Geo-Net-PT 02

1. BASIC INFORMATION

1.1 Lexicon type (wordform, explanatory, terminological lexicon, wordnet, etc.)

Geo-Net-PT 02 is a public Geospatial Ontology of Portugal (see Chaves *et al.*, 2007), a computational resource (see Rodrigues *et al.*, 2006 and Rodrigues, 2009) for applications demanding geographic information about Portugal, and contains 701,209 concepts stored in a GKB system, most of them administrative features and place names. Some of these concepts have additional types to ease the reuse in the Web of Data: 390,664 administrative and physical features and footprints are classified as *geo:SpatialThing* and 23,666 network features are classified as *sioc:Space*. Geo-Net-PT 02 identifies 22,980 owners of domains, which are classified as *sioc:User* instances. The administrative and physical features are classified by 81 feature types. Postal code, street layout and settlement are the most common feature types found in the geo-administrative domain. Hydrography and touristic resources, such as museums and hotels, are the most common feature types found in the geo-physical domain.
The Geo-Net-PT 02 is an extension to the Geo-Net-PT 01 ontology presented in Chaves *et al.* (2005). It respects the recommended international standards for publishing ontologies (for more about the resource, see `http://dmir.inesc-id.pt/project/Geo-Net-PT_02_in_English` and Lopez-Pellicer *et al.* (2010).
This resource was created by the XLDB Team of the University of Lisbon, Faculty of Sciences, under the GREASE (Geographic Reasoning for Search Engines)[1] project (see Lopez-Pellicer *et al.*, 2009), and contains all the geographic administrative data of Portugal (*distritos*, *concelhos* and *ruas*, among others), and domains of websites of the Portuguese Web and their geographic scopes. Currently, it is maintained by the REACTION project (`http://dmir.inesc-id.pt/project/Reaction`).

Geo-Net-PT 02 adds the following feature types (classes of geographic entities: 5656, in total) to the physical domain:

Geographic Entity Type	Count	Geographic Entity Type	Count
Rio (River)	2421	Recurso turístico (Touristic resource)	84
Praia (Beach)	558	Biótopo	58
Museu (Museum)	507	Linha férrea (Railroad)	38
Sítio arqueológico (Archeological site)	414	Área protegida (Protected area)	31

[1] Available at `http://xldb.fc.ul.pt/wiki/Grease`.

Hotel	381	Serra (Mountain range)	25
Região Natural (Natural region)	304	Marina	26
Castelo (Castle)	256	Parque Natural (Natural Park)	12
Nascente (Spring)	220	Estuário (Estuary)	8
Aldeia histórica (Historic village)	217	Monumento Natural (Natural monument)	5
Albufeira (Reservoir)	90	Parque Nacional (National Park)	1

Besides the physical geography information, the data of all geographic domains have been annotated with additional information:

- multilingual names (Lisboa[PT] is also Lisbon[EN], Lisbonne[FR] e Lissabon[DE])
- footprints for both administrative and physical features
- provenance information for each datum included in the ontology.

The resource also provides:
1. An alignment with Yahoo! GeoPlanet (TM) (see Ferreira *et al.*, 2010), between features in the Administrative with "Where On Earth Identifiers" (WOEID) from GeoPlanet (TM). The alignment produced is relatively complete, containing almost all the administrative divisions of Portugal (districts, municipalities and civil parishes) as well as a high number of other towns:

Feature Type	Coverage (%)
Distrito (District)	100.0
Concelho (Municipality)	100.0
Ilha (Islands)	100.0
Freguesia (Civil Parish)	83.7
Localidade (Settlement)	27.6
Zona (Zone)	13.0

2. Geo-CHAVE-PT, a collection of Portuguese news articles with the mention of locations, in the territory of Portugal, annotated with a reference to Geo-Net-PT. The articles were selected from the CHAVE collection (available at `www.linguateca.pt/chave/`) which contains Portuguese news articles published between 1994 and 1995, from *CETEMPúblico* (106,827 documents) and *Folha de São Paulo* (103,919 documents) (see Chaves, 2009). The articles were scanned to map each identified toponym to the geographic concepts, that it might represent in Geo-Net-PT. A total of 195 news articles were selected to be manually annotated. Only toponyms which are part of the portuguese territory and have a geographic concept in Geo-Net-PT were considered:

Categories	Articles	Geographic Entities
Local	124	972
National	35	218
Society	14	124
Diverse	3	26
Economy	4	21
Sport	4	17
Science	4	24
Culture	7	61
Total	**195**	**1463**

The table above shows the diversity of geographic feature types present in the annotated texts. The feature type *Localidade* is the one with most instances:

Type of Geographic Entity	Number of Entities	Percentage
Localidade	774	52.9 %
Concelho	253	17.29 %
Freguesia	108	7.38 %
País	98	6.7 %
Zona	29	1.98 %
Rua	29	1.98 %
Distrito	25	1.71 %
Praça	24	1.64 %
NUT2	24	1.64 %
Província	21	1.44 %

Avenida	13	0.89 %
NUT3	9	0.62 %
Cais	9	0.62 %
Largo	7	0.48 %
Estrada	7	0.48 %
Travessa	5	0.34 %
Castelo (Physical)	5	0.34 %
Ilha	4	0.27 %
Região (Physical)	3	0.21 %
Planeta (Physical)	3	0.21 %
Outro	3	0.21 %
Rio (Physical)	2	0.14 %
Região	2	0.14 %
Calçada	2	0.14 %
Bairro	2	0.14 %
Serra (Physical)	1	0.07 %
Parque	1	0.07 %

1.2 Representation of the lexicon (flat files, database, markup)

Geo-Net-PT is available in the formats of the W3C (RDF, TTL, N3). For more information about RDF see `http://www.w3.org/TR/REC-rdf-syntax/`; TTL see `http://www.w3.org/TeamSubmission/turtle/`; and N3 see `http://www.w3.org/DesignIssues/Notation3.html`.

1.3 Character encoding

The characters are in UTF-8 code.

2. ADMINISTRATIVE INFORMATION

2.1 Contact person (name, address, affiliation, position, telephone, fax, e-mail)

Name: Mário J. Gaspar da Silva
Address:
INESC-ID - Instituto de Engenharia de Sistemas e Computadores Investigação e Desenvolvimento em Lisboa
Rua Alves Redol, 9
1000-029 Lisboa
Position: Head researcher
Affiliation: INESC-ID - Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento em Lisboa
Telephone: + 351 21 3100300
Fax: + 351 21 3145843
E-mail: mjs@inesc-id.pt

2.2 Delivery medium (if relevant; description of the content of each piece of medium)

The resource will be available on the META-SHARE platform.

2.3 Copyright statement and information on IPR

3. TECHNICAL INFORMATION

3.1 Directories and files

Contents of the Geo-Net-PT distribution:

A.

I) The definition pf the structure of Geo-Net-PT vocabulary. It is available in TURTLE and RDF/XML formats.
a) geo-net.owl (8.2K, format RDF/XML)
b) geo-net.ttl (5.2K, format TURTLE)

II) Geo-Net-PT vocabulary is an extension of the Geo-Net vocabulary. The goal of this extension is to annotate specific characteristics of the Geo-Net-PT data.
It is available in Turtle and RDF/XML formats.
a) geo-net-pt.owl (5.8K, format RDF/XML)

b) geo-net-pt.ttl (3.9K, format TURTLE)

III) The feature types of Geo-Net-PT 02. This collection uses SKOS terms (see more at: `http://www.w3.org/TR/skos-reference/`). It is available in TURTLE and RDF/XML formats.

a) geo-net-pt-02-skos.rdf (73K, format RDF/XML)
b) geo-net-pt-02-skos.ttl (44K, format TURTLE)

IV) The instances of Geo-Net-PT 02. It is available in TURTLE, RDF/XML and N3 formats.

a) geo-net-pt-02-plain.n3 (565M, format N3)
b) geo-net-pt-02-plain.rdf (580M, format RDF/XML)
c) geo-net-pt-02-plain.ttl (418M, format TURTLE)

V) The instances of Geo-Net-PT 02. Includes provenance of relations annotated using escrows objects. It is available in TURTLE, RDF/XML and N3 formats.

a) geo-net-pt-02-escrow.n3 (696M, format N3)
b) geo-net-pt-02-escrow.rdf (827M, format RDF/XML)
c) geo-net-pt-02-escrow.ttl (578M, format TURTLE)

VI) The instances of Geo-Net-PT 02. Includes provenance of relations annotated using escrows objects. It is available in TURTLE, RDF/XML and N3 formats.

a) geo-net-pt-02-reified.n3 (686M, format N3)
b) geo-net-pt-02-reified.rdf (867M, format RDF/XML)
c) geo-net-pt-02-reified.ttl (557M, format TURTLE)

VII) The alignment files of Geo-Net-PT 02. with Yahoo! GeoPlanet (TM) are available in TURTLE, RDF/XML and N3 formats.

a) GeoNetPT02-yahoo-alignment-places.N3 (6.6M, format N3)
b) GeoNetPT02-yahoo-alignment-places.RDF (6.6M, format RDF/XML)
c) GeoNetPT02-yahoo-alignment-places.TTL (3.9M, format TURTLE)
d) GeoNetPT02-yahoo-alignment-postal-codes.N3 (31M, format N3)
e) GeoNetPT02-yahoo-alignment-postal-codes.RDF (37M, format RDF/XML)
f) GeoNetPT02-yahoo-alignment-postal-codes.TTL (26M, format TURTLE)

In addition to these 21 files from <Geo-Net-PT 02> archive, there are on HTML file with the Geo-Net-Pt 02 organization and a brief description, and two text files describing the license, in English and Portuguese. There is also the <Geo-CHAVE-PT01> archive with 199 XML files from *CETEMPúblico* (available at `http://www.linguateca.pt/cetempublico/`), only news with Portuguese references, and one README file.

3.2 Data structure of an entry

The first three items (see 3.1) describes the annotation in terms of the structure, semantic relations, and types needed for the constrution of the resource. For the remaining items, see above the different stuctures per formats:

IV) The instances of Geo-Net-PT 02. It is available in TURTLE, RDF/XML and N3 formats.

a) N3 format:

```
gnpt02:amor-AF1681 a        geo:SpatialThing .
gnpt02:amor-AF1681 a        gn:Feature .
gnpt02:amor-AF1681 dcterms:alternative "100901"^^xsd:string .
gnpt02:amor-AF1681 dcterms:title "Amor"@pt .
gnpt02:amor-AF1681 rdfs:label "Amor (Freguesia)"@pt .
gnpt02:amor-AF1681 gn:inDomain gnpt02:GeoAdministrative .
gnpt02:amor-AF1681 gn:type  gnpt02:freguesia-ATFRG .
gnpt02:amor-AF1681 gn:lineage gnpt02:SRC-WIKI-FRE .
gnpt02:amor-AF1681 gnpt:preferred gnpt02:amor-pt .
gnpt02:amor-AF1681 gn:footprint gnpt02:AFP2487 .
gnpt02:amor-AF1681 gnpt:identifier gnpt02:100901-DICOFRE .
gnpt02:amor-AF1681 gnpt:isAdjacentTo gnpt02:marinha_grande-AF160 .
{...}
gnpt02:amor-AF1681 gnpt:isAdjacentTo gnpt02:marinha_grande-AF1914 .
gnpt02:amor-AF1681 gnpt:isPartOf gnpt02:leiria-AF143 .
gnpt02:amor-pt a        gn:PlaceName .
gnpt02:amor-pt dcterms:title "Amor"@pt .
gnpt02:amor-pt rdfs:label "Amor"@pt .
gnpt02:amor-pt gn:inDomain gnpt02:GeoAdministrative .
gnpt02:amor-pt gn:inDomain gnpt02:GeoPhysical .
gnpt02:amor-pt gn:lemma "Amor"^^xsd:string .
gnpt02:amor-pt gn:languageCode "pt"^^xsd:string .
```

b) RDF/XML format:

```
<rdf:Description rdf:about="#amor-AF1681">
 <rdf:type rdf:resource="http://www.w3.org/2003/01/geo/wgs84_pos#SpatialThing"/>
 <rdf:type rdf:resource="http://xldb.di.fc.ul.pt/xldb/publications/2009/10/geo-net#Feature"/>
 <dcterms:alternative rdf:datatype="http://www.w3.org/2001/XMLSchema#string">100901</dcterms:alternative>
 <dcterms:title xml:lang="pt">Amor</dcterms:title>
 <rdfs:label xml:lang="pt">Amor (Freguesia)</rdfs:label>
 <gn:inDomain rdf:resource="#GeoAdministrative"/>
 <gn:type rdf:resource="#freguesia-ATFRG"/>
 <gn:lineage rdf:resource="#SRC-WIKI-FRE"/>
 <gnpt:preferred rdf:resource="#amor-pt"/>
 <gn:footprint rdf:resource="#AFP2487"/>
 <gnpt:identifier rdf:resource="#100901-DICOFRE"/>
 <gnpt:isAdjacentTo rdf:resource="#marinha_grande-AF160"/>
 {...}
 <gnpt:isAdjacentTo rdf:resource="#marinha_grande-AF1914"/>
 <gnpt:isPartOf rdf:resource="#leiria-AF143"/>
</rdf:Description>
```

c) TURTLE format:

```
gnpt02:amor-AF1681
    a       gn:Feature , geo:SpatialThing ;
    rdfs:label "Amor (Freguesia)"@pt ;
    dcterms:alternative "100901"^^xsd:string ;
    dcterms:title "Amor"@pt ;
    gn:footprint gnpt02:AFP2487 ;
    gn:inDomain gnpt02:GeoAdministrative ;
    gn:lineage gnpt02:SRC-WIKI-FRE ;
    gn:type gnpt02:freguesia-ATFRG ;
    gnpt:identifier gnpt02:100901-DICOFRE ;
    gnpt:isAdjacentTo  gnpt02:marrazes-AF1694 , gnpt02:marinha_grande-AF1914 , gnpt02:ortigosa-AF1698 , gnpt02:barosa-
AF1685 , gnpt02:marinha_grande-AF160 , gnpt02:regueira_de_pontes-AF1701 , gnpt02:mon        te_real-AF1697 ;
    gnpt:isPartOf gnpt02:leiria-AF143 ;
    gnpt:preferred gnpt02:amor-pt .
```

V) The instances of Geo-Net-PT 02. It includes provenance of relations annotated using escrows objects. It is available in TURTLE, RDF/XML and N3 formats.

a) N3 format:

```
gnpt02:amor-AF1681 a        geo:SpatialThing .
gnpt02:amor-AF1681 a        gn:Feature .
gnpt02:amor-AF1681 dcterms:alternative "100901"^^xsd:string .
gnpt02:amor-AF1681 dcterms:title "Amor"@pt .
gnpt02:amor-AF1681 rdfs:label "Amor (Freguesia)"@pt .
gnpt02:amor-AF1681 gn:inDomain gnpt02:GeoAdministrative .
gnpt02:amor-AF1681 gn:type  gnpt02:freguesia-ATFRG .
gnpt02:amor-AF1681 gn:lineage gnpt02:SRC-WIKI-FRE .
gnpt02:amor-AF1681 gnpt:preferred gnpt02:amor-pt .
gnpt02:amor-AF1681 gn:footprint gnpt02:AFP2487 .
gnpt02:amor-AF1681 gnpt:identifier _:b106112 .
gnpt02:amor-AF1681 gnpt:isAdjacentTo _:b106113 .
{...}
gnpt02:amor-AF1681 gnpt:isPartOf _:b106126 .
gnpt02:amor-pt a        gn:PlaceName .
gnpt02:amor-pt dcterms:title "Amor"@pt .
gnpt02:amor-pt rdfs:label "Amor"@pt .
gnpt02:amor-pt gn:inDomain gnpt02:GeoAdministrative .
gnpt02:amor-pt gn:inDomain gnpt02:GeoPhysical .
gnpt02:amor-pt gn:lemma "Amor"^^xsd:string .
gnpt02:amor-pt gn:languageCode "pt"^^xsd:string .
```

b) RDF/XML format:

```
<rdf:Description rdf:about="#amor-AF1681">
    <rdf:type rdf:resource="http://www.w3.org/2003/01/geo/wgs84_pos#SpatialThing"/>
    <rdf:type rdf:resource="http://xldb.di.fc.ul.pt/xldb/publications/2009/10/geo-net#Feature"/>
    <dcterms:alternative rdf:datatype="http://www.w3.org/2001/XMLSchema#string">100901</dcterms:alternative>
    <dcterms:title xml:lang="pt">Amor</dcterms:title>
    <rdfs:label xml:lang="pt">Amor (Freguesia)</rdfs:label>
    <gn:inDomain rdf:resource="#GeoAdministrative"/>
    <gn:type rdf:resource="#freguesia-ATFRG"/>
    <gn:lineage rdf:resource="#SRC-WIKI-FRE"/>
    <gnpt:preferred rdf:resource="#amor-pt"/>
```

```
    <gn:footprint rdf:resource="#AFP2487"/>
    <gnpt:identifier rdf:nodeID="A592259"/>
    <gnpt:isAdjacentTo rdf:nodeID="A592260"/>
    {...}
    <gnpt:isAdjacentTo rdf:nodeID="A521278"/>
    <gnpt:isPartOf rdf:nodeID="A592271"/>
  </rdf:Description>
```

c) TURTLE format:

```
gnpt02:amor-AF1681
    a       gn:Feature , geo:SpatialThing ;
    rdfs:label "Amor (Freguesia)"@pt ;
    dcterms:alternative "100901"^^xsd:string ;
    dcterms:title "Amor"@pt ;
    gn:footprint gnpt02:AFP2487 ;
    gn:inDomain gnpt02:GeoAdministrative ;
    gn:lineage gnpt02:SRC-WIKI-FRE ;
    gn:type gnpt02:freguesia-ATFRG ;
    gnpt:identifier
          [ a       gnpt:Escrow ;
            gn:inDomain gnpt02:GeoAdministrative ;
            gn:lineage gnpt02:SRC-CAOP ;
            gnpt:escrowFor gnpt02:100901-DICOFRE
          ] ;
    gnpt:isAdjacentTo
          [ a       gnpt:Escrow ;
            gn:inDomain gnpt02:GeoAdministrative ;
            gn:lineage gnpt02:SRC-AR-ADJ ;
            gnpt:escrowFor gnpt02:barosa-AF1685
          ] ;
  {...}
    gnpt:isAdjacentTo
          [ a       gnpt:Escrow ;
            gn:inDomain gnpt02:GeoAdministrative ;
            gn:lineage gnpt02:SRC-AR-ADJ ;
            gnpt:escrowFor gnpt02:marinha_grande-AF160
          ] ;
    gnpt:isPartOf
          [ a       gnpt:Escrow ;
            gn:inDomain gnpt02:GeoAdministrative ;
            gn:lineage gnpt02:SRC-WIKI-FRE ;
            gnpt:escrowFor gnpt02:leiria-AF143
          ] ;
    gnpt:preferred gnpt02:amor-pt .
```

VI) The instances of Geo-Net-PT 02. It includes provenance of relations annotated using escrows objects. It is available in TURTLE, RDF/XML and N3 formats.

a) N3 format:

```
gnpt02:amor-AF1681 a       geo:SpatialThing .
gnpt02:amor-AF1681 a       gn:Feature .
gnpt02:amor-AF1681 dcterms:alternative "100901"^^xsd:string .
gnpt02:amor-AF1681 dcterms:title "Amor"@pt .
gnpt02:amor-AF1681 rdfs:label "Amor (Freguesia)"@pt .
gnpt02:amor-AF1681 gn:inDomain gnpt02:GeoAdministrative .
gnpt02:amor-AF1681 gn:type  gnpt02:freguesia-ATFRG .
```

```
gnpt02:amor-AF1681 gn:lineage gnpt02:SRC-WIKI-FRE .
gnpt02:amor-AF1681 gnpt:preferred gnpt02:amor-pt .
gnpt02:amor-AF1681 gn:footprint gnpt02:AFP2487 .
gnpt02:amor-pt a        gn:PlaceName .
gnpt02:amor-pt dcterms:title "Amor"@pt .
gnpt02:amor-pt rdfs:label "Amor"@pt .
gnpt02:amor-pt gn:inDomain gnpt02:GeoAdministrative .
gnpt02:amor-pt gn:inDomain gnpt02:GeoPhysical .
gnpt02:amor-pt gn:lemma "Amor"^^xsd:string .
gnpt02:amor-pt gn:languageCode "pt"^^xsd:string .
```

b) RDF/XML format:

```
  <rdf:Description rdf:about="#amor-AF1681">
    <rdf:type rdf:resource="http://www.w3.org/2003/01/geo/wgs84_pos#SpatialThing"/>
    <rdf:type rdf:resource="http://xldb.di.fc.ul.pt/xldb/publications/2009/10/geo-net#Feature"/>
   <dcterms:alternative rdf:datatype="http://www.w3.org/2001/XMLSchema#string">100901</dcterms:alternative>
    <dcterms:title xml:lang="pt">Amor</dcterms:title>
    <rdfs:label xml:lang="pt">Amor (Freguesia)</rdfs:label>
    <gn:inDomain rdf:resource="#GeoAdministrative"/>
    <gn:type rdf:resource="#freguesia-ATFRG"/>
    <gn:lineage rdf:resource="#SRC-WIKI-FRE"/>
    <gnpt:preferred rdf:resource="#amor-pt"/>
    <gn:footprint rdf:resource="#AFP2487"/>
  </rdf:Description>
```

c) TURTLE format:

```
gnpt02:amor-AF1681
      a       gn:Feature , geo:SpatialThing ;
      rdfs:label "Amor (Freguesia)"@pt ;
      dcterms:alternative "100901"^^xsd:string ;
      dcterms:title "Amor"@pt ;
      gn:footprint gnpt02:AFP2487 ;
      gn:inDomain gnpt02:GeoAdministrative ;
      gn:lineage gnpt02:SRC-WIKI-FRE ;
      gn:type gnpt02:freguesia-ATFRG ;
      gnpt:preferred gnpt02:amor-pt .
```

VII) The alignment files of Geo-Net-PT 02 with Yahoo! GeoPlanet (TM), available in N3, RDF/XML and TURTLE formats.

a) N3 format:

```
<http://where.yahooapis.com/v1/place/736158> <http://purl.org/dc/terms/#identifier> "736158" .
<http://where.yahooapis.com/v1/place/736158> <http://www.w3.org/2004/02/skos/core#prefLabel> "Amor (Town)" .
```

b) RDF/XML format:

```
  <rdf:Description rdf:about="http://xldb.di.fc.ul.pt/xldb/publications/2009/10/geo-net-pt-02#amor-AF167576">
```

```
    <skos:closeMatch>
      <rdf:Description rdf:about="http://where.yahooapis.com/v1/place/736158">
        <dc:identifier>736158</dc:identifier>
        <skos:prefLabel>Amor (Town)</skos:prefLabel>
      </rdf:Description>
    </skos:closeMatch>
  </rdf:Description>
```

c) TURTLE format:

```
<http://where.yahooapis.com/v1/place/736158>
      dc:identifier "736158" ;
      skos:prefLabel "Amor (Town)" .
```

d) N3 format:

```
 <http://where.yahooapis.com/v1/place/12867496> <http://purl.org/dc/terms/#identifier> "12867496" .
 <http://where.yahooapis.com/v1/place/12867496> <http://www.w3.org/2004/02/skos/core#prefLabel> "2400 (Postal Code)" .
```

e) RDF/XML format:

```
  <rdf:Description rdf:about="http://xldb.di.fc.ul.pt/xldb/publications/2009/10/geo-net-pt-02#171542">
    <skos:broaderMatch>
      <rdf:Description rdf:about="http://where.yahooapis.com/v1/place/12867496">
        <dc:identifier>12867496</dc:identifier>
        <skos:prefLabel>2400 (Postal Code)</skos:prefLabel>
      </rdf:Description>
    </skos:broaderMatch>
  </rdf:Description>
```

f) TURTLE format:

```
 <http://where.yahooapis.com/v1/place/12867496>
      dc:identifier "12867496" ;
      skos:prefLabel "2400 (Postal Code)" .
```

B. **Geo-CHAVE-PT01:**

```
−<root>
   <title>PUBLICO-19940401-100</title>
 −<text>
     Pesetas falsas A polícia de
     <LOCAL_GeoNetPT02 f_id="32033"
     t_id="LOC:ADM">Beja</LOCAL_GeoNetPT02>
     deteve na terça-feira, num banco da cidade, dois homens que tentavam trocar duas
     notas de 5000 pesetas falsas Ambos acabaram por ser detidos depois de o gerente
     bancário ter constado que as notas eram contrafeitas e ter chamado a polícia. No
     últimos meses, em diversas localidades do
     <LOCAL_GeoNetPT02 f_id="418734" t_id="PRO:ADM">Baixo
     Alentejo</LOCAL_GeoNetPT02>
     , foram descobertas notas falsas de cinco e dez contos, tendo a Polícia Judiciária, que
     procede a averiguações, chegado a deter um homem residente na
     <LOCAL_GeoNetPT02 f_id="40109"
     t_id="LOC:ADM">Vidigueira</LOCAL_GeoNetPT02>
     .
   </text>
 </root>
```

3.3 Lexicon size (nmb. of lexical items, KB occupied on disk)

The ontology is composed by 701,209 concepts with 941.6 MB compressed (6.2 GB uncompressed) for disk storage. Concerning the Geo-CHAVE-PT, there are 1463 entities across 195 news articles, plus four articles with one toponym each, with 378.3 KB compressed (748.4 KB uncompressed) for disk storage.

4. CONTENT INFORMATION

4.1 The natural language(s) of the lexicon

The language of the resource is Portuguese.

4.2 Entry Type

The entries are a geographycal entities in three different formats (see 3.2).

4.3 Attributes and their values

Concerning to Geo-Net-PT 02 annotation, for plain, escrow, and reified formats, for each entry, there is a set of attributes: *label*, *alternative*, *title*, *footprint*, *domain*, *lineage/resource*, *type*, and *preferred*. See below, a TURTLE sample:

```
gnpt02:amor-AF1681
    a       gn:Feature , geo:SpatialThing ;
    rdfs:label "Amor (Freguesia)"@pt ;
    dcterms:alternative "100901"^^xsd:string ;
    dcterms:title "Amor"@pt ;
    gn:footprint gnpt02:AFP2487 ;
    gn:inDomain gnpt02:GeoAdministrative ;
    gn:lineage gnpt02:SRC-WIKI-FRE ;
    gn:type gnpt02:freguesia-ATFRG ;
```

gnpt:preferred *gnpt02:amor-pt* .

For yahoo format, places and postal codes, there are three fields: *yahoo api url*, *identifier*, *prefLabel* (place/postal code). Below, an example in the TURTLE format:

<http://where.yahooapis.com/v1/place/736158>
dc:identifier *"736158"* ;
skos:prefLabel *"Amor (Town)"* .

Note that Geo-CHAVE-PT was annotated by human annotators which picked, from the set of possible geographic concepts with the toponym identified by REMBRANDT (for more informations, look at `http://dmir.inesc-id.pt/project/Geo-Net-PT_02_in_English`), the one that they reasoned that was being referenced. They also performed some manual cleaning tasks: during the annotation process, some toponyms not detected by REMBRANDT were manually annotated and entities that in the given context don't represent a toponym, were discarded.
The toponyms are annotated with the <LOCAL_GeoNetPT02> tag which contains two attributes, *f_id* which corresponds to the Geo-Net-PT02 identifier and *t_id* which corresponds to the feature type (see 3.1, B).

4.4 Coverage of the lexicon

The ontology works on the geographic language.

4.5 Intended application of the lexicon

Geo-Net-PT is specially useful for applications that need to obtain the descriptions of geographic resources matching a given name or resource type.

4.6 POS assignment

Not applicable.

4.7 Reliability (automatically/manually constructed)

Geo-Net-PT is a geospatial ontology representing the Portuguese territory and the relations between the several locations within in. Yahoo! GeoPlanet (TM) is a geospatial ontology that covers the whole world. To diminish the effects of repeated information, we propose an alignment between the administrative part of these two ontologies based on name similarity and physical closeness. Results: After running the matching process, 16,814 matches were found, corresponding to 33% of the considered features in Geo-Net-PT and to 75% in Yahoo! GeoPlanet (TM). Among these, there are correct matches for each of the 18 districts and 308 municipalities. Only 1% of the matches failed the validation process. This alignment represents a step further for the exploitation of geospatial ontologies, since it enables the mapping of annotation in documents and other resources from GeoPlanet (TM), a broad ontology, to Geo-Net-PT, a specific ontology for Portuguese geography.

5. RELEVANT REFERENCES AND OTHER INFORMATION

Rodrigues, C., Chaves, M., and Silva, M. J. (2006). ”Uma Representação Ontológica da Geografia Física de Portugal.” In *IX Encontro de Utilizadores de Informação Geográfica - ESIG 2006 Oeiras*, Portugal.

Rodrigues, C. (2009). *An Ontology of the Physical Geography of Portugal Master Thesis.* Faculty of Sciences, University of Lisbon.

Lopez-Pellicer, F. J., Chaves, M., Rodrigues, C., and Silva, M. J. (2009). ”Geographic Ontologies Production in Grease-II.” In *Technical Report. TR 09-18*. University of Lisbon, Faculty of Sciences, LASIGE.

Lopez-Pellicer, F. J., Silva, M. J., and Chaves, M. (2010). ”Linkable Geographic Ontologies.” In *Proceedings of the 6th Workshop on Geographic Information Retrieval*, Zurich, Switzerland, pp. 18-19.

Ferreira, J. D., Batista, D., Couto, F., and Silva, M. J. (2010). ”The Geo-Net-PT/Yahoo! GeoPlanet (TM) concordance”. In *Technical Report. TR 10-05*. University of Lisbon, Faculty of Sciences, LASIGE.

Chaves, M., Silva, M. J., and Martins B. (2005). ”GKB – Geographic Knowledge Base.” In *Technical report*, Departamento de Informática da Faculdade de Ciências da Universidade de Lisboa. Available at: http://www.di.fc.ul.pt/tech-reports/05-12.pdf. DI/FCUL TR 05-12.

Chaves, M., Rodrigues, C., and Silva, M. J. (2007). ”Data Model for Geographic Ontologies Generation.” In *XATA 2007 - XML: Aplicações e Tecnologias Associadas*, pp. 47-58.

Chaves, M. (2009). *Uma Metodologia para Construção de Geo-Ontologias PhD Thesis*. Faculty of Sciences, University of Lisbon.