NLDB 2019, Salford, UK
26, June 2019

An Arabic-Multilingual Database with **a Lexicographic Search Engine**

Mustafa Jarrar
Birzeit University
Palestine

An Arabic-Multilingual Database with a Lexicographic Search Engine

Mustafa Jarrar(✉) and Hamzeh Amayreh

Birzeit University, Birzeit, Palestine
mjarrar@birzeit.edu, hamayreh@staff.birzeit.edu

Abstract. We present a lexicographic search engine built on top of the largest Arabic multilingual database, allowing people to search and retrieve translations, synonyms, definitions, and more. The database currently contains about 150 Arabic multilingual lexicons that we have been digitizing, restructuring, and normalizing over 9 years. It comprises most types of lexical resources, such as modern and classical lexicons, thesauri, glossaries, lexicographic datasets, and (bi/)tri-lingual dictionaries. This is in addition to the Arabic Ontology – an Arabic WordNet with ontologically cleaned content, which is being used to reference and interlink lexical concepts. The search engine was developed with the state-of-the-art design features and according to the W3C's recommendation and best practices for publishing data on the web, as well as the W3C's Lemon RDF model. The search engine is publicly available at (https://ontology.birzeit.edu).

Keywords: Arabic · Multilingual lexicons · Online dictionary · Language resources · Lexical semantics · Lexicographic search · W3C lemon · RDF · NLP

1 Introduction and Motivation

The increasing demands to use and reuse dictionaries (of all types) in modern appli-

Mustafa Jarrar, Hamzeh Amayreh: **An Arabic-Multilingual Database with a Lexicographic Search Engine.** The 24th International Conference on Applications of Natural Language to Information Systems (NLDB 2019). Pages(234--246), LNCS 11608, Springer. 2019.

http://www.jarrar.info/publications/JA19.pdf

Why Lexical Resources!

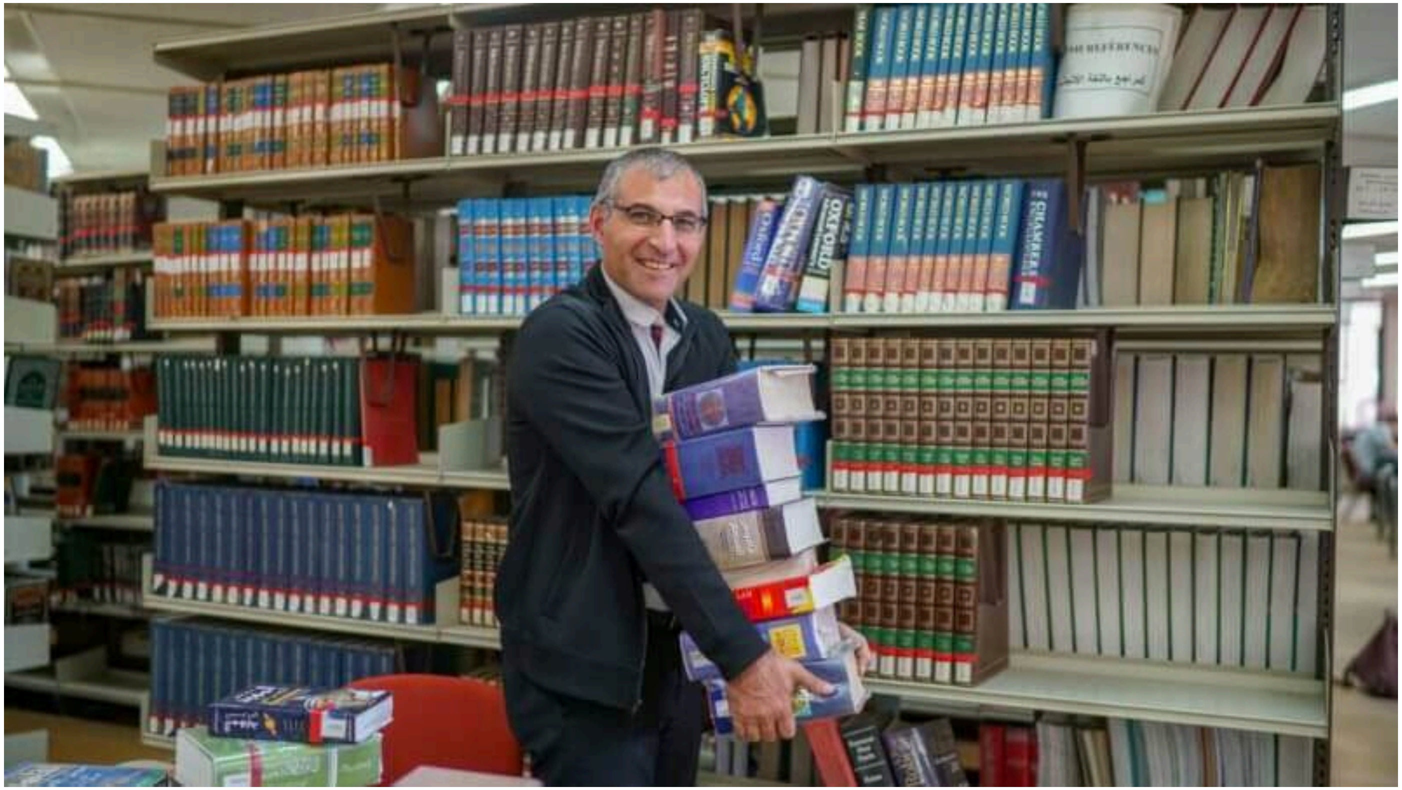
- The importance of lexical resources (dictionaries, thesauri, wordnets, linguistic ontologies) is increasing in many application areas, such as:
    - NLP tasks and applications
    - Information search and retrieval
    - Multilingual big data
    - Multilingual semantic web
    - Data integration
    - among many others.

- Lack of Arabic Lexical resources for human use!

- Lack of Arabic Lexical resources for NLP!

Solution

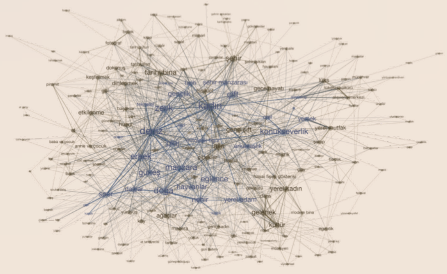
Digitize, Collect, Build, then clean and link



- Make available online for people
- Make available through APIs for NLP applications

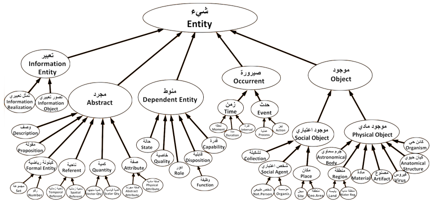
Lexical Resources at Birzeit University

Lexicographic Database



The largest Arabic-multilingual database (semantics & morphology)

Arabic Ontology



Classification of the meanings of the Arabic Terms - formal Arabic Wordnet

Dialect Corpus

كراس
Curras
Corpus for Palestinian Arabic
مدونة اللهجة العامية الفلسطينية

Search | Search

☑ Stem ☑ MSA ☑ Surface ☑ Gloss
● Whole Word ○ Substring

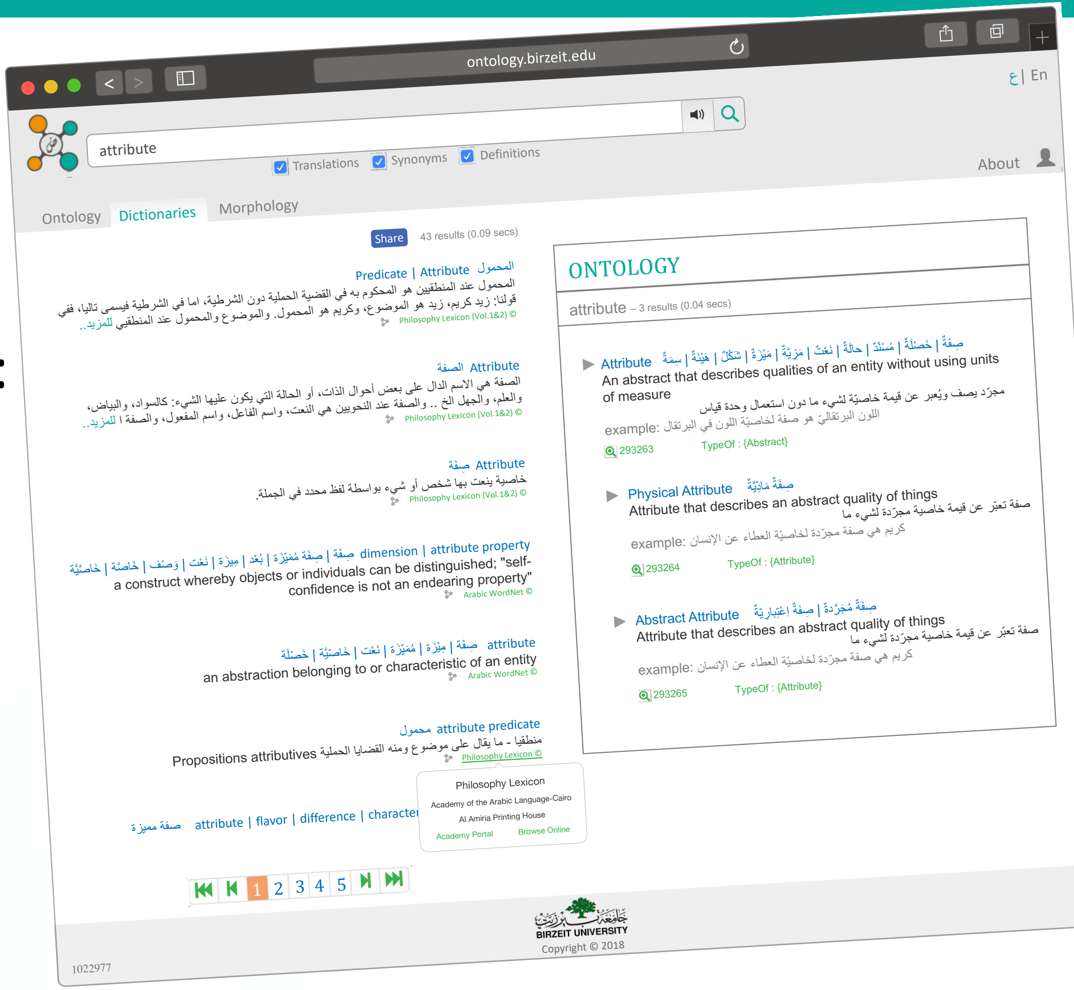
A large text in dialect, each word is annotated with 16 features

Linguistic Big Data Graph via a **Lexicographic Search Engine**

The Lexicographic Database

The Lexicographic Database

- **The largest lexicographic Arabic database**
- **Contains most lexicon types**: glossaries, thesauri, bi/trilingual dictionaries, morph datasets, **Arabic Ontology**, and more.
- **Covers most domains:** science, technology, law, business, art, philosophy, ...



43 results (0.09 secs)

Predicate | Attribute المحمول
المحمول عند المنطقيين هو المحكوم به في القضية الحملية دون الشرطية، اما في الشرطية فيسمى تاليا، ففي قولنا: زيد كريم، زيد هو الموضوع، وكريم هو المحمول. والموضوع والمحمول عند المنطقيي للمزيد..
Philosophy Lexicon (Vol.1&2) ©

Attribute الصفة
الصفة هي الاسم الدال على بعض أحوال الذات، أو الحالة التي يكون عليها الشيء: كالسواد، والبياض، والعلم، والجهل الخ .. والصفة عند النحويين هي النعت، واسم الفاعل، واسم المفعول، والصفة ا للمزيد..
Philosophy Lexicon (Vol.1&2) ©

Attribute صِفَة
خاصية ينعت بها شخص أو شيء بواسطة لفظ محدد في الجملة.
Philosophy Lexicon (Vol.1&2) ©

attribute property | dimension صِفَة | صِفَة مُمَيِّزَة | بُعْد | مِيزَة | نَعْت | وَصْف | خَاصَّة | خَاصِّيَّة
a construct whereby objects or individuals can be distinguished; "self-confidence is not an endearing property"
Arabic WordNet ©

attribute صِفَة | مِيْزَة | مُمَيِّزَة | نَعْت | خَاصِّيَّة | خَصْلَة
an abstraction belonging to or characteristic of an entity
Arabic WordNet ©

attribute predicate محمول
منطقيا - ما يقال على موضوع ومنه القضايا الحملية Propositions attributives
Philosophy Lexicon ©

Philosophy Lexicon
Academy of the Arabic Language-Cairo
Al Amiria Printing House
Academy Portal Browse Online

character | difference | flavor | attribute صفة مميزة

⏮ ⏴ 1 2 3 4 5 ⏵ ⏭

ONTOLOGY

attribute – 3 results (0.04 secs)

▶ Attribute صِفَةٌ | خَصْلَةٌ | مُسْنَدٌ | حالَةٌ | نَعْتٌ | مَزِيَّةٌ | مَيْزَةٌ | شَكْلٌ | هَيْئَةٌ | سِمَةٌ
An abstract that describes qualities of an entity without using units of measure
مجرّد يصف ويُعبر عن قيمة خاصيّة لشيء ما دون استعمال وحدة قياس
example: اللون البرتقاليّ هو صفة لخاصيّة اللون في البرتقال
293263 TypeOf : {Abstract}

▶ Physical Attribute صِفَةٌ مَادِيَّةٌ
Attribute that describes an abstract quality of things
صفة تعبّر عن قيمة خاصية مجرّدة لشيء ما
example: كريم هي صفة مجرّدة لخاصيّة العطاء عن الإنسان
293264 TypeOf : {Attribute}

▶ Abstract Attribute صِفَةٌ مُجَرَّدَةٌ | صِفَةٌ إِعْتِبارِيّةٌ
Attribute that describes an abstract quality of things
صفة تعبّر عن قيمة خاصية مجرّدة لشيء ما
example: كريم هي صفة مجرّدة لخاصيّة العطاء عن الإنسان
293265 TypeOf : {Attribute}

BIRZEIT UNIVERSITY
Copyright © 2018

1022977

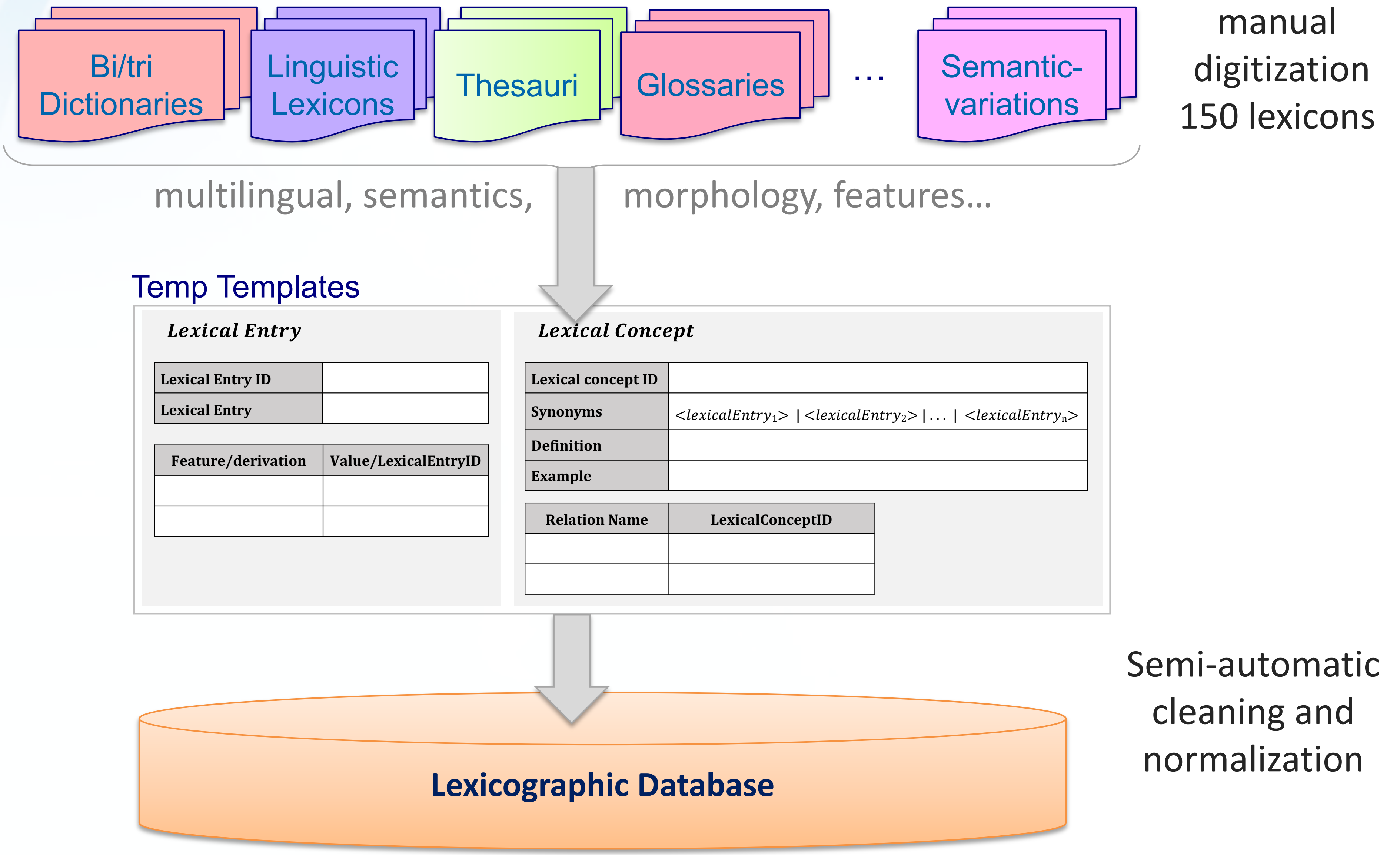
https://ontology.birzeit.edu

Some Statistics

Currently!

Category	Lexical Concepts	Lexical entries	Synsets	Translations pairs	Glosses	Semantic relations
Total (Millions)	1.1 M	2.4 M	1.8 M	1.5 M	0.7 M	0.5 M
Sub Counts		1,100K Arabic 1,100K English 200K French 3K Others 1,300K Single-word 1,000K Multi-word	800K Arabic 800K English 200K French 50K Others	1,000K English-Arabic 300K English-French 200K French-Arabic	400K Arabic 300K English 1K Others	170K Sub-super links 29K Part-of links 260K Has-Domain links 30K Other links

Constructing the Database (9 years)



manual digitization
150 lexicons

multilingual, semantics, morphology, features…

Temp Templates

Lexical Entry

Lexical Entry ID	
Lexical Entry	

Feature/derivation	**Value/LexicalEntryID**

Lexical Concept

Lexical concept ID	
Synonyms	$\langle lexicalEntry_1 \rangle \mid \langle lexicalEntry_2 \rangle \mid \ldots \mid \langle lexicalEntry_n \rangle$
Definition	
Example	

Relation Name	**LexicalConceptID**

Lexicographic Database

Semi-automatic cleaning and normalization

Cleaning and Normalization

- Lexicons are typically designed to be printed and used as hard copies.
- Big challenges when converting them into a machine processable format.

Examples of challenges:

Challenges induced by ordering:

“accelerator (linear...)”, “affinity (chemical)”, “drawing (final)”, “earth (the)”, and “crush (to)”, “tube (pipe)”, “academy (of art)”

Subterm synonymy:

“liquid drier, drier”, “calomel electrode, calomel”, “kelvin's scale, kelvin's absolute scale”

Long multiword lexical entries:

“buildings or other structures recurrent taxes on land”. Such cases of long and “poor”

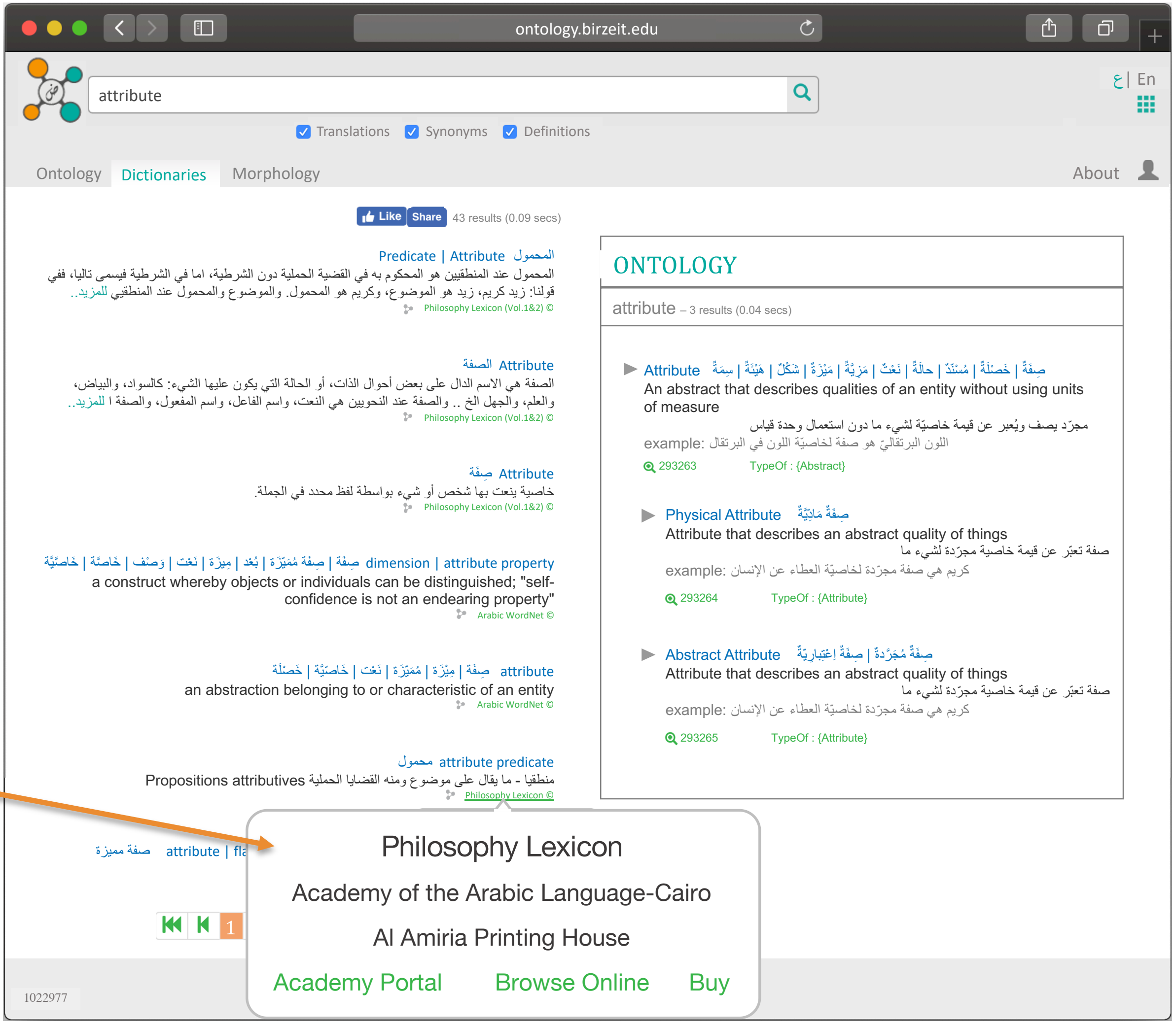
Character set: Same characters and symbols have different encodings across different languages (e.g., the dash, quotations, punctuations, and whitespaces),

➢ **See 30 parsers at:**

Hamzeh Amayreh, Mohammad Dwaikat, and Mustafa Jarrar. **Lexicon Digitization -A Framework for Structuring, Normalizing and Cleaning Lexical Entries**, Technical Report, Birzeit University 2018.

Obtaining Copyrights

- Obtained permission from each lexicons owner (individually contacted).
- Most accepted!
- Show lexicon name and © copyright symbol beside each result.
- Promote lexicons (click to see lexicon info)



attribute

☑ Translations ☑ Synonyms ☑ Definitions

43 results (0.09 secs)

Predicate | Attribute المحمول

المحمول عند المنطقيين هو المحكوم به في القضية الحملية دون الشرطية، اما في الشرطية فيسمى تاليا، ففي
قولنا: زيد كريم، زيد هو الموضوع، وكريم هو المحمول. والموضوع والمحمول عند المنطقيي للمزيد..

Philosophy Lexicon (Vol.1&2) ©

Attribute الصفة

الصفة هي الاسم الدال على بعض أحوال الذات، أو الحالة التي يكون عليها الشيء: كالسواد، والبياض،
والعلم، والجهل الخ .. والصفة عند النحويين هي النعت، واسم الفاعل، واسم المفعول، والصفة ا للمزيد..

Philosophy Lexicon (Vol.1&2) ©

Attribute صِفَة

خاصية ينعت بها شخص أو شيء بواسطة لفظ محدد في الجملة.

Philosophy Lexicon (Vol.1&2) ©

attribute property | dimension صِفَة | صِفَة مُمَيِّزَة | بُعْد | مِيزَة | نَعْت | وَصْف | خَاصَّة | خَاصَّيَّة

a construct whereby objects or individuals can be distinguished; "self-confidence is not an endearing property"

Arabic WordNet ©

attribute صِفَة | مِيْزَة | مُمَيِّزَة | نَعْت | خَاصِّيَّة | خَصْلَة

an abstraction belonging to or characteristic of an entity

Arabic WordNet ©

attribute predicate محمول

منطقيا - ما يقال على موضوع ومنه القضايا الحملية Propositions attributives

Philosophy Lexicon ©

attribute | fla صفة مميزة

1

1022977

Philosophy Lexicon

Academy of the Arabic Language-Cairo

Al Amiria Printing House

Academy Portal Browse Online Buy

ONTOLOGY

attribute – 3 results (0.04 secs)

▶ Attribute صِفَةٌ | خَصْلَةٌ | مُسْنَدٌ | حالَةٌ | نَعْتٌ | مَزِيَّةٌ | مَيْزَةٌ | شَكْلٌ | هَيْئَةٌ | سِمَةٌ

An abstract that describes qualities of an entity without using units of measure

مجرّد يصف ويُعبر عن قيمة خاصيّة لشيء ما دون استعمال وحدة قياس

example: اللون البرتقاليّ هو صفة لخاصيّة اللون في البرتقال

293263 TypeOf : {Abstract}

▶ Physical Attribute صِفَةٌ مَادِّيَّةٌ

Attribute that describes an abstract quality of things

صفة تعبّر عن قيمة خاصية مجرّدة لشيء ما

example: كريم هي صفة مجرّدة لخاصيّة العطاء عن الإنسان

293264 TypeOf : {Attribute}

▶ Abstract Attribute صِفَةٌ مُجَرَّدَةٌ | صِفَةٌ اِعْتِبارِيَّةٌ

Attribute that describes an abstract quality of things

صفة تعبّر عن قيمة خاصية مجرّدة لشيء ما

example: كريم هي صفة مجرّدة لخاصيّة العطاء عن الإنسان

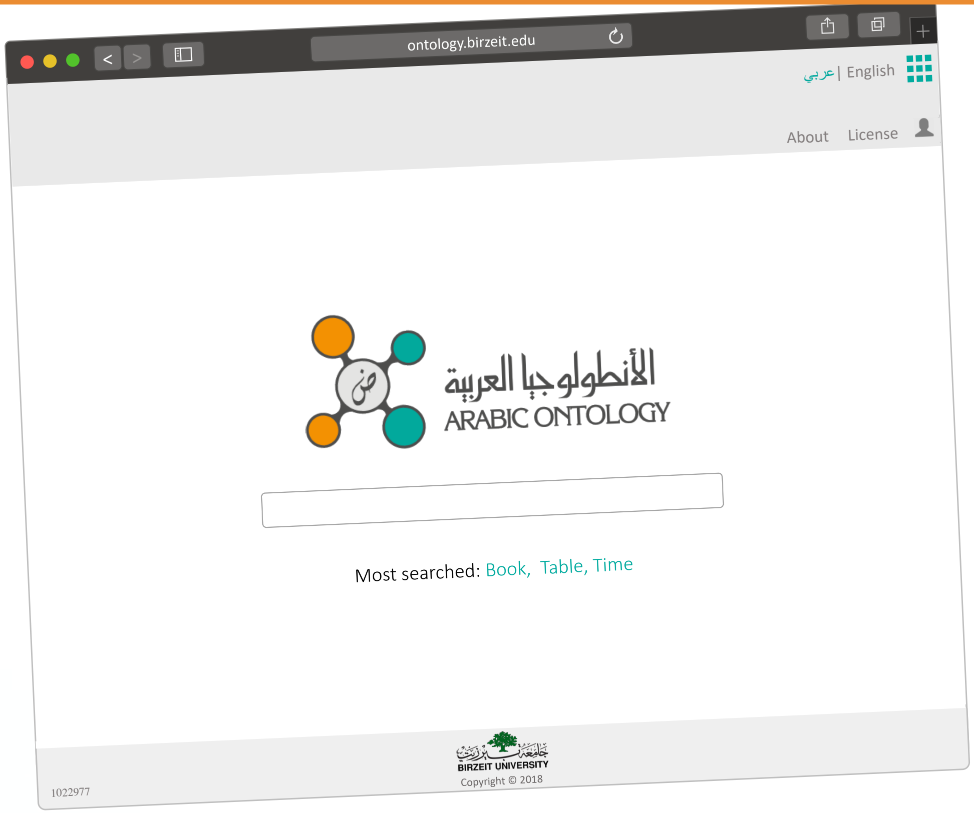
293265 TypeOf : {Attribute}

Lexicographic Search Engine

Lexicographic Search Engine

- **Free access to people**: students, translators, researchers, Arabic learners …

- **API accessible** for NLP applications.



ontology.birzeit.edu

عربي | English

About License

الأنطولوجيا العربية

ARABIC ONTOLOGY

Most searched: Book, Table, Time

BIRZEIT UNIVERSITY

Copyright © 2018

1022977

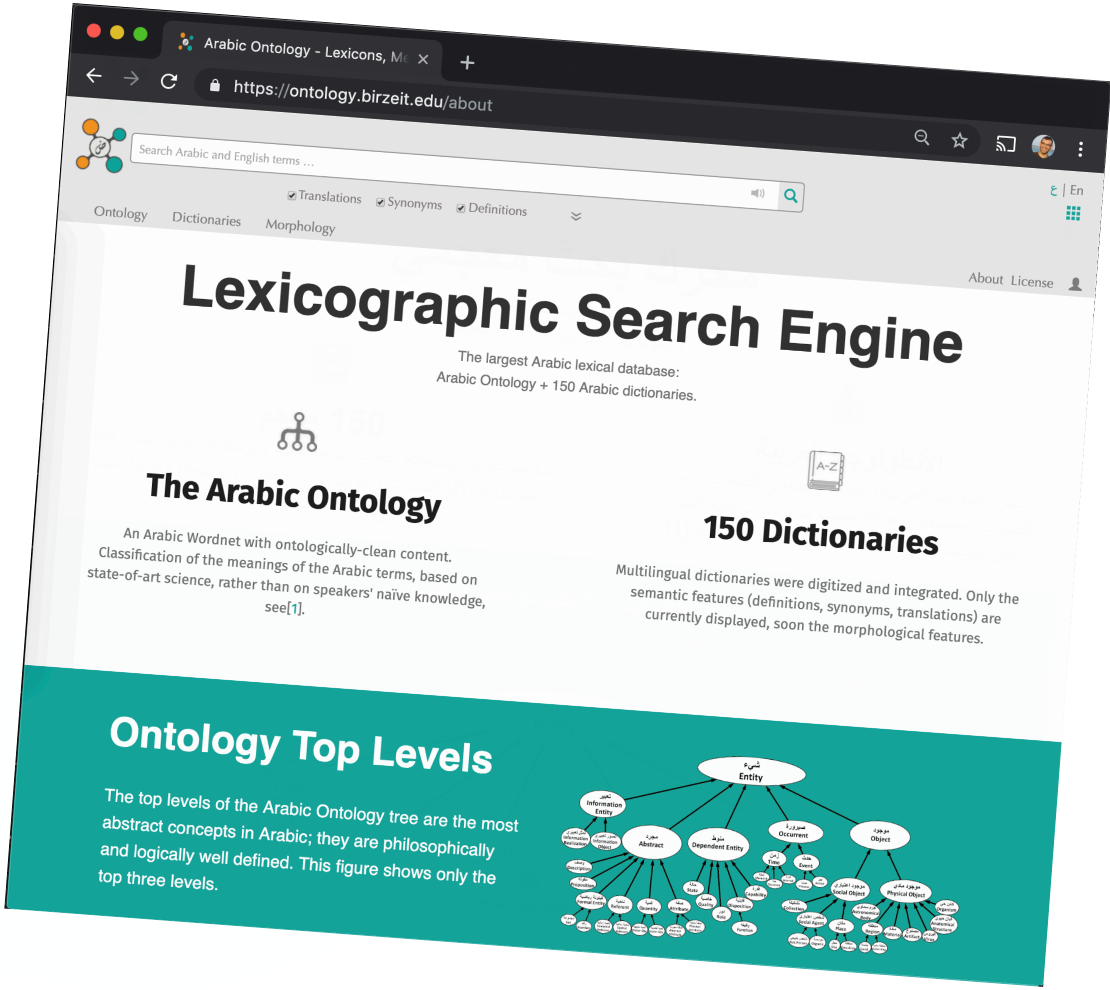
https://ontology.birzeit.edu

Reference:

Mustafa Jarrar, Hamzeh Amayreh: **An Arabic-Multilingual Database with a Lexicographic Search Engine.** NLDB 2019. Pages(234--246), LNCS 11608, Springer. 2019.

Lexicographic Search Engine

- **Search 150 lexicons** for definitions, synonyms, specialized translations, morphology, ontology...
- **Accurate**! compared with machine translation.
- **The first of its kind**! e.g., there are no similar search engines for English lexicons!



Arabic Ontology - Lexicons, Me
https://ontology.birzeit.edu/about
Search Arabic and English terms ...
☑ Translations
☑ Synonyms
☑ Definitions
Ontology
Dictionaries
Morphology
ع | En
About
License

Lexicographic Search Engine

The largest Arabic lexical database:
Arabic Ontology + 150 Arabic dictionaries.

The Arabic Ontology

An Arabic Wordnet with ontologically-clean content. Classification of the meanings of the Arabic terms, based on state-of-art science, rather than on speakers' naïve knowledge, see[1].

150 Dictionaries

Multilingual dictionaries were digitized and integrated. Only the semantic features (definitions, synonyms, translations) are currently displayed, soon the morphological features.

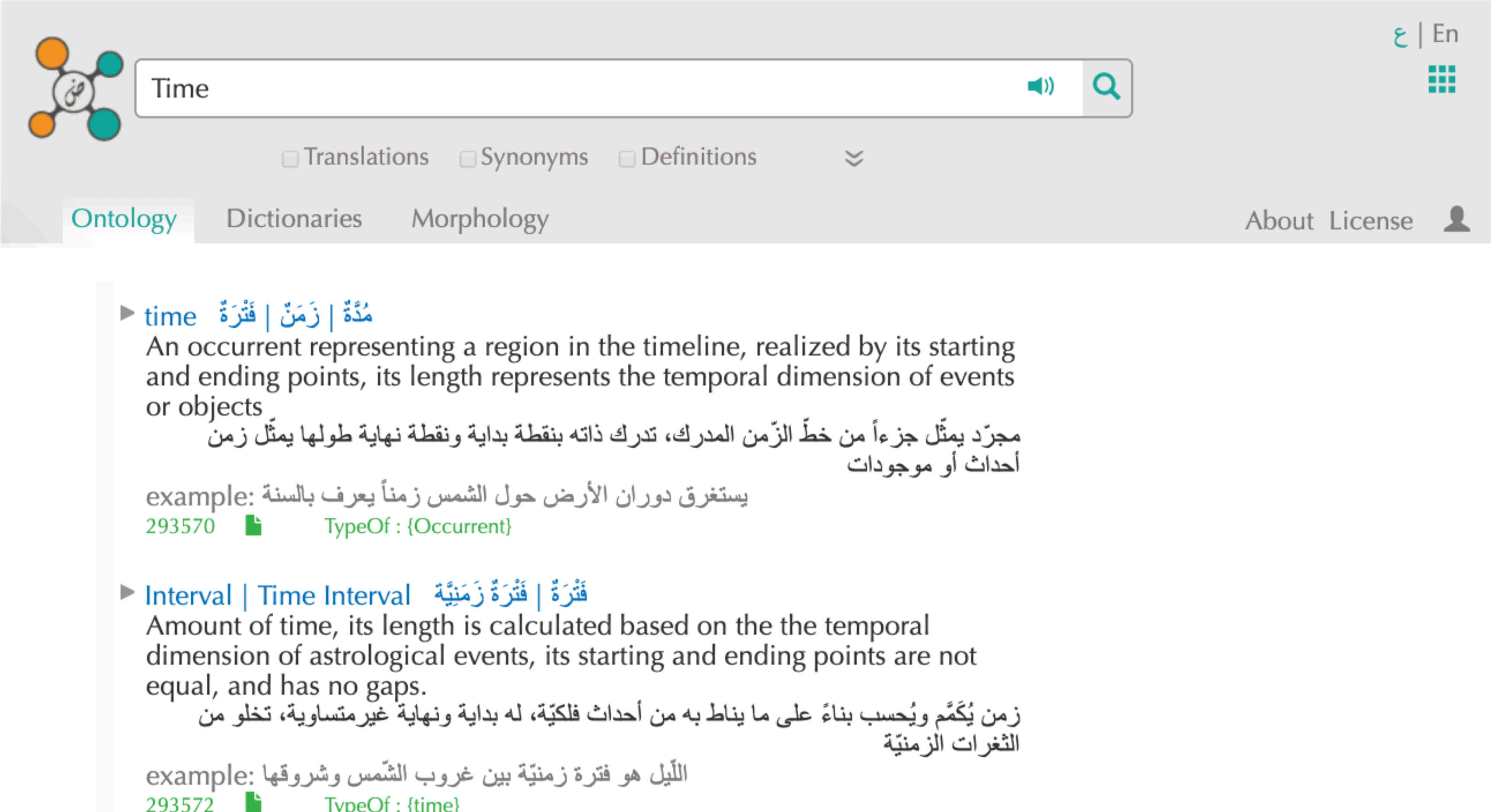
Ontology Top Levels

The top levels of the Arabic Ontology tree are the most abstract concepts in Arabic; they are philosophically and logically well defined. This figure shows only the top three levels.

https://ontology.birzeit.edu

Search Taps

Ontology tab: results in this tab are ontology concepts retrieved only from the Arabic ontology. The tab also allows expanding and exploring the ontology tree.



Time

▸ time مُدَّةٌ | زَمَنٌ | فَتْرَةٌ

An occurrent representing a region in the timeline, realized by its starting and ending points, its length represents the temporal dimension of events or objects

مجرّد يمثّل جزءاً من خطّ الزّمن المدرك، تدرك ذاته بنقطة بداية ونقطة نهاية طولها يمثّل زمن أحداث أو موجودات

example: يستغرق دوران الأرض حول الشمس زمناً يعرف بالسنة

293570 TypeOf : {Occurrent}

▸ Interval | Time Interval فَتْرَةٌ | فَتْرَةٌ زَمَنِيَّة

Amount of time, its length is calculated based on the the temporal dimension of astrological events, its starting and ending points are not equal, and has no gaps.

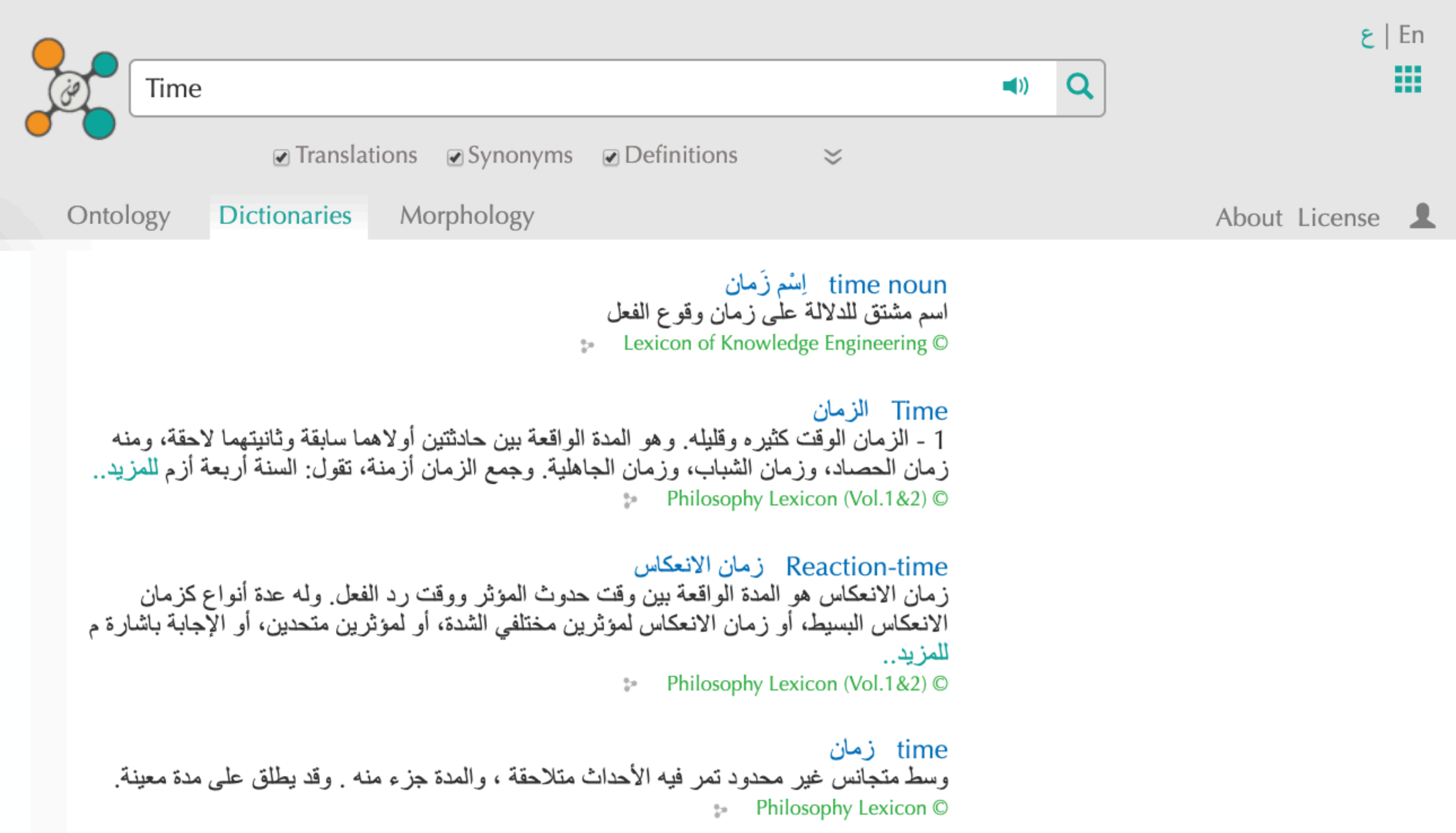
زمن يُكَمَّم ويُحسب بناءً على ما يناط به من أحداث فلكيّة، له بداية ونهاية غيرمتساوية، تخلو من الثغرات الزمنيّة

example: اللّيل هو فترة زمنيّة بين غروب الشّمس وشروقها

293572 TypeOf : {time}

Search Taps

Dictionaries tab: results in this tab are lexical concepts retrieved from the lexicons.



Time

☑ Translations ☑ Synonyms ☑ Definitions

time noun إِسْم زَمان
اسم مشتق للدلالة على زمان وقوع الفعل
Lexicon of Knowledge Engineering ©

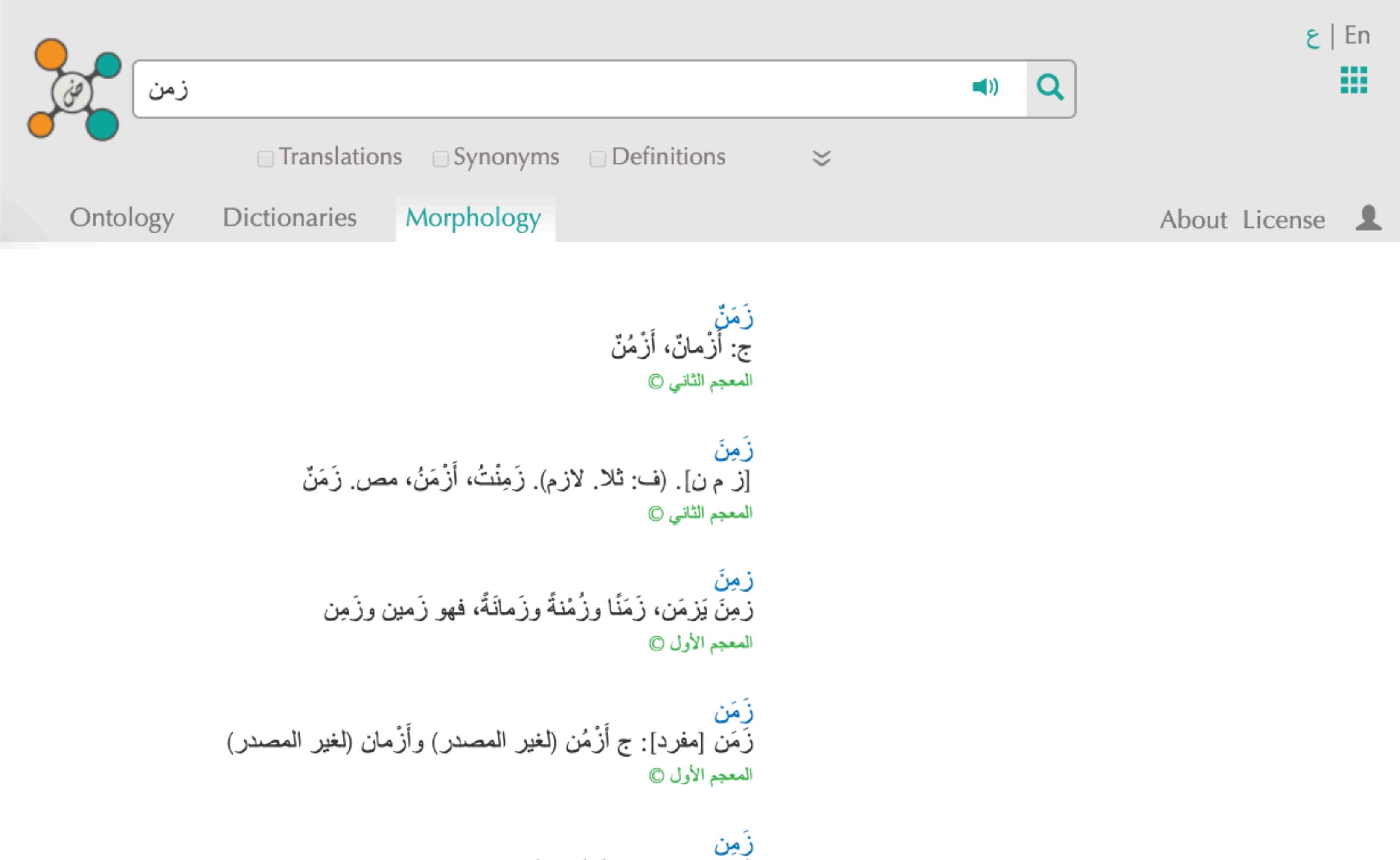
Time الزمان
1 - الزمان الوقت كثيره وقليله. وهو المدة الواقعة بين حادثتين أولاهما سابقة وثانيتهما لاحقة، ومنه
زمان الحصاد، وزمان الشباب، وزمان الجاهلية. وجمع الزمان أزمنة، تقول: السنة أربعة أزم للمزيد..
Philosophy Lexicon (Vol.1&2) ©

Reaction-time زمان الانعكاس
زمان الانعكاس هو المدة الواقعة بين وقت حدوث المؤثر ووقت رد الفعل. وله عدة أنواع كزمان
الانعكاس البسيط، أو زمان الانعكاس لمؤثرين مختلفي الشدة، أو لمؤثرين متحدين، أو الإجابة باشارة م
للمزيد..
Philosophy Lexicon (Vol.1&2) ©

time زمان
وسط متجانس غير محدود تمر فيه الأحداث متلاحقة ، والمدة جزء منه . وقد يطلق على مدة معينة.
Philosophy Lexicon ©

Search Taps

Morphology tab: results are linguistic features, lemma(s), inflections, and derivations of the searched term (partially implemented!).



زَمَنٌ
ج: أَزْمانٌ، أَزْمُنٌ
المعجم الثاني ©

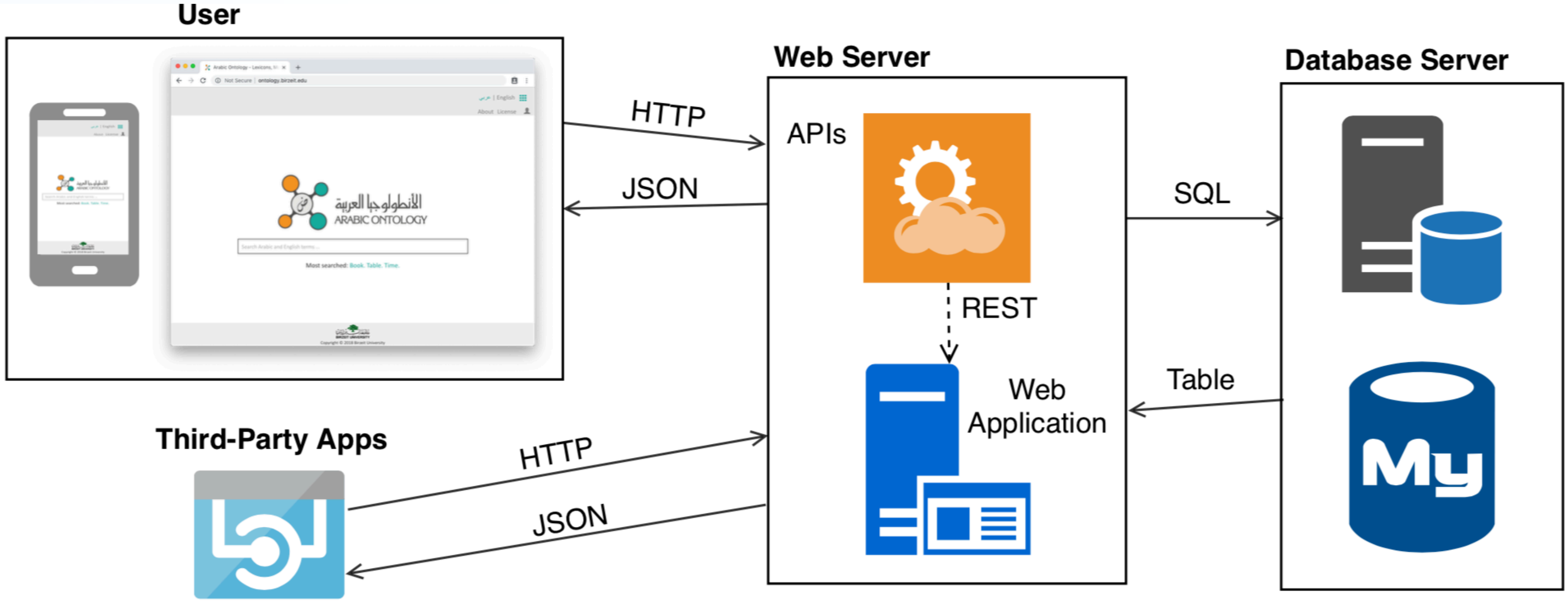
زَمِنَ
[ز م ن]. (فـ: ثلا. لازم). زَمِنْتُ، أَزْمَنُ، مص. زَمَنٌ
المعجم الثاني ©

زمِنَ
زمِنَ يَزمَن، زَمَنًا وزُمْنةً وزَمانَةً، فهو زَمين وزَمِن
المعجم الأول ©

زَمَن
زَمَن [مفرد]: ج أَزْمُن (لغير المصدر) وأَزْمان (لغير المصدر)
المعجم الأول ©

زَمِن

Search Engine Architecture

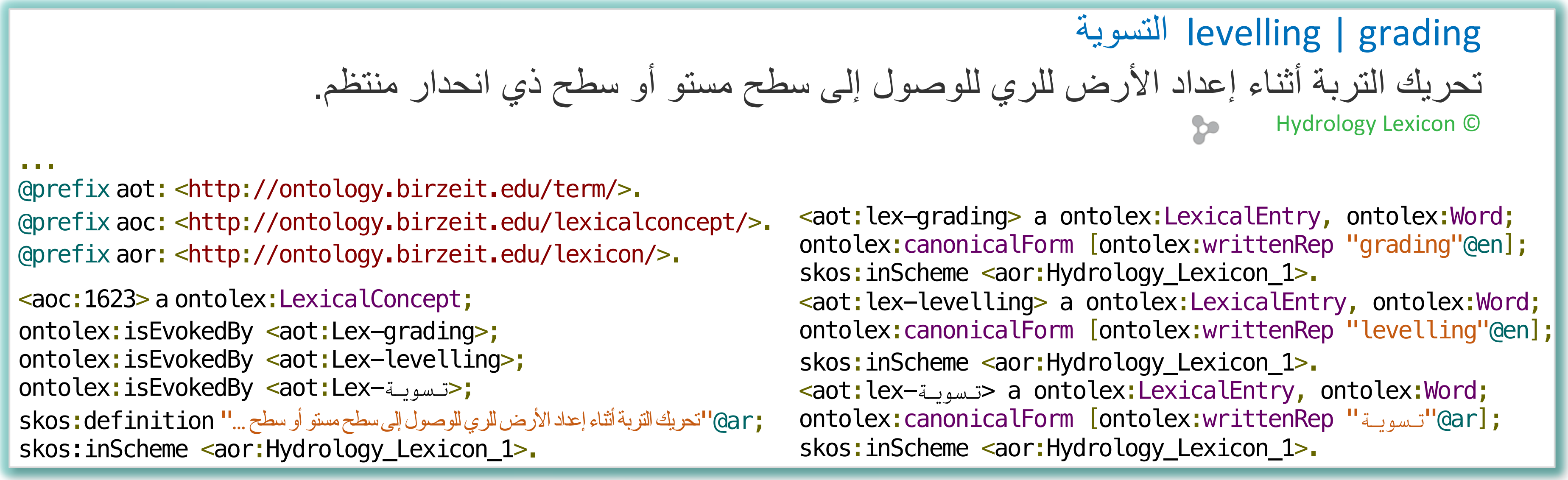


Conformance with W3C Standards

✓ W3C’s RDF Lemon Model

Represent (lexical entries, concepts, synsets, …) using the Lemon RDF model

To interlink it with the Linguistic Linked Open Data Cloud



التسوية levelling | grading

تحريك التربة أثناء إعداد الأرض للري للوصول إلى سطح مستو أو سطح ذي انحدار منتظم.

Hydrology Lexicon ©

```
...
@prefix aot: <http://ontology.birzeit.edu/term/>.
@prefix aoc: <http://ontology.birzeit.edu/lexicalconcept/>.
@prefix aor: <http://ontology.birzeit.edu/lexicon/>.

<aoc:1623> a ontolex:LexicalConcept;
ontolex:isEvokedBy <aot:Lex-grading>;
ontolex:isEvokedBy <aot:Lex-levelling>;
ontolex:isEvokedBy <aot:Lex-تـسـويـة>;
skos:definition "تحريك التربة أثناء إعداد الأرض للري للوصول إلى سطح مستو أو سطح..."@ar;
skos:inScheme <aor:Hydrology_Lexicon_1>.
```

```
<aot:lex-grading> a ontolex:LexicalEntry, ontolex:Word;
ontolex:canonicalForm [ontolex:writtenRep "grading"@en];
skos:inScheme <aor:Hydrology_Lexicon_1>.
<aot:lex-levelling> a ontolex:LexicalEntry, ontolex:Word;
ontolex:canonicalForm [ontolex:writtenRep "levelling"@en];
skos:inScheme <aor:Hydrology_Lexicon_1>.
<aot:lex-تـسـويـة> a ontolex:LexicalEntry, ontolex:Word;
ontolex:canonicalForm [ontolex:writtenRep "تـسـويـة"@ar];
skos:inScheme <aor:Hydrology_Lexicon_1>.
```

Based On:

Mustafa Jarrar, Hamzeh Amayreh, John McCarae: **Progress on Representing Arabic Lexicons in Lemon**. The 2nd Conference on Language, Data and Knowledge (LDK 2019), Germany. 2019.

Conformance with W3C Standards

✓ **W3C's Best Practices for Publishing Linked Data** including the Cool URIs, simplicity, stability, and linking

URLs Schema:

- Each term is given a URL: `http://{domain}/term/{term}`

  http://ontology.birzeit.edu/term/virus

- Each lexical concept is given a URL:

  `http://{domain}/lexicalconcept/{lexicalConceptID}`

  https://ontology.birzeit.edu/lexicalconcept/304000682

- Each concept in the Arabic Ontology has a URL:

  `http://{domain}/concept/{ConceptID | Term}`

  https://ontology.birzeit.edu/concept/293262

- Each Semantic relation is given a URL:

  `http://{domain}/concept/{RelationName}/{ConceptID}`

  https://ontology.birzeit.edu/concept/instances/293121

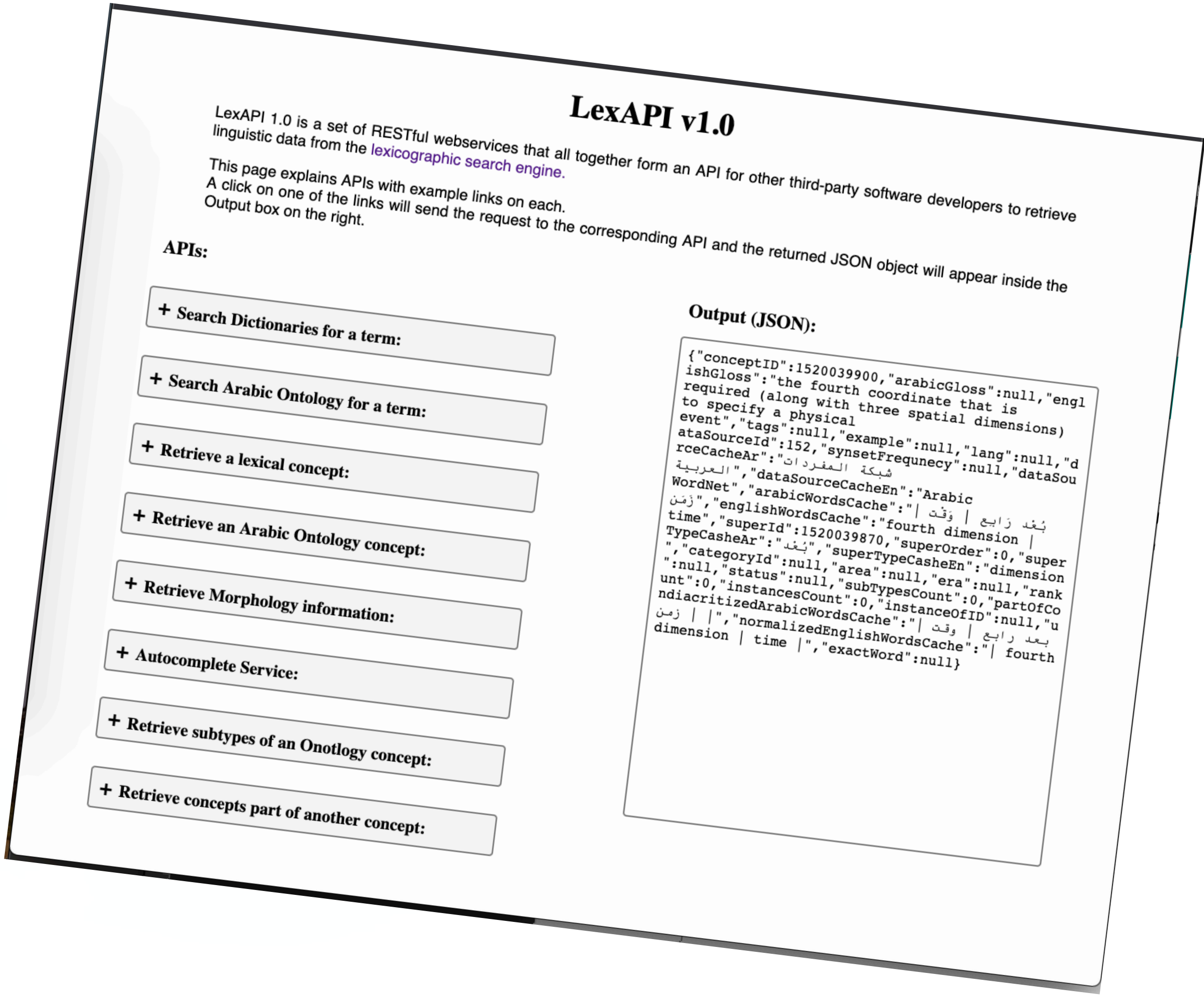
- The W3C Lemon representation of each lexical concept is given a URL: `http://{domain}/lemon/lexicalconcept/{lexicalConceptID}`

  https://ontology.birzeit.edu/lemon/lexicalconcept/304000682

API Access

RESTful web services

Ask us for an API Key!



LexAPI v1.0

LexAPI 1.0 is a set of RESTful webservices that all together form an API for other third-party software developers to retrieve linguistic data from the lexicographic search engine.

This page explains APIs with example links on each.
A click on one of the links will send the request to the corresponding API and the returned JSON object will appear inside the Output box on the right.

APIs:

- + Search Dictionaries for a term:
- + Search Arabic Ontology for a term:
- + Retrieve a lexical concept:
- + Retrieve an Arabic Ontology concept:
- + Retrieve Morphology information:
- + Autocomplete Service:
- + Retrieve subtypes of an Onotlogy concept:
- + Retrieve concepts part of another concept:

Output (JSON):

```
{"conceptID":1520039900,"arabicGloss":null,"engl
ishGloss":"the fourth coordinate that is
required (along with three spatial dimensions)
to specify a physical
event","tags":null,"example":null,"lang":null,"d
ataSourceId":152,"synsetFrequnecy":null,"dataSou
rceCacheAr":"شبكة المفردات
العربية","dataSourceCacheEn":"Arabic
WordNet","arabicWordsCache":"| بُعْد رَابِع | وَقْت
زَمَن","englishWordsCache":"fourth dimension |
time","superId":1520039870,"superOrder":0,"super
TypeCasheAr":"بُعْد","superTypeCasheEn":"dimension
","categoryId":null,"area":null,"era":null,"rank
":null,"status":null,"subTypesCount":0,"partOfCo
unt":0,"instancesCount":0,"instanceOfID":null,"u
ndiacritizedArabicWordsCache":"| بعد رابع | وقت
زمن | |","normalizedEnglishWordsCache":"| fourth
dimension | time |","exactWord":null}
```

Ranking Metrics

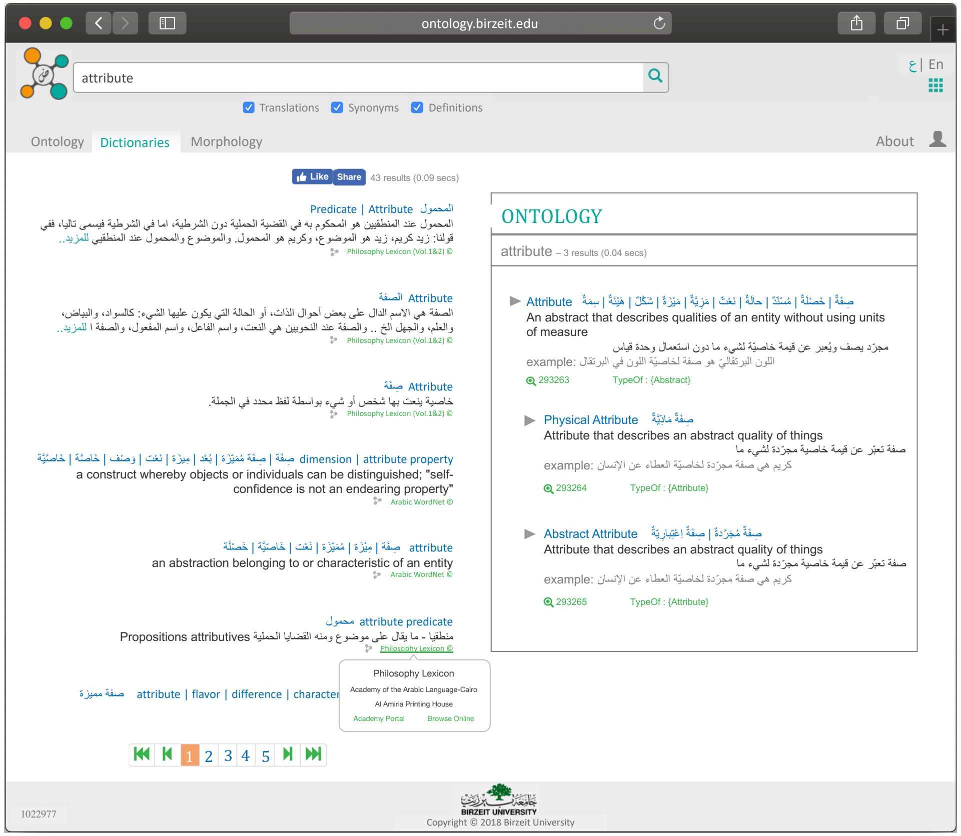
We developed three strategies:

- **Citation strategy (R_{cit})** frequency of the lexical concept terms:

$$R = \sum_{n=1}^{|A|} \sum_{m=1}^{k} F_{anm}$$

$$R_{cit} = \frac{R - R_{min}}{R_{max} - R_{min}}$$

- **Lexicon renown ranking strategy (R_{ren}):** experts assigned each lexicon a rank based on its renown.



attribute

☑ Translations ☑ Synonyms ☑ Definitions

43 results (0.09 secs)

Predicate | Attribute المحمول
المحمول عند المنطقيين هو المحكوم به في القضية الحملية دون الشرطية، اما في الشرطية فيسمى تاليا، ففي قولنا: زيد كريم، زيد هو الموضوع، وكريم هو المحمول. والموضوع والمحمول عند المنطقيي للمزيد..
Philosophy Lexicon (Vol.1&2) ©

Attribute الصفة
الصفة هي الاسم الدال على بعض أحوال الذات، أو الحالة التي يكون عليها الشيء: كالسواد، والبياض، والعلم، والجهل الخ .. والصفة عند النحويين هي النعت، واسم الفاعل، واسم المفعول، والصفة ا للمزيد..
Philosophy Lexicon (Vol.1&2) ©

Attribute صِفَة
خاصية ينعت بها شخص أو شيء بواسطة لفظ محدد في الجملة.
Philosophy Lexicon (Vol.1&2) ©

attribute property | dimension صِفَة | صِفَة مُمَيّزَة | بُعْد | مِيزَة | نَعْت | وَصْف | خَاصَّة | خَاصَّيَّة
a construct whereby objects or individuals can be distinguished; "self-confidence is not an endearing property"
Arabic WordNet ©

attribute صِفَة | مِيْزَة | مُمَيّزَة | نَعْت | خَاصّيَّة | خَصْلَة
an abstraction belonging to or characteristic of an entity
Arabic WordNet ©

attribute predicate محمول
منطقيا - ما يقال على موضوع ومنه القضايا الحملية Propositions attributives
Philosophy Lexicon ©

Philosophy Lexicon
Academy of the Arabic Language-Cairo
Al Amiria Printing House
Academy Portal Browse Online

attribute | flavor | difference | character صفة مميزة

1 2 3 4 5

ONTOLOGY

attribute – 3 results (0.04 secs)

▶ Attribute صِفَةٌ | خَصْلَةٌ | مُسْنَدٌ | حالَةٌ | نَعْتٌ | مَزِيَّةٌ | مَيْزَةٌ | شَكْلٌ | هَيْئَةٌ | سِمَةٌ
An abstract that describes qualities of an entity without using units of measure
مجرّد يصف ويُعبر عن قيمة خاصيّة لشيء ما دون استعمال وحدة قياس
example: اللون البرتقاليّ هو صفة لخاصيّة اللون في البرتقال
293263 TypeOf : {Abstract}

▶ Physical Attribute صِفَةٌ مَادِّيَّةٌ
Attribute that describes an abstract quality of things
صفة تعبّر عن قيمة خاصية مجرّدة لشيء ما
example: كريم هي صفة مجرّدة لخاصيّة العطاء عن الإنسان
293264 TypeOf : {Attribute}

▶ Abstract Attribute صِفَةٌ مُجَرَّدَةٌ | صِفَةٌ إعْتِبارِيَّةٌ
Attribute that describes an abstract quality of things
صفة تعبّر عن قيمة خاصية مجرّدة لشيء ما
example: كريم هي صفة مجرّدة لخاصيّة العطاء عن الإنسان
293265 TypeOf : {Attribute}

1022977

BIRZEIT UNIVERSITY

Copyright © 2018 Birzeit University

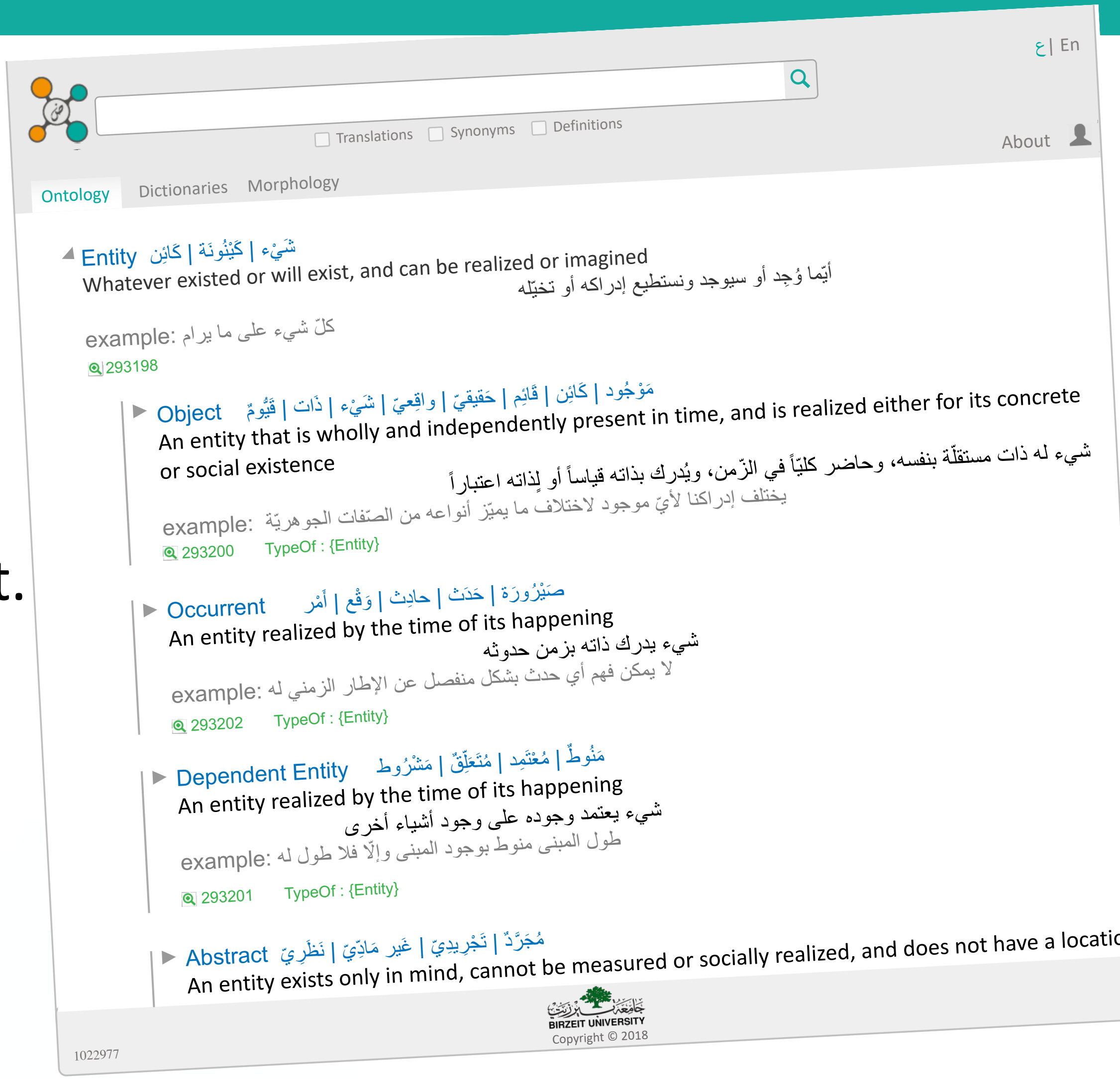
- **Hybrid ranking strategy (R_{hyb})** is a combination metric:

$$R_{hyb} = R_{ren} + R_{cit}$$

The Arabic Ontology

Arabic Ontology

- Classification of the meanings of the Arabic terms, specified in D. Logic
- Can be used as a formal Arabic Wordnet -with ontologically-clean content.
- Linked with WordNet, WikiData, BFO, DOLCE
- Benchmarked to scientific advances rather than to speakers' naïve beliefs as wordnets do.



◢ Entity شَيْء | كَيْنُونَة | كَائِن
Whatever existed or will exist, and can be realized or imagined
أيّما وُجِد أو سيوجد ونستطيع إدراكه أو تخيّله

example: كلّ شيء على ما يرام
293198

▶ Object مَوْجُود | كَائِن | قَائِم | حَقيقيّ | واقِعيّ | شَيْء | ذَات | قَيُّومٌ
An entity that is wholly and independently present in time, and is realized either for its concrete or social existence
شيء له ذات مستقلّة بنفسه، وحاضر كليّاً في الزّمن، ويُدرك بذاته قياساً أو لِذاته اعتباراً

example: يختلف إدراكنا لأيّ موجود لاختلاف ما يميّز أنواعه من الصّفات الجوهريّة
293200 TypeOf : {Entity}

▶ Occurrent صَيْرُورَة | حَدَث | حادِث | وَقْع | أَمْر
An entity realized by the time of its happening
شيء يدرك ذاته بزمن حدوثه

example: لا يمكن فهم أي حدث بشكل منفصل عن الإطار الزمني له
293202 TypeOf : {Entity}

▶ Dependent Entity مَنُوطٌ | مُعْتَمِد | مُتَعَلِّقٌ | مَشْرُوط
An entity realized by the time of its happening
شيء يعتمد وجوده على وجود أشياء أخرى

example: طول المبنى منوط بوجود المبنى وإلّا فلا طول له
293201 TypeOf : {Entity}

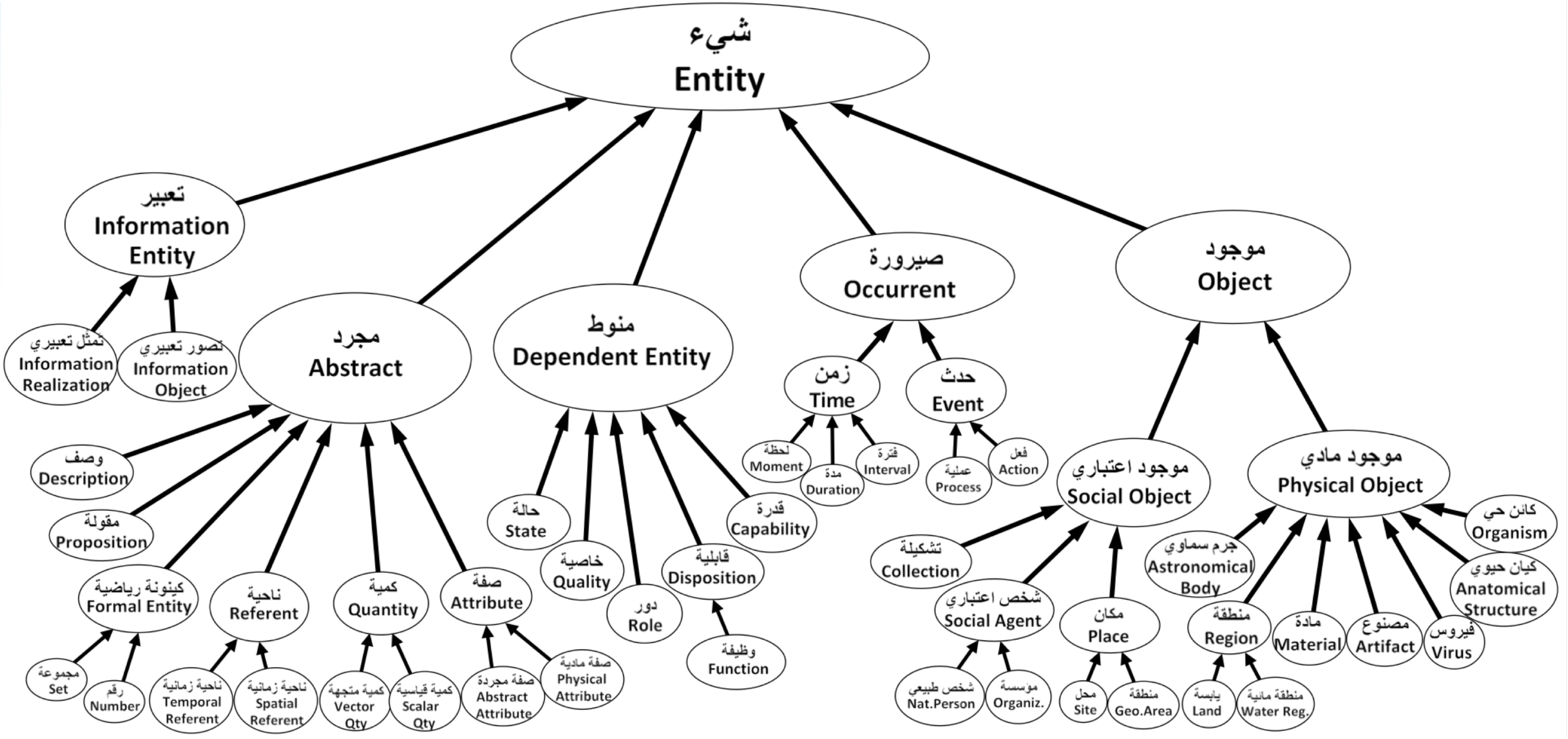
▶ Abstract مُجَرَّدٌ | تَجْرِيدِيّ | غَير مَادِّيّ | نَظَرِيّ
An entity exists only in mind, cannot be measured or socially realized, and does not have a locatic

BIRZEIT UNIVERSITY
Copyright © 2018

1022977

https://ontology.birzeit.edu/concept/293198

Top Levels of the Arabic Ontology



شيء
Entity

تعبير
Information Entity

تمثل تعبيري
Information Realization

تصور تعبيري
Information Object

مجرد
Abstract

وصف
Description

مقولة
Proposition

كينونة رياضية
Formal Entity

مجموعة
Set

رقم
Number

ناحية
Referent

ناحية زمانية
Temporal Referent

ناحية زمانية
Spatial Referent

كمية
Quantity

كمية متجهة
Vector Qty

كمية قياسية
Scalar Qty

صفة
Attribute

صفة مجردة
Abstract Attribute

صفة مادية
Physical Attribute

منوط
Dependent Entity

حالة
State

خاصية
Quality

دور
Role

قابلية
Disposition

وظيفة
Function

قدرة
Capability

صيرورة
Occurrent

زمن
Time

لحظة
Moment

مدة
Duration

فترة
Interval

حدث
Event

عملية
Process

فعل
Action

موجود
Object

موجود اعتباري
Social Object

تشكيلة
Collection

شخص اعتباري
Social Agent

شخص طبيعي
Nat.Person

مؤسسة
Organiz.

مكان
Place

محل
Site

منطقة
Geo.Area

موجود مادي
Physical Object

جرم سماوي
Astronomical Body

منطقة
Region

يابسة
Land

منطقة مائية
Water Reg.

مادة
Material

مصنوع
Artifact

فيروس
Virus

كيان حيوي
Anatomical Structure

كائن حي
Organism

Based on:

Mustafa Jarrar: **The Arabic Ontology - An Arabic Wordnet with Ontologically Clean Content**. Applied Ontology Journal, IOS Press. (Forthcoming).

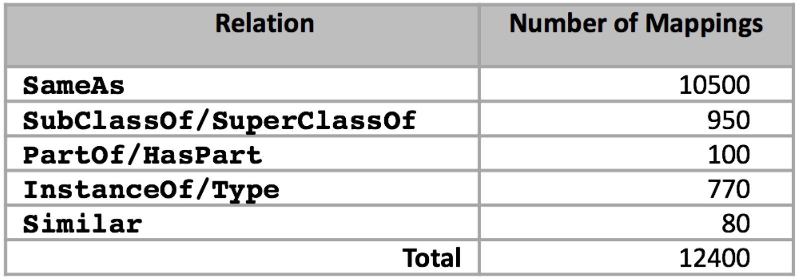
Linking Lexicons with the Arabic Ontology

- Lexical concepts (in lexicons) are interlinked with the entities in the ontology.

- Given two entities e_1 and e_2, a *mapping correspondence* between them is defined as the following:

$$< e_1, e_2, R, P, C>$$

- Progress so far:



Relation	Number of Mappings
SameAs	10500
SubClassOf/SuperClassOf	950
PartOf/HasPart	100
InstanceOf/Type	770
Similar	80
Total	12400

- In this way, lexical concepts across all lexicons would be semantically linked

Work in Progress

- Lemmatize each lexical entry in every lexicon
- Lemon-izing and Interlinking Arabic resources with the Linguistic Linked Open Data Cloud
- Building an Arabic Knowledge Graph
- Collect and interlink dialect corpora

References

1. Mustafa Jarrar. The Arabic Ontology - An Arabic Wordnet with Ontologically Clean Content. Applied Ontology Journal, 2019 [Forthcoming].
2. Mustafa Jarrar, Hamzeh Amayreh: An Arabic-Multilingual Database with a Lexicographic Search Engine. The 24th International Conference on Applications of Natural Language to Information Systems (NLDB 2019). Pages(234--246), LNCS 11608, Springer. 2019.
3. Mustafa Jarrar, Hamzeh Amayreh, John McCarae: Representing Arabic Lexicons in Lemon - a Preliminary Study. The 2nd Conference on Language, Data and Knowledge (LDK 2019), Germany. 2019.
4. Hamzeh Amayreh, Mohammad Dwaikat, and Mustafa Jarrar. Lexicon Digitization -A Framework for Structuring, Normalizing and Cleaning Lexical Entries, Technical Report, Birzeit University 2018.
5. Mustafa Jarrar: Search Engine for Arabic Lexicons. Proceedings of the 5th Conference on Translation and the Problematics of Cross-cultural Understanding. The Forum for Arab and International Relations. Doha, Qatar. 2018
6. Mustafa Jarrar, Fadi Zaraket, Rami Asia, Hamzeh Amayreh: Diacritic-Based Matching of Arabic Words. ACM Asian and Low-Resource Language Information Processing. Volume 18, No 2, Pages(10:1--10:21), ACM, December 2018.
7. Mustafa Jarrar, Nizar Habash, Faeq Alrimawi, Diyam Akra, and Nasser Zalmout. Curras: An Annotated Corpus for the Palestinian Arabic Dialect. Journal Language Resources and Evaluation, 51(3):745–775, 2017.
8. Mustafa Jarrar, Nizar Habash, Diyam Akra, Nasser Zalmout: Building a Corpus for Palestinian Arabic: a Preliminary Study. In proceedings of the EMNLP 2014 Workshop on Arabic Natural Language Processing. Association for Computational Linguistics (ACL), Pages (18-27). October 25, 2014, Doha, Qatar. ISBN: 978-1-937284-96-1
9. Mustafa Jarrar. Building a Formal Arabic Ontology (Invited Paper). In Proceedings of the Experts Meeting on Arabic Ontologies and Semantic Networks. ALECSO, Arab League, 2011