1970's [16, 17]. The original paragraphs are written in Chinese; here we provide the English translations.

The nine paragraphs each describes the location of an aboriginal settlement. These nine settlements were identified by Liao from which people had migrated to Kele Settlement, a “multi-settlement community” in Hualian County at east-coast Taiwan [17, 146–147]. Liao described the locations of the nine settlements by the following:

1. “The land of Batakan . . . is located at the left bank of midstream Liwu River, and is south of Sanjhuei Mountain. It is above the cliff on the riverbank opposite to now Jinheng Station on Central Cross-Island Road.” [16, p. 121]
2. “Bolowan Settlement is located at the right bank of downstream Liwu River, and is north of Ta Mountain. The land is about 1 kilometer to the south of now Sipan Station on Central Cross-Island Road. It extends to the east to an unnamed river (called Puluowan River by the aborigines), and to the west to Yanzihkou.” [16, p. 123]
3. “The land of Kalugi . . . is located at the left bank of Liwu River, and is about 3 kilometers to the west of the junction of its branch Wahei-er River. It is slightly west of now Syuejiachang Station on Central Cross-Island Road.” [16, p. 178]
4. “The land of Lausi . . . is located at the left bank of midstream Liwu River, is around the area slightly northeast of the junction of its branch Laosi (called Pako by the aborigines) River. That is, it is east of and above now Heliou Station on Central Cross-Island Road.” [16, p. 149]
5. “Lodox Settlement . . . is located at the left bank of Liwu River, is about 7–8 kilometers to the north of the junction of its branch Tuosai River, and is southeast of Jiming Mountain. That is, it is at the place now called Shangmeiyuan.” [16, p. 196]
6. “Mogolisi Settlement is located at the right bank of midstream Tuosai River which is a branch of Liwu River. It is southeast of Jhongyangjian Mountain, and belongs to the second terrace of now Shangmeiyuan (now renamed to Jhucunfennong). It extends to the south to an unnamed river (called Mokelisi River by the aborigines), and to the west to the mountain belly opposite to the junction of Tuosai River and its branch Dalama River.” [16, p. 195]
7. “Sikilian Settlement is located at left bank of midstream Liwu River, and is about 1.6 kilometers to the northwest of the junction of its branch Wahei-er River. It is opposite to the small terrace slightly east of and below the Syuejiachang Station on Central Cross-Island Road. That is, it is at the mountain belly north of Mantou Mountain.” [16, p. 178]



Fig. 5. Settlements are positioned according to geospatial narratives about them.

8. “Sowasal Settlement is located at the left bank of downstream Tuosai River, and is on the highland in between the river and the right bank of downstream Sikalahan River which is its branch. It is about 5 kilometers to the northwest of Tiansiang Station on Central Cross–Island Road, is above the riverbank opposite to Hueitouwan, and is at a place now called Lianhua Pond.” [16, p. 155]
9. “Tomowan Settlement is located at the right bank of downstream Liwu River, and is northeast of Ta Mountain. It is at the mountain belly south of now Ning-an Bridge on Central Cross–Island Road. It extends to the east to Changchun Bridge, and to the west to Baisha Bridge.” [16, p. 124]

We also provide in Figure 5 a topographical map of this region. The map is an abstract illustration. It shows the locations of the many rivers, mountains, bridges, stations, and other places that have been mentioned in the narratives and serve as the objects of reference when describing the the locations of the nine settlements. Results from this research are to be applied in situation like this in which geospatial knowledge deduced from the narratives is used to position the locations of the settlements on the map. Figure 6 is a map taken directly from another part of Liao’ ethnography [17, p. 135] on which Kanaongan Settlement is positioned at the northern river bank of the Dacingshuei estuary, in spite of

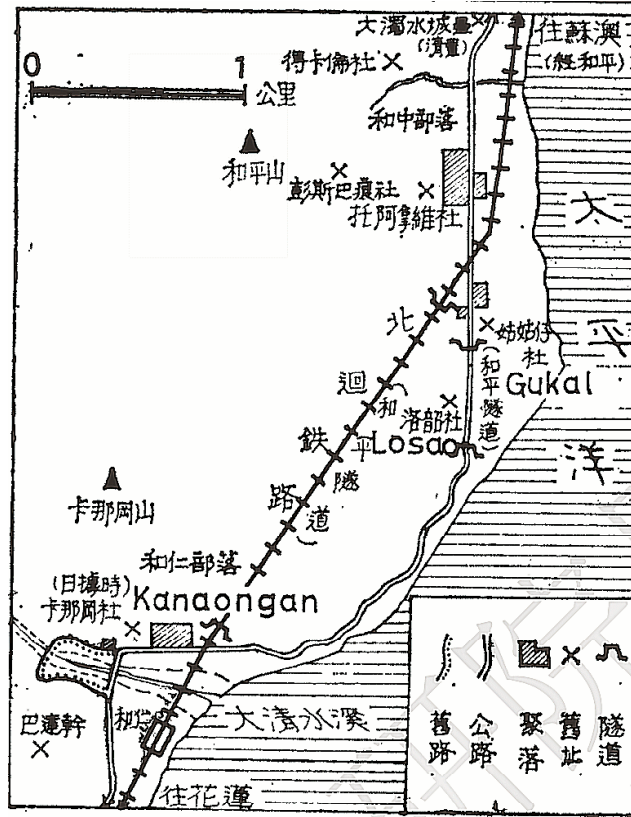


Fig. 6. An ethnographic map showing the location of Kanaongan Settlement.

the narrative in the same ethnography that it is located at the right bank of Dacingshuei River [17, p. 137]. See Section 1 for a discussion.