Representing Arabic Lexicons in Lemon
– Preliminary Study

Mustafa Jarrar*, Hamzeh Amayreh*, John P. McCrae+

*Birzeit University, Palestine
+National University of Ireland Galway, Ireland

Abstract: Represent 150 Arabic-multilingual lexicons using Lemon - to enable them to be used by NLP applications, and interlinked with the Open Linguistic Data Cloud.

Types of Lexicon we addressed

- **Dictionary**: a list of lexical entries, each with some bi/trilingual translations.
- **Thesaurus**: sets of synonymous lexical entries. Each set is lexicalized in one or more languages.
- **Glossary**: a domain-specific lexicon. Each lexical entry is defined in a few lines. Advanced glossaries provide also synonyms, translation(s), and relations.
- **Linguistic Lexicon**: entries with linguistic features, and senses that might be combined in a description.
- **Semantic-variations lexicon**: pairs of semantically close lexical entries and the differences between their meanings, (e.g. like ~ love, pain ~ ache).

Tentative Representation in Lemon

- **Lexical entry**: a translation term in a dictionary, a synonym in a thesaurus, a term in a glossary, or a headword in a linguistic lexicon.
- **Lexical concept**: a gloss in a glossary, a set of synonyms in a thesaurus, or a translations set in a dictionary.
- **Ontology concepts**: entities in the Arabic Ontology, also linked with lexical concepts using the isConceptOf property.
- **Relations**: semantic relations like *related*, *border/narrower*, etc) represented using conceptRel.
- **Linguistic features**: Glosses/definitions are skos:definition. POS, root and inflictions are using other Lemon properties.

Example

country بلد | دولة

موجود اعتباري يُعرَف بحُدُوده السياسية المُتفق عليها له شَعْب ويُشكّل منظومة مستقل ذات حُكُومَة ومُؤسَّسَات مُنظَّمَة.

BZU Thesaurus ©

```
...
@prefix aot: <http://ontology.birzeit.edu/term/>.
@prefix  ao: <http://ontology.birzeit.edu/concept/>.
@prefix aoc: <http://ontology.birzeit.edu/lexicalconcept/>.
@prefix aor: <http://ontology.birzeit.edu/lexicon/>.
<aoc:1623> a ontolex:LexicalConcept;
 ontolex:isEvokedBy <aot:Lex-country>;
 ontolex:isEvokedBy <aot:Lex-دولة>;
 ontolex:isEvokedBy <aot:Lex-بلد>;
 skos:definition "موجود اعتباري يُعرَف بحُدُوده السياسية المُتفق عليها له شَعْب ويُشكّل ..."@ar;
 skos:inScheme <aor:BZU_Thesaurus_43>;
 ontolex:Concept <ao:293121>.

<aot:lex-country> a ontolex:LexicalEntry, ontolex:Word;
 ontolex:canonicalForm [ontolex:writtenRep "country"@en];
 skos:inScheme <aor:BZU_Thesaurus_43>.
<aot:lex-دولة> a ontolex:LexicalEntry, ontolex:Word;
 ontolex:canonicalForm [ontolex:writtenRep "دولة"@ar];
 skos:inScheme <aor:BZU_Thesaurus_43>.
<aot:lex-بلد> a ontolex:LexicalEntry, ontolex:Word;
 ontolex:canonicalForm [ontolex:writtenRep "بـلـد"@ar];
 skos:inScheme <aor:BZU_Thesaurus_43>.
```

Major Challenges

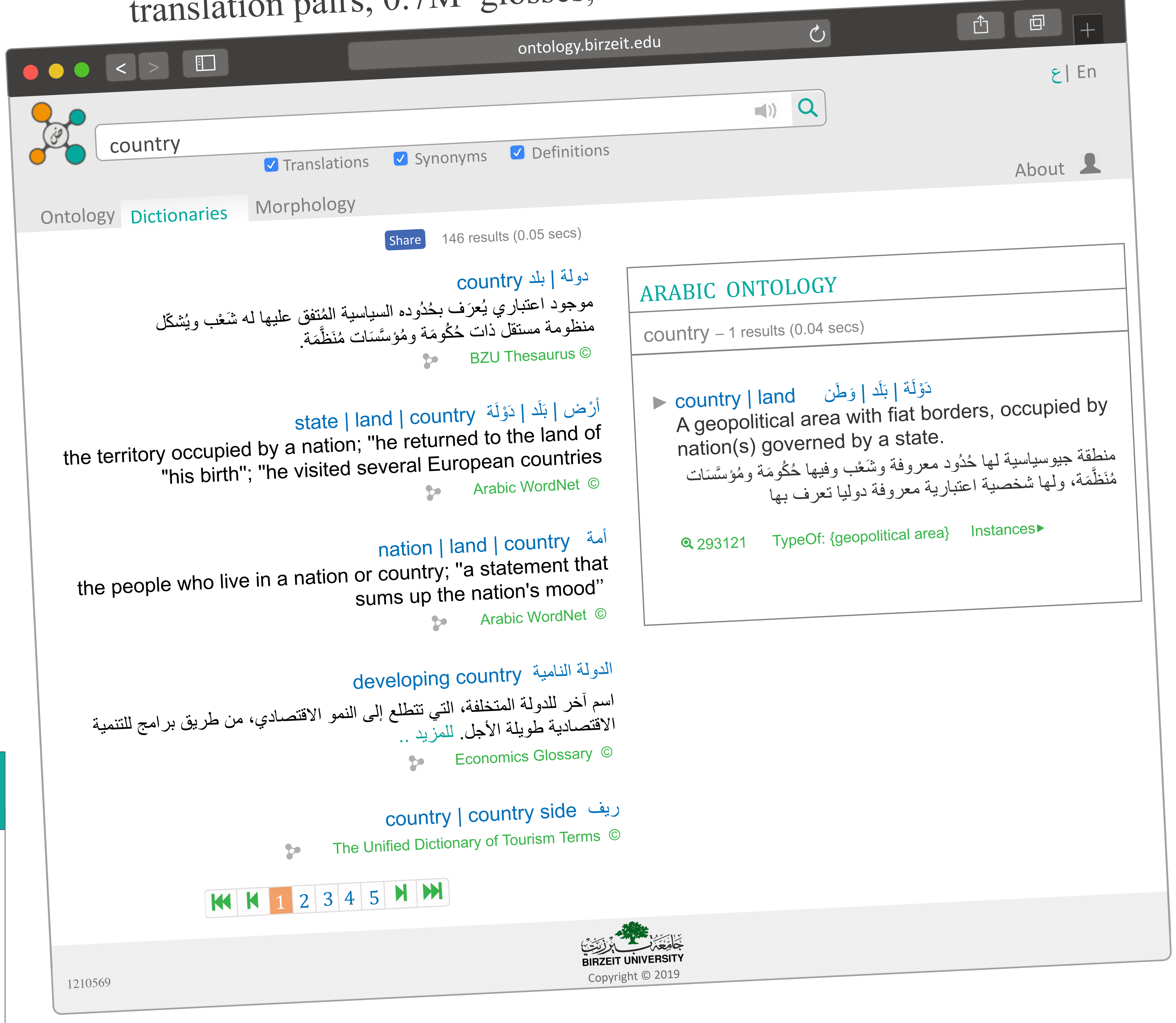
Arabic lexical entries are less often lemmas

- Many Arabic lexicons do not strictly follow lemmatization conventions.
- lexical entries in Arabic lexicons might be partially or not at all diacritized (difficult to be disambiguated).

Lexicographic Search Engine

http://ontology.birzeit.edu

The largest Arabic-Multilingual lexicographic database: **150 lexicons + Arabic Ontology**, 1.1M lexical concepts, 2.4M lexical entries, 1.5M translation pairs, 0.7M glosses, and 0.5M semantic relations.



ontology.birzeit.edu

ع | En

country

☑ Translations ☑ Synonyms ☑ Definitions

About

Ontology Dictionaries Morphology

Share 146 results (0.05 secs)

country بلد | دولة
موجود اعتباري يُعرَف بحُدُوده السياسية المُتفق عليها له شَعْب ويُشكّل منظومة مستقل ذات حُكُومَة ومُؤسَّسَات مُنظَّمَة.
BZU Thesaurus ©

state | land | country أرْض | بَلَد | دَوْلَة
the territory occupied by a nation; "he returned to the land of "his birth"; "he visited several European countries
Arabic WordNet ©

nation | land | country أمة
the people who live in a nation or country; "a statement that sums up the nation's mood"
Arabic WordNet ©

developing country الدولة النامية
اسم آخر للدولة المتخلفة، التي تتطلع إلى النمو الاقتصادي، من طريق برامج للتنمية الاقتصادية طويلة الأجل. للمزيد ..
Economics Glossary ©

country | country side ريف
The Unified Dictionary of Tourism Terms ©

1 2 3 4 5

ARABIC ONTOLOGY

country – 1 results (0.04 secs)

▶ country | land دَوْلَة | بَلَد | وَطَن
A geopolitical area with fiat borders, occupied by nation(s) governed by a state.
منطقة جيوسياسية لها حُدُود معروفة وشَعْب وفيها حُكُومَة ومُؤسَّسَات مُنَظَّمَة، ولها شخصية اعتبارية معروفة دوليا تعرف بها

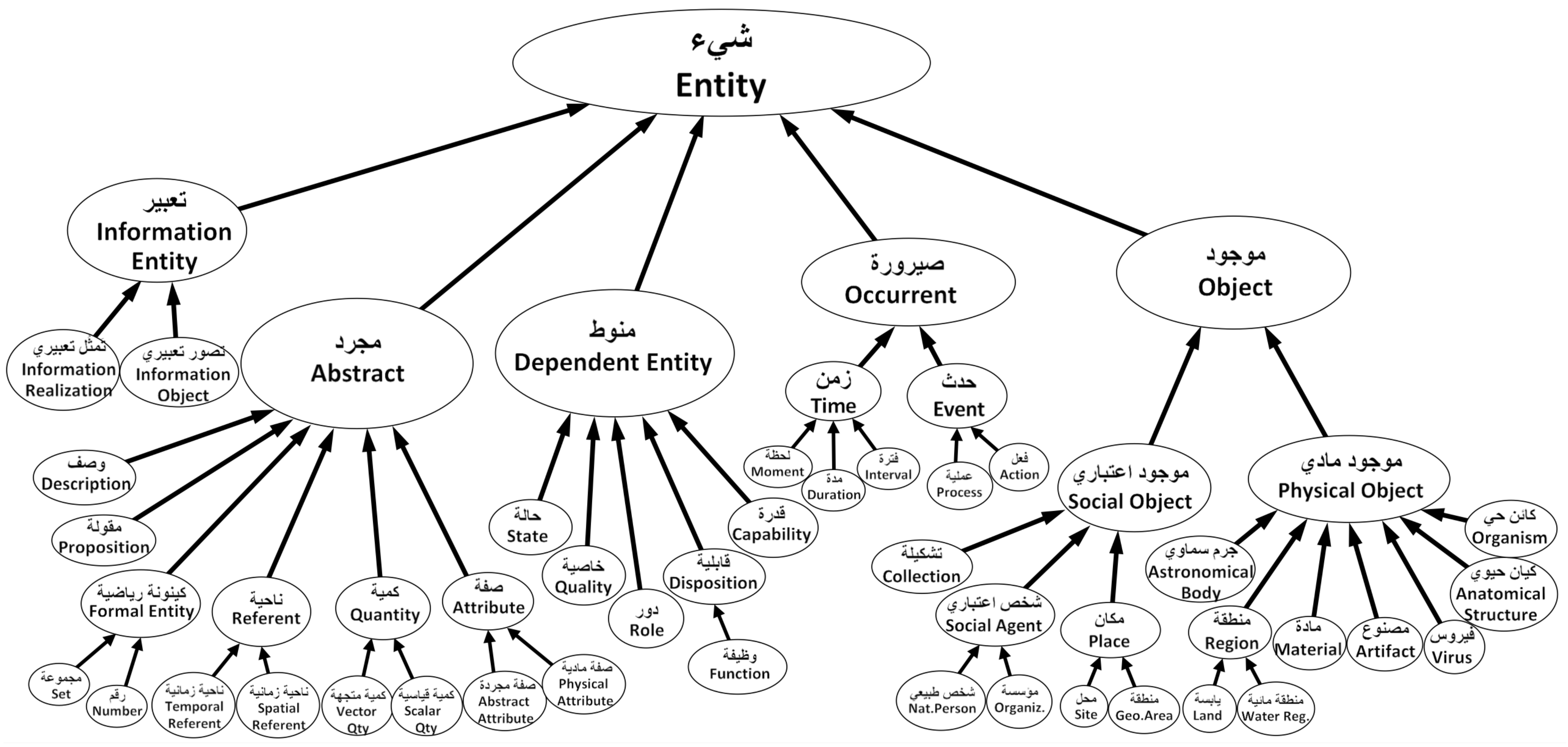
293121 TypeOf: {geopolitical area} Instances▸

1210569

BIRZEIT UNIVERSITY
Copyright © 2019

Arabic Ontology

An Arabic Wordnet with Ontologically Clean Content (~1300 Concepts)



To correctly represent Arabic entries in Lemon

Each lexical entry needs to be carefully lemmatized first

The lemma for each lexical entry, in each of the 150 lexicons should be specified, which would enable lexicons to be interlinked based on their lemmas.

Also: extend the Lemon morph module to cover Arabic-specific features, e.g., imperfect and imperative verbs, verbal nouns, intensive participle, place nouns, time nouns, instrumental noun.