A Multi-Layer Markup Language for Geospatial Semantic Annotations

L. Moncla (LIUPPA, IAAA), M. Gaio (LIUPPA)

IAAA
Sistemas de Información
Avanzados
Universidad Zaragoza

M. Gaio
mauro.gaio@univ-pau.fr

CONTENTS

OUTLINE

Introduction

Motivation

Two categories of markup languages

- encoding spatial data (GML, KML, . . .)
- annotation of spatial or spatio-temporal information in texts (SpatialML, ISO-Space, . . .)

Generic markup languages

- interchange and encoding of textual information (TEI)

Introduction

Named Entity Recognition

Objective Annotating complex Named Entities

(1) the owner of the restaurant of Neuvic Lake

- Stanford Named Entity Recognizer [1] annotates one entity :
  - 'Neuvic Lake' → **location**
- Open Calais [2] annotates two NEs :
  - 'Neuvic Lake' → **natural feature**
  - 'restaurant of Neuvic Lake' → **facility**
- NER tool of the Cognitive Computation Group [3] of the University of Illinois annotates only one NE :
  - 'Neuvic Lake' → **organization**
- FreeLing [4] annotates one NE :
  - 'Neuvic Lake' → **geographical location**

1. http://nlp.stanford.edu/software/CRF-NER.shtml
2. http://new.opencalais.com/
3. http://cogcomp.cs.illinois.edu/page/demo_view/NER
4. http://nlp.lsi.upc.edu/freeling/

Introduction

Named Entity Recognition

Objective Annotating complex Named Entities

(1) the owner of the restaurant of Neuvic Lake

- Stanford Named Entity Recognizer [1] annotates one entity :
  - 'Neuvic Lake' → **location**
- Open Calais [2] annotates two NEs :
  - 'Neuvic Lake' → **natural feature**
  - 'restaurant of Neuvic Lake' → **facility**
- NER tool of the Cognitive Computation Group [3] of the University of Illinois annotates only one NE :
  - 'Neuvic Lake' → **organization**
- FreeLing [4] annotates one NE :
  - 'Neuvic Lake' → **geographical location**

1. http://nlp.stanford.edu/software/CRF-NER.shtml
2. http://new.opencalais.com/
3. http://cogcomp.cs.illinois.edu/page/demo_view/NER
4. http://nlp.lsi.upc.edu/freeling/

OUTLINE

Generic Markup Language

Expanded Named Entity (ENE)

Proper name :

- Two categories : **pure** and **descriptive**

Expanded Named Entity (ENE) :

- pure proper name associated with descriptive expansion
- definition of several level of overlapping

Generic Markup Language

Expanded Named Entity (ENE)

Proper name :

- Two categories : **pure** and **descriptive**

Expanded Named Entity (ENE) :

- pure proper name associated with descriptive expansion
- definition of several level of overlapping

A generic markup language

Expanded Named Entity (ENE)

Level 0

(2) Paris

Level 1

(3) ville de Paris

Level 2

(4) Parque Natural del Delta del Ebro

Level 3

(5) the owner of the restaurant of Neuvic Lake

A generic markup language

Expanded Named Entity (ENE)

Level 0

(2) Paris

Level 1

(3) ville de Paris

Level 2

(4) Parque Natural del Delta del Ebro

Level 3

(5) the owner of the restaurant of Neuvic Lake

A generic markup language

Expanded Named Entity (ENE)

Level 0

(2) Paris

Level 1

(3) ville de [Paris]

Level 2

(4) Parque Natural del [Delta del [Ebro]]

Level 3

(5) the owner of [the restaurant of [Neuvic] Lake]

A generic markup language

Expanded Named Entity (ENE)

Level 0

(2) Paris

Level 1

(3) ville de [Paris]

Level 2

(4) Parque Natural del [Delta del [Ebro]]

Level 3

(5) the owner of [the restaurant of [Neuvic] Lake]

A generic markup language

Expanded Named Entity (ENE)

Two main categories of ENE :

Absolute : Paris

```
<rs type="expandedName" n="0">
  <name xml:id="n1">
    <w type="NPr">Paris</w>
  </name>
</rs>
```

Relative : au nord de Paris

```
<rs type="relative" n="0">
  <term type="offset" subtype="orientation">
    <w lemma="au" type="PREPDET">au</w>
    <w lemma="nord" type="ADJ">nord</w>
    <w lemma="de" type="PREP">de</w>
  </term>
  <name xml:id="n1">
    <w type="NPr">Paris</w>
  </name>
</rs>
```

A generic markup language

Expanded Named Entity (ENE)

Two main categories of ENE :

Absolute : Paris

```
<rs type="expandedName" n="0">
  <name xml:id="n1">
    <w type="NPr">Paris</w>
  </name>
</rs>
```

Relative : au nord de Paris

```
<rs type="relative" n="0">
  <term type="offset" subtype="orientation">
    <w lemma="au" type="PREPDET">au</w>
    <w lemma="nord" type="ADJ">nord</w>
    <w lemma="de" type="PREP">de</w>
  </term>
  <name xml:id="n1">
    <w type="NPr">Paris</w>
  </name>
</rs>
```

A generic markup language

Expanded Named Entity (ENE)

Hierarchical encapsulation of ENE

ENE level 0 : Paris

```
<rs type="expandedName" n="0">
  <name xml:id="n1">
    <w type="NPr">Paris</w>
  </name>
</rs>
```

ENE level 1 : Cardiff Dowtown

```
<rs type="expandedName" n="1">
  <name xml:id="n3">
    <w type="NPr">Cardiff</w>
  </name>
  <term xml:id="t1" type="N">
    <w lemma="dowtown" type="N">Downtown</w>
  </term>
</rs>
```

A generic markup language

Expanded Named Entity (ENE)

Hierarchical encapsulation of ENE

ENE level 0 : Paris

```
<rs type="expandedName" n="0">
  <name xml:id="n1">
    <w type="NPr">Paris</w>
  </name>
</rs>
```

ENE level 1 : Cardiff Dowtown

```
<rs type="expandedName" n="1">
  <name xml:id="n3">
    <w type="NPr">Cardiff</w>
  </name>
  <term xml:id="t1" type="N">
    <w lemma="dowtown" type="N">Downtown</w>
  </term>
</rs>
```

A generic markup language

Text segmentation

Textual elements	Tagset	Attributes
sentence	`<s>`	
word	`<w>`	lemma, type (POS), subtype (semantics)
punctuation	`<pc>`	force, type
phrase	`<phr>`	type, subType, function

OUTLINE

Towards a geospatial semantic ML

Overview

1. Turn generic elements into specific elements (spatial)

Textual elements	Generic layer	Geospatial layer

spatial and temporal relations
measure expressions
expansion of ENEs
	`<term>`	`<offset>`
		`<measure>`
		`<geogFeat>`

ENE *(level > 0)*	`<rs>`	`<geogName>`
		`<placeName>`
		`<place>`

2. Add attributes for encoding semantics
   - type, subtype, function, ...
3. Add new elements for annotating spatial information
   - `<location>`, `<geo>`, ...
4. Add element for encoding uncertainty
   - `<certainty>`

Towards a geospatial semantic ML

Overview

1. Turn generic elements into specific elements (spatial)

Textual elements	Generic layer	Geospatial layer

spatial and temporal relations
measure expressions
expansion of ENEs
	`<term>`	`<offset>`
		`<measure>`
		`<geogFeat>`

ENE *(level > 0)*	`<rs>`	`<geogName>`
		`<placeName>`
		`<place>`

2. Add attributes for encoding semantics
   - type, subtype, function, ...
3. Add new elements for annotating spatial information
   - `<location>`, `<geo>`, ...
4. Add element for encoding uncertainty
   - `<certainty>`

Towards a geospatial semantic ML

Overview

1. Turn generic elements into specific elements (spatial)

Textual elements	Generic layer	Geospatial layer

spatial and temporal relations
measure expressions
expansion of ENEs
	`<term>`	`<offset>`
		`<measure>`
		`<geogFeat>`

ENE *(level > 0)*	`<rs>`	`<geogName>`
		`<placeName>`
		`<place>`

2. Add attributes for encoding semantics
   - type, subtype, function, ...
3. Add new elements for annotating spatial information
   - `<location>`, `<geo>`, ...
4. Add element for encoding uncertainty
   - `<certainty>`

Towards a geospatial semantic ML

Overview

1. Turn generic elements into specific elements (spatial)

Textual elements	Generic layer	Geospatial layer

spatial and temporal relations
measure expressions
expansion of ENEs
	`<term>`	`<offset>`
		`<measure>`
		`<geogFeat>`

ENE *(level > 0)*	`<rs>`	`<geogName>`
		`<placeName>`
		`<place>`

2. Add attributes for encoding semantics
   - type, subtype, function, ...
3. Add new elements for annotating spatial information
   - `<location>`, `<geo>`, ...
4. Add element for encoding uncertainty
   - `<certainty>`

Towards a geospatial semantic ML

Overview : the <geogFeat> element

Descriptive expansions of ENEs (`<term type="N">`) are turned into `<geogFeat>` elements.

Example :

```
<geogFeat>
  <w lemma="lac" type="N">lac</w>
</geogFeat>
<name>
  <w type="NPr">Grattaleu</w>
</name>
---
<geogFeat>
  <w lemma="torrent" type="N">torrent</w>
</geogFeat>
<w lemma="de" type="PREP">de</w>
<w lemma="la" type="DET">la</w>
<name>
  <w type="NPr">Leisse</w>
</name>
```

Towards a geospatial semantic ML

Overview : the <geogFeat> element

Descriptive expansions of ENEs (<term type="N">) are turned into <geogFeat> elements.

Example :

```
<geogFeat>
  <w lemma="lac" type="N">lac</w>
</geogFeat>
<name>
  <w type="NPr">Grattaleu</w>
</name>
---
<geogFeat>
  <w lemma="torrent" type="N">torrent</w>
</geogFeat>
<w lemma="de" type="PREP">de</w>
<w lemma="la" type="DET">la</w>
<name>
  <w type="NPr">Leisse</w>
</name>
```

Towards a geospatial semantic ML

Overview : the <geogName> element

Expanded Spatial Named Entities (ESNEs)

- `<rs type="expandedName">` are turned into `<geogName>` elements.

Example :

```
<geogName type="S" subtype="RHSE" n="2">
  <geogFeat>
    <w lemma="refuge" type="N">refuge</w>
  </geogFeat>
  <w lemma="du" type="PREPDET">du</w>
  <geogName type="T" subtype="PASS" n="1">
    <geogFeat>
      <w lemma="col" type="N">Col</w>
    </geogFeat>
    <w lemma="de" type="PREP">de</w>
    <w lemma="le" type="DET">la</w>
    <name>
      <w lemma="Vanoise" type="NPr">Vanoise</w>
    </name>
  </geogName>
</geogName>
```

Towards a geospatial semantic ML

Overview : the <geogName> element

Expanded Spatial Named Entities (ESNEs)

- `<rs type="expandedName">` are turned into `<geogName>` elements.

Example :

```
<geogName type="S" subtype="RHSE" n="2">
  <geogFeat>
    <w lemma="refuge" type="N">refuge</w>
  </geogFeat>
  <w lemma="du" type="PREPDET">du</w>
  <geogName type="T" subtype="PASS" n="1">
    <geogFeat>
      <w lemma="col" type="N">Col</w>
    </geogFeat>
    <w lemma="de" type="PREP">de</w>
    <w lemma="le" type="DET">la</w>
    <name>
      <w lemma="Vanoise" type="NPr">Vanoise</w>
    </name>
  </geogName>
</geogName>
```

Towards a geospatial semantic ML

Overview : the `<offset>` element

Spatial relations (Offset)

- `<term type="offset">` are turned into `<offset>` elements.
- Categorization of spatial relations
  - topological, directional and distances

Example :

```
<offset type="adjacency" subtype="near">
  <w lemma="pres" type="ADV">pres</w>
  <w lemma="du" type="PREPDET">des</w>
</offset>
---
<offset type="orientation" subtype="north">
  <w lemma="au" type="PREPDET">au</w>
  <w lemma="nord" type="ADJ">nord</w>
  <w lemma="de" type="PREP">de</w>
</offset>
---
<offset type="direction_final">
  <w lemma="jusque" type="PREP">jusqu</w>
  <w lemma="au" type="PREPDET">au</w>
</offset>
```

Towards a geospatial semantic ML

Overview : the `<offset>` element

Spatial relations (Offset)

- `<term type="offset">` are turned into `<offset>` elements.
- Categorization of spatial relations
  - topological, directional and distances

Example :

```
<offset type="adjacency" subtype="near">
  <w lemma="pres" type="ADV">pres</w>
  <w lemma="du" type="PREPDET">des</w>
</offset>
---
<offset type="orientation" subtype="north">
  <w lemma="au" type="PREPDET">au</w>
  <w lemma="nord" type="ADJ">nord</w>
  <w lemma="de" type="PREP">de</w>
</offset>
---
<offset type="direction_final">
  <w lemma="jusque" type="PREP">jusqu</w>
  <w lemma="au" type="PREPDET">au</w>
</offset>
```

Towards a geospatial semantic ML

Overview : the <placeName> element

<placeName>

- absolute or relative place name
  - <rs type="expandedName"> are turned into <placeName type="absolute"> elements.
  - <rs type="relative"> are turned into <placeName type="relative"> elements.

Example :

```
<placeName type="relative">
  <measure type="distance" unit="m" quantity="200">
    <w type="NUM">200</w>
    <w lemma="mètre" type="N">mètres</w>
  </measure>
  <offset type="orientation" subtype="north">
    <w lemma="au" type="PREPDET">au</w>
    <w lemma="nord" type="ADJ">nord</w>
    <w lemma="de" type="PREP">de</w>
  </offset>
  <name type="place">
    <w type="NPr">Pau</w>
  </name>
</placeName>
```

Towards a geospatial semantic ML

Overview : the <placeName> element

<placeName>

- absolute or relative place name
  - <rs type="expandedName"> are turned into <placeName type="absolute"> elements.
  - <rs type="relative"> are turned into <placeName type="relative"> elements.

Example :

```
<placeName type="relative">
  <measure type="distance" unit="m" quantity="200">
    <w type="NUM">200</w>
    <w lemma="mètre" type="N">mètres</w>
  </measure>
  <offset type="orientation" subtype="north">
    <w lemma="au" type="PREPDET">au</w>
    <w lemma="nord" type="ADJ">nord</w>
    <w lemma="de" type="PREP">de</w>
  </offset>
  <name type="place">
    <w type="NPr">Pau</w>
  </name>
</placeName>
```

Towards a geospatial semantic ML

Overview : the <phr> element

The `subtype` attribute indicates the semantic of the verbal phrase

- motion, perception

The `function` attribute indicates the motion class

- Six motion classes (motion verbs)
  - leave, hit, reach, external, internal, and cross

Example :

```
<phr type="verbal" subtype="motion" function="reach">
  <w lemma="parvenir" type="V" subtype="motion:final">parvenu</w>
  ...
</phr>
```

Towards a geospatial semantic ML

Overview : the <phr> element

The `subtype` attribute indicates the semantic of the verbal phrase

- motion, perception

The `function` attribute indicates the motion class

- Six motion classes (motion verbs)
  - leave, hit, reach, external, internal, and cross

Example :

```
<phr type="verbal" subtype="motion" function="reach">
  <w lemma="parvenir" type="V" subtype="motion:final">parvenu</w>
  ...
</phr>
```

Towards a geospatial semantic ML

Encoding geometric properties of spatial features

- Encoding geometric properties
  - `<location>`,`<geo>`, . . . from the TEI Guidelines
  - GML, KML, . . .

Example :

```
<placeName type="absolute">
  <name>
    <w type="NPr">Pau</w>
    <location>
      <country key="FR" />
      <bloc type="continent" key="EU" />
      <geo>43.301667 -0.368611</geo>
    </location>
  </name>
</placeName>
```

- Mapping via gazetteer unique identifiers
  - RDF identifiers to interlink with resources of the Web of Linked Data

Towards a geospatial semantic ML

Encoding geometric properties of spatial features

- Encoding geometric properties
  - `<location>`,`<geo>`, . . . from the TEI Guidelines
  - GML, KML, . . .

Example :

```
<placeName type="absolute">
  <name>
    <w type="NPr">Pau</w>
    <location>
      <country key="FR" />
      <bloc type="continent" key="EU" />
      <geo>43.301667 -0.368611</geo>
    </location>
  </name>
</placeName>
```

- Mapping via gazetteer unique identifiers
  - RDF identifiers to interlink with resources of the Web of Linked Data

Towards a geospatial semantic ML

Encoding geometric properties of spatial features

- Encoding geometric properties
  - `<location>`,`<geo>`, . . . from the TEI Guidelines
  - GML, KML, . . .

Example :

```
<placeName type="absolute">
  <name>
    <w type="NPr">Pau</w>
    <location>
      <country key="FR" />
      <bloc type="continent" key="EU" />
      <geo>43.301667 -0.368611</geo>
    </location>
  </name>
</placeName>
```

- Mapping via gazetteer unique identifiers
  - RDF identifiers to interlink with resources of the Web of Linked Data

Towards a geospatial semantic ML

Indication of uncertainty

Attributes for `<certainty>` tag

Attribute name	Description/value
`target`	URI data pointer
`locus`	*name, start, end, location, value*
`assertedValue`	alternative value
`degree`	degree of confidence

Example :

```
<placeName xml:id="pl1">
  <certainty target="#pl1" locus="name" assertedValue="rs"
      degree="0.6" />
  <name>
    <w type="NPr">Paris</w>
  </name>
</placeName>
```

Towards a geospatial semantic ML

Indication of uncertainty

Attributes for `<certainty>` tag

Attribute name	Description/value
`target`	URI data pointer
`locus`	*name, start, end, location, value*
`assertedValue`	alternative value
`degree`	degree of confidence

Example :

```
<placeName xml:id="pl1">
  <certainty target="#pl1" locus="name" assertedValue="rs"
      degree="0.6" />
  <name>
    <w type="NPr">Paris</w>
  </name>
</placeName>
```

Towards a geospatial semantic ML

Examples

French cabane du Tracuit

```
<geogName>
  <geogFeat>
    <w lemma="cabane" type="N">cabane</w>
  </geogFeat>
  <w lemma="de" type="PREPDET">du</w>
  <name>
    <w type="NPr">Tracuit</w>
  </name>
</geogName>
```

German Tracuithütte

```
<geogName>
  <w type="NPr">
    <m xml:id="m1">Tracuit</m>
    <m xml:id="m2" baseForm="hütte">hütte</m>
  </w>
  <geogFeat><ptr target="#m2" /></geogFeat>
  <name><ptr target="#m1" /></name>
</geogName>
```

Towards a geospatial semantic ML

Examples

French cabane du Tracuit

```
<geogName>
  <geogFeat>
    <w lemma="cabane" type="N">cabane</w>
  </geogFeat>
  <w lemma="de" type="PREPDET">du</w>
  <name>
    <w type="NPr">Tracuit</w>
  </name>
</geogName>
```

German Tracuithütte

```
<geogName>
  <w type="NPr">
    <m xml:id="m1">Tracuit</m>
    <m xml:id="m2" baseForm="hütte">hütte</m>
  </w>
  <geogFeat><ptr target="#m2" /></geogFeat>
  <name><ptr target="#m1" /></name>
</geogName>
```

Towards a geospatial semantic ML

Examples


```
<w lemma="on" type="PRO">On</w>
<phr xml:id="phr1" type="verb" subtype="motion" function="reach">
  <w lemma="parvenir" type="V" subtype="motion_final">parvient</w>
  <w lemma="au" type="PREPDET">au</w>
  <placeName xml:id="pn1" type="absolute">
    <location>
      <geo xml:id="geo1">51.969604 -2.893146</geo>
      <country key="FR" />
      <bloc type="continent" key="EU" />
      <certainty target="#geo1" locus="value" degree="1.0"/>
    </location>
    <w lemma="du" type="PREPDET">du</w>
    <geogName type="T" subtype="PASS" n="1">
      <geogFeat>
        <w lemma="col" type="N">Col</w>
      </geogFeat>
      <w lemma="de" type="PREP">de</w>
      <w lemma="le" type="DET">la</w>
      <name>
        <w lemma="Vanoise" type="NPr">Vanoise</w>
      </name>
    </geogName>
  </placeName>
</phr>
```

OUTLINE

Thank you for your attention

CONTACT

Ludovic Moncla (LIUPPA, IAAA)
ludovic.moncla@univ-pau.fr

Mauro Gaio (LIUPPA)
mauro.gaio@univ-pau.fr

This work has been supported by