Linguistic Ontologies for Knowledge Graphs

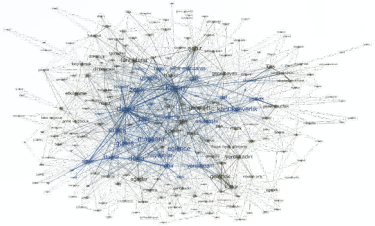
Mustafa Jarrar
Birzeit University
Palestine


Building linguistic resources for NLP

جامعة بيرزيت
BIRZEIT UNIVERSITY

Lexical Resources at Birzeit University

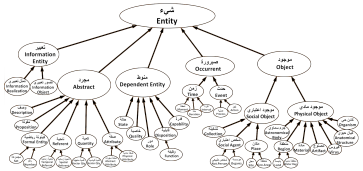
Lexicographic Database



150 lexicons

Very large Arabic-multilingual database

Arabic Ontology



Formal Arabic Wordnet

with ontologically clean content

Dialect Corpora

Annotated corpora

each word is annotated with many morph features

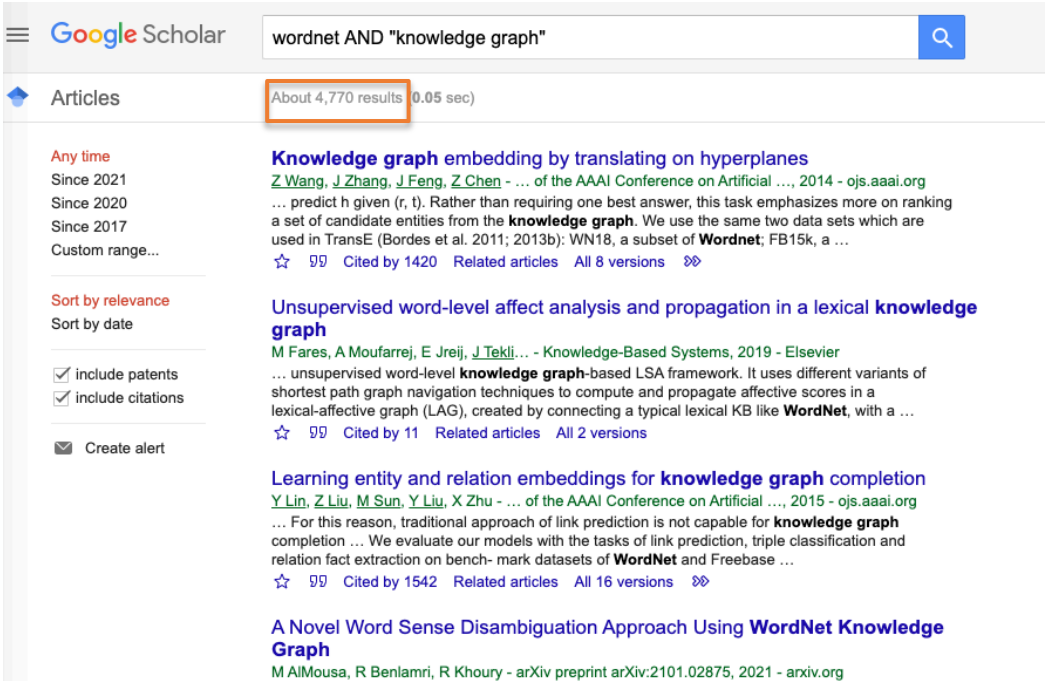
Big Linguistic Data Graph

https://ontology.birzeit.edu

Motivation

- Wordnets are used to understand and retrieve unstructured information in e.g., NLP and IR.

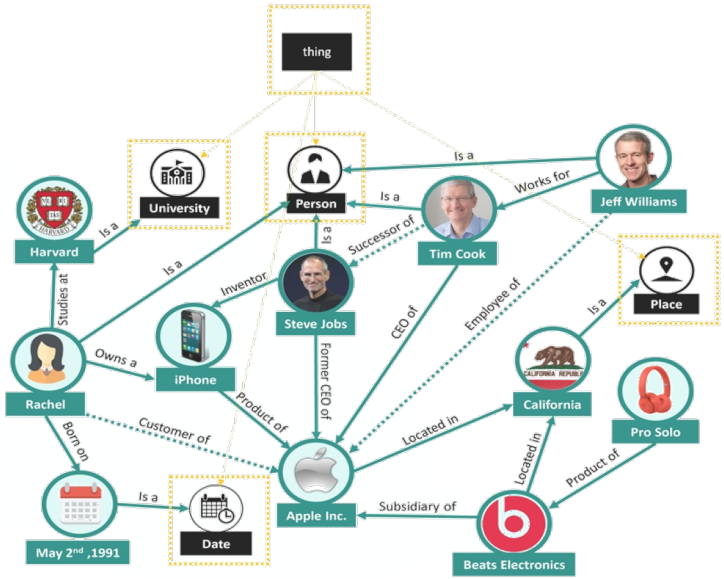
- New demands to use wordnets like ontologies: to manage and retrieve structured data in e.g., Knowledge Graphs, multilingual Big Data, and medical informatics.



Motivation

- Wordnets are used to understand and retrieve unstructured information in e.g., NLP and IR.

- New demands to use wordnets like ontologies: to manage and retrieve structured data in e.g., Knowledge Graphs, multilingual Big Data, and medical informatics.



Source: https://blogs.timbr.ai/introducing-timbr-sql-kg

Motivation

But,

Ontologies are typically application-specific rich axiomatizations;
Wordnets are general-purpose mental lexicons, and thus axiomatizing wordnet would be a **rigidification**.

- How to build a linguistic ontology as a wordnet - to better serve new application scenarios

Wordnets Vs Ontologies

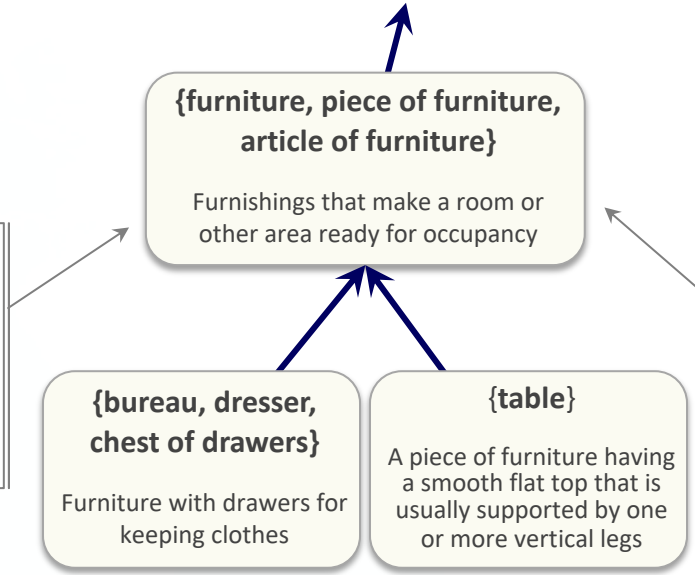
Synset vs Concept

Wordnet

made of synsets
(linguistic concepts)

Synset: *signifies a concept*; a thought in our mind.

Individuals are also linguistic concepts (ISO 1087-1:2000)



Ontology

made of concepts
(classes of individuals).

Concept: *class of individuals*; characteristics its instances have in common.

In BFO, Universal/Defined Class.

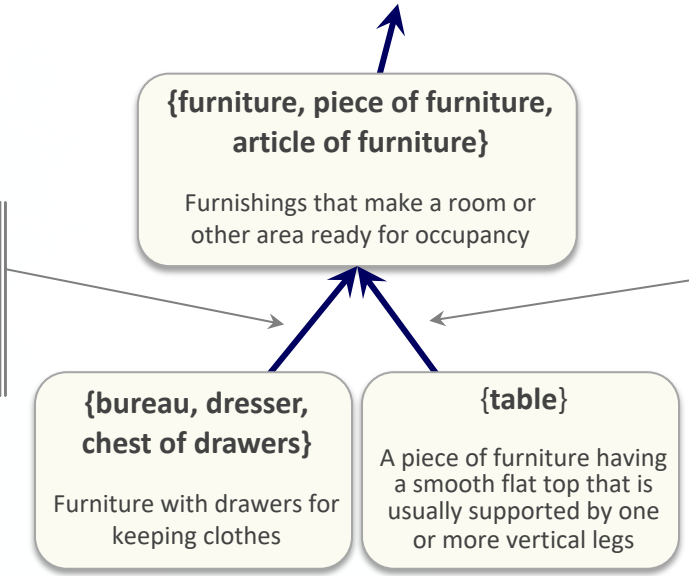
Proposed definition (for linguistic ontologies)

Definition: concept (Jarrar, 2021):
Given a concept c, its intensional interpretation c^I is defined on a *domain space* $\langle D, W \rangle$ as a function $c^I : W \rightarrow 2^D$, where D is a domain and W is a set of maximal states of affairs on D. For a concept c, the set $E_c = \{c^I(w) \mid w \in W\}$ is the set of the admissible extensions of c. Two concepts having the same set of *admissible* instances, in all states of affairs, are considered the same concept.

Hyponymy vs Subsumption

Wordnet

Hyponymy: If native speakers accept a sentence like: B is a kind of A



Ontology

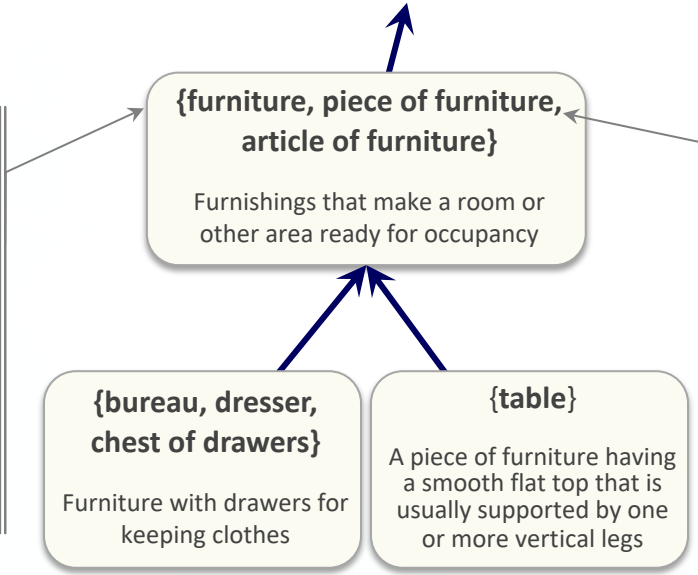
Subsumption: a subset relation between concepts. Every instance in c_1 is also an instance in c_2.

Formally: c_1 subsumes c_2, *iff* every instance of c_2 is an instance of c_1, in every possible state of affairs.

Synonymy

Wordnet

Synonyms: "two expressions are synonymous in a linguistic context C if the substitution of one for the other in C does not alter the truth value" (Miller et al., 1990).



Ontology

Synonyms: alternative labels/names of concepts.

Proposed definition (for linguistic ontologies)

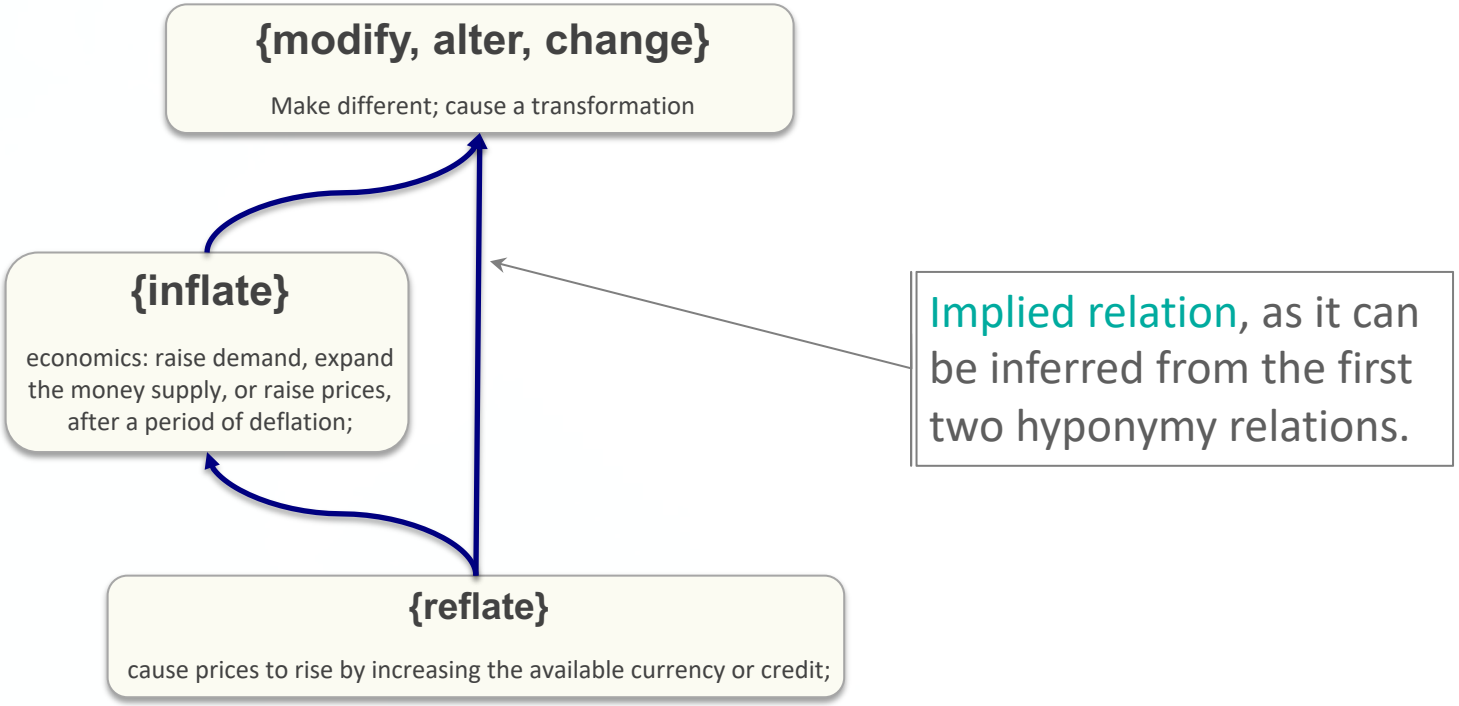
Definition: Synonymy Relation (Jarrar, 2021)
Given two terms t_1 and t_2 lexicalizing concepts c_1 and c_2, respectively, then t_1 and t_2 are considered to be synonymous *iff* $c_1 = c_2$. In this way, synonymy can be defined as an equivalence relation $=_c$ between terms lexicalizing the same concept, thus it is a reflexive, symmetric and transitive relation.

More ontology/formal issues in wordnet:

Which might not be correct from a formal/ontological perspective

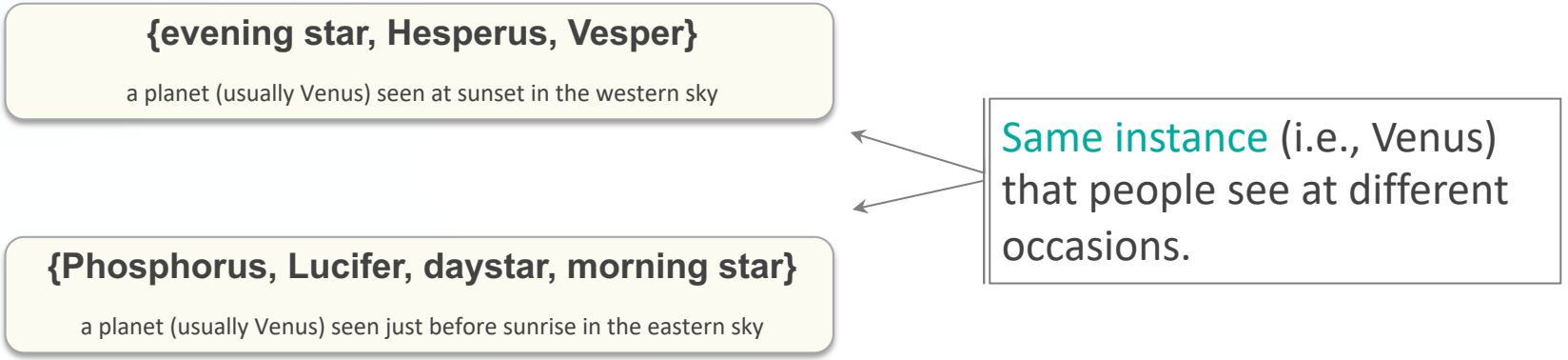
Formal/Ontological issues in wordnet

No benefits for including implied relations



Formal/Ontological issues in wordnet

Might be different linguistic concepts, but ontologically it is the same instance.

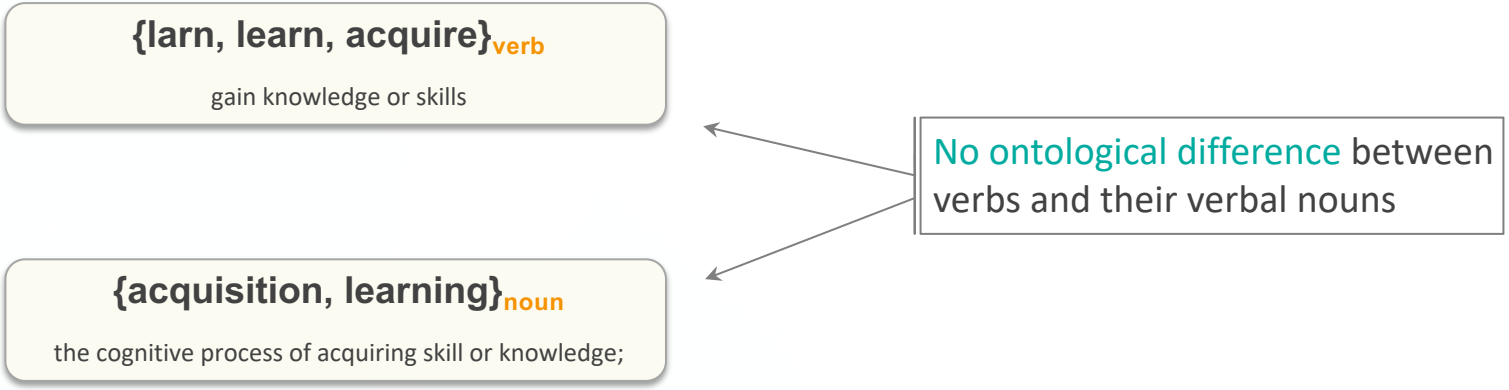


Formal/Ontological issues in wordnet

Verbs are linguistic rather than ontological categories.

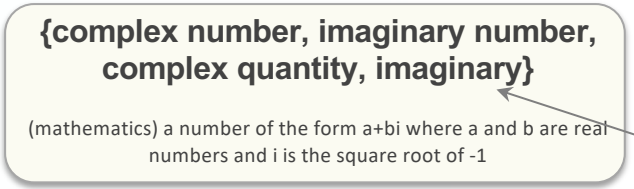
Ontologies capture the events that verbs denote rather than verbs themselves.

➔ We say (he learns, he learned, he is learning, the learning he ..,) referring to the exact same learning event.



Formal/Ontological issues in wordnet

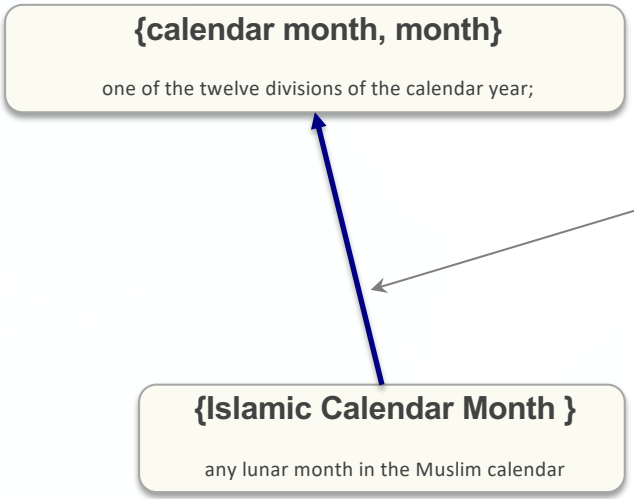
Who/How to decide on the accuracy?



not synonyms, Imaginary number is only a special case of a Complex number. Similarly, WordNet provides a poor classification of the types of numbers, e.g., Real, Rational, Natural, and Integer numbers are all subsumed by Number, while they subsume each other.

Formal/Ontological issues in wordnet

Who/How to decide on the accuracy?



Inaccurate, because Month is defined as one of the twelve divisions of the *calendar year*, which is a Gregorian year as defined in the gloss (i.e., an average month is about 30.43 days); however, an average of an Islamic month is 29.53 days. In fact, both months belong to two different calendar systems. A Gregorian month is 1/12 of a Gregorian year, while an Islamic year is a multiple of 12 lunar months.

Application Ontology vs Linguistic Ontology /Formal Wordnet

- Typically rich axiomatization
- Each term refers to one concept (no polysemy).
- Synonymy is not a target.
- Benchmarked to application's knowledge.
- Used by a certain application or a class of applications.

- Light-weight axiomatization, and cannot be rigid.
- Each term refers to one or more concepts (Polysemy).
- Synonymy is important.
- Benchmarked general knowledge
- Used for general purposes.

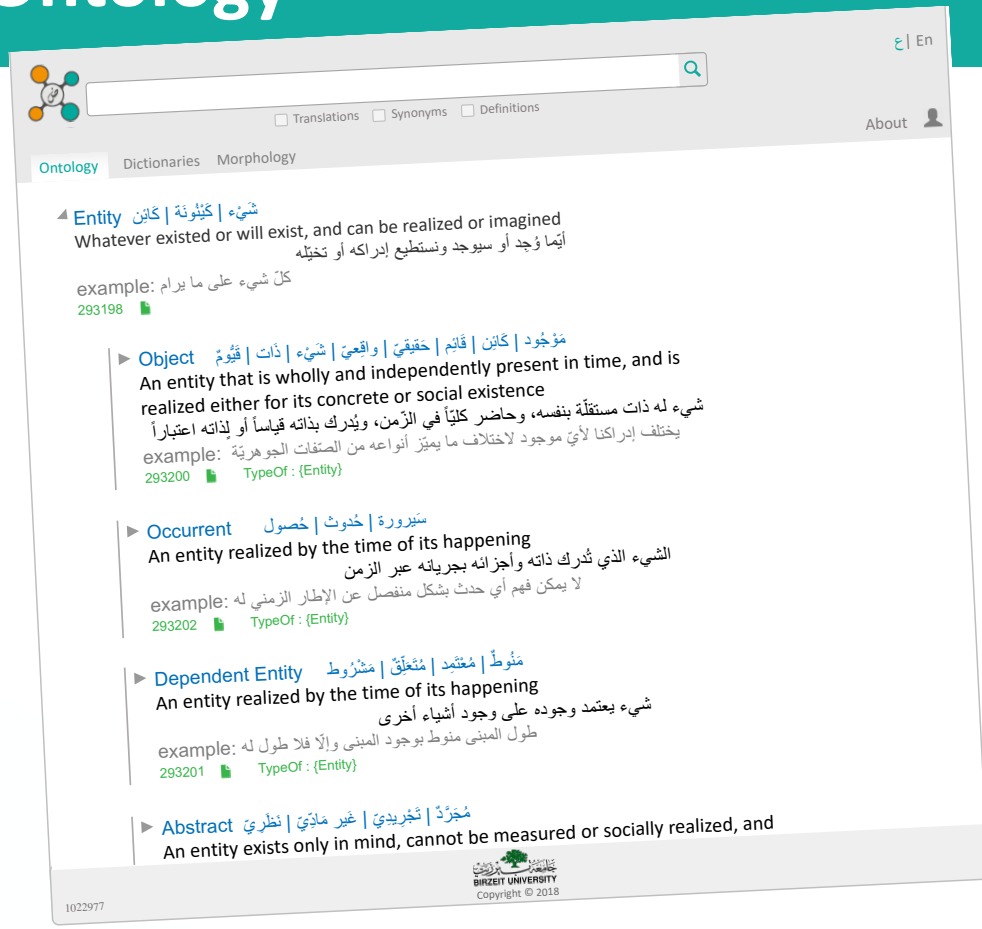
The Arabic Ontology

Arabic Ontology

- Characterization of the intended meaning (i.e., concepts) that the Arabic terms convey.

- Formal Arabic Wordnet - with ontologically-clean content.

- Linked with PWN, Wikidata, BFO, DOLCE
- Linked with many lexicons



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

Arabic Ontology

- Current size so far (but the numbers are dynamic)

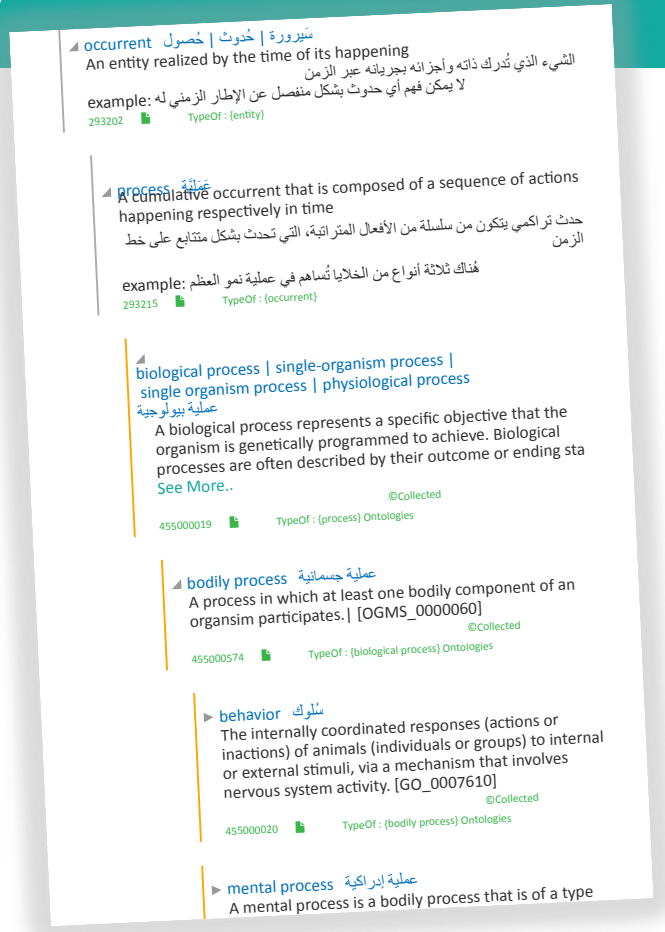
  1800 fully-done concepts (mostly top levels)

  17K partially investigated (ready for NLP applications)

  Some branches are elaborated, other not yet.

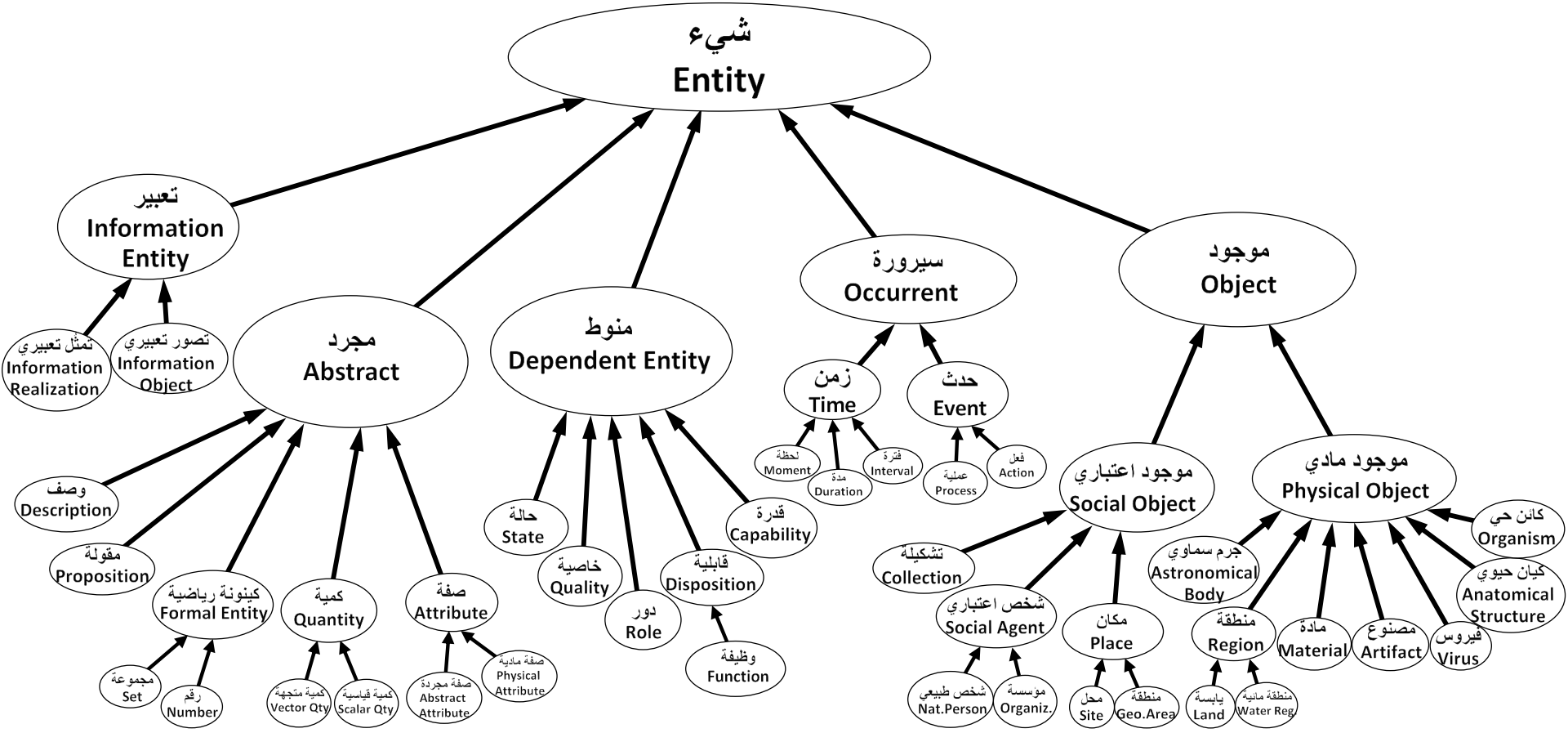
- English labels are not our target - provided for readability and communication.

- Methodology: Built top-down and bottom-up at the same time.



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

Top Levels of the Arabic Ontology

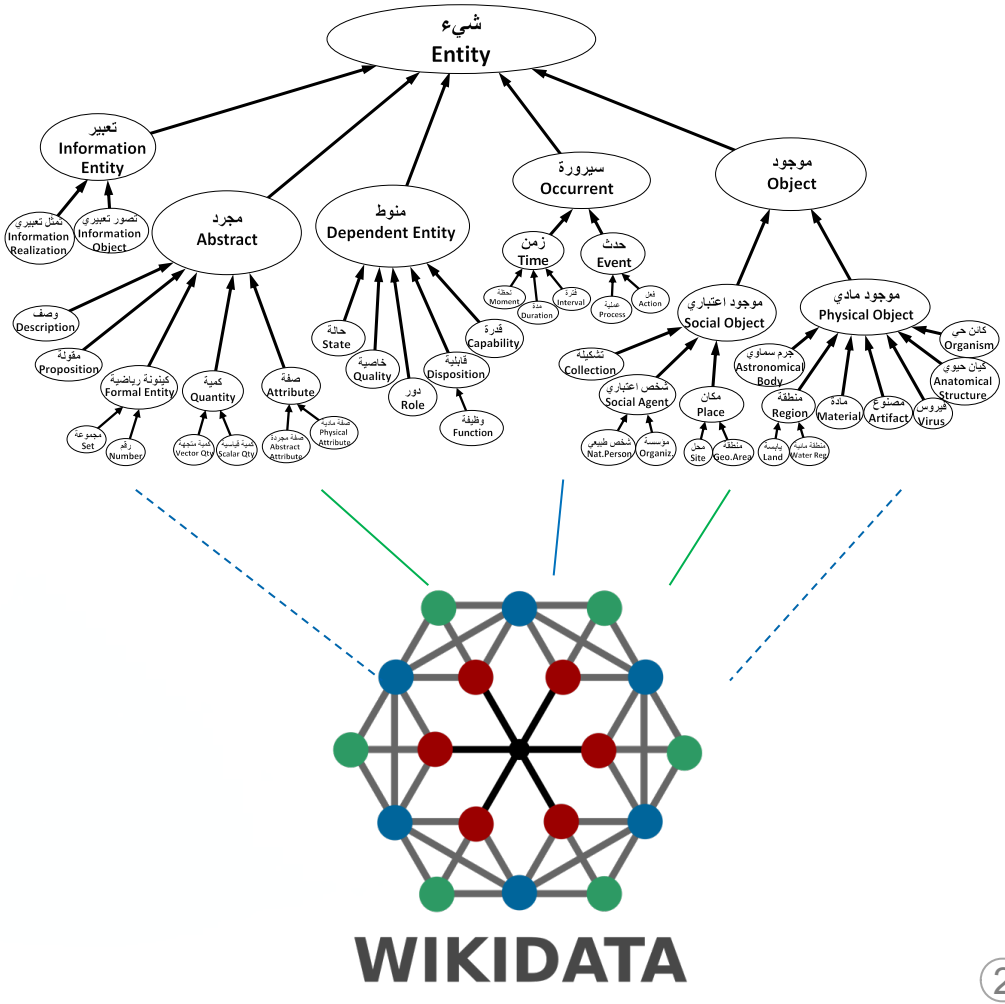


Linking with Wikidata

Every Concept in the ontology is mapped mapped with a Wikidata node.

Thus, all instances in Wikidata are instances in the ontology

Thus, a knowledge Graph from the Arabic Ontology viewpoint.

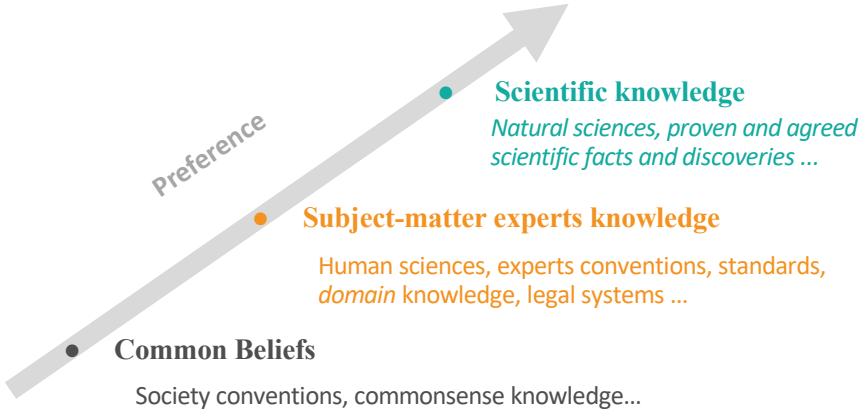


Benchmarking the ontology content

- **What should the ontology capture and adhere to?**
- **On what basis the correctness of the ontology content can be benchmarked?**

- Should concepts be defined/classified based on what Arabic speakers commonly believe!
- Should we adopt a certain lexicon and formalize it!
- Should we rely on what the scientific literature accepts!
- Should we build the ontology based on what we, the ontology builders, believe!

Benchmarking Methodology



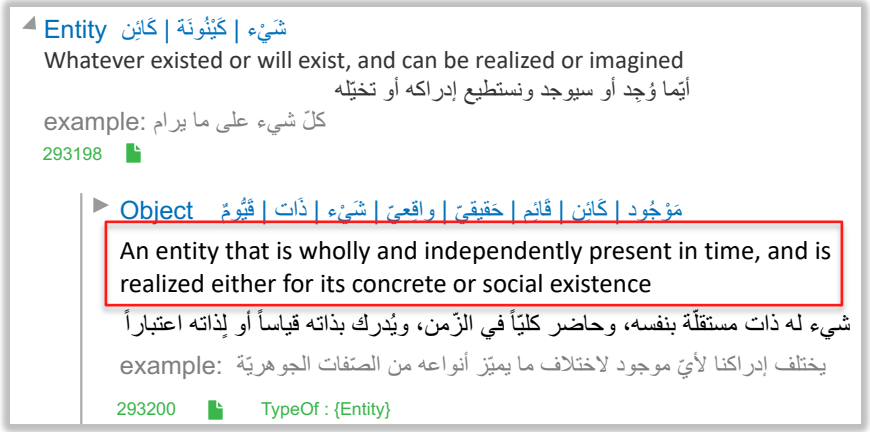
Benchmarked against the following, in order:

1. **Scientific knowledge**, which scientists typically accept on the basis of experimentation and verification and commonly agree about. If no mature answer is found in the state-of-art scientific discoveries, then against,
2. **Subject-matter experts' and domain knowledge and conventions**. If no answer can be synthesized or attained from experts' knowledge, then against,
3. **Commonsense knowledge**, repeatedly found in quality lexicons and literature.

Gloss Formulation Methodology

The purpose of a gloss state the critical and distinguishing characteristics that all instances of a concept have in common, in an informal but controlled way as the following:

Step 1: Start with the supertype of the concept being defined. e.g., “Object: An entity that…”, “Physical Object: An object that …”.

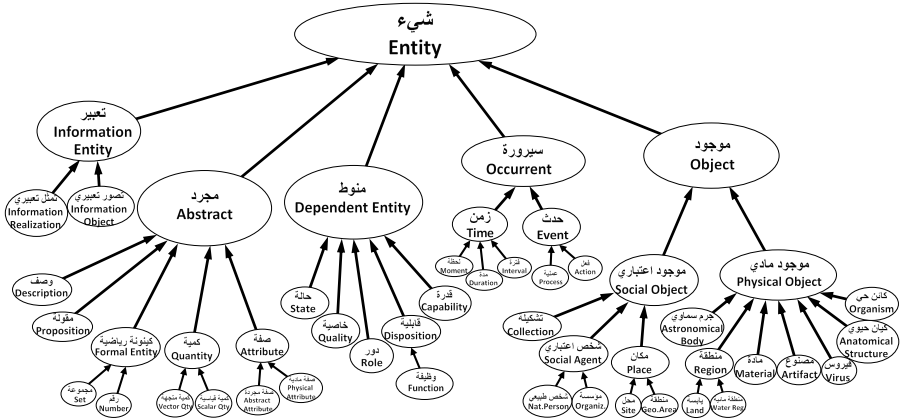


Step 2: List only the most distinguishing and intrinsic characteristics that specialize the concept from its supertype, and that differentiate it from other concepts in the same level.

Step 3: Write the distinguishing characteristics in the form of a sequence of propositions to help the reader to easily mentally rebuild the concept being defined in a declarative and non-narrative manner.

Comprehensiveness Evaluation

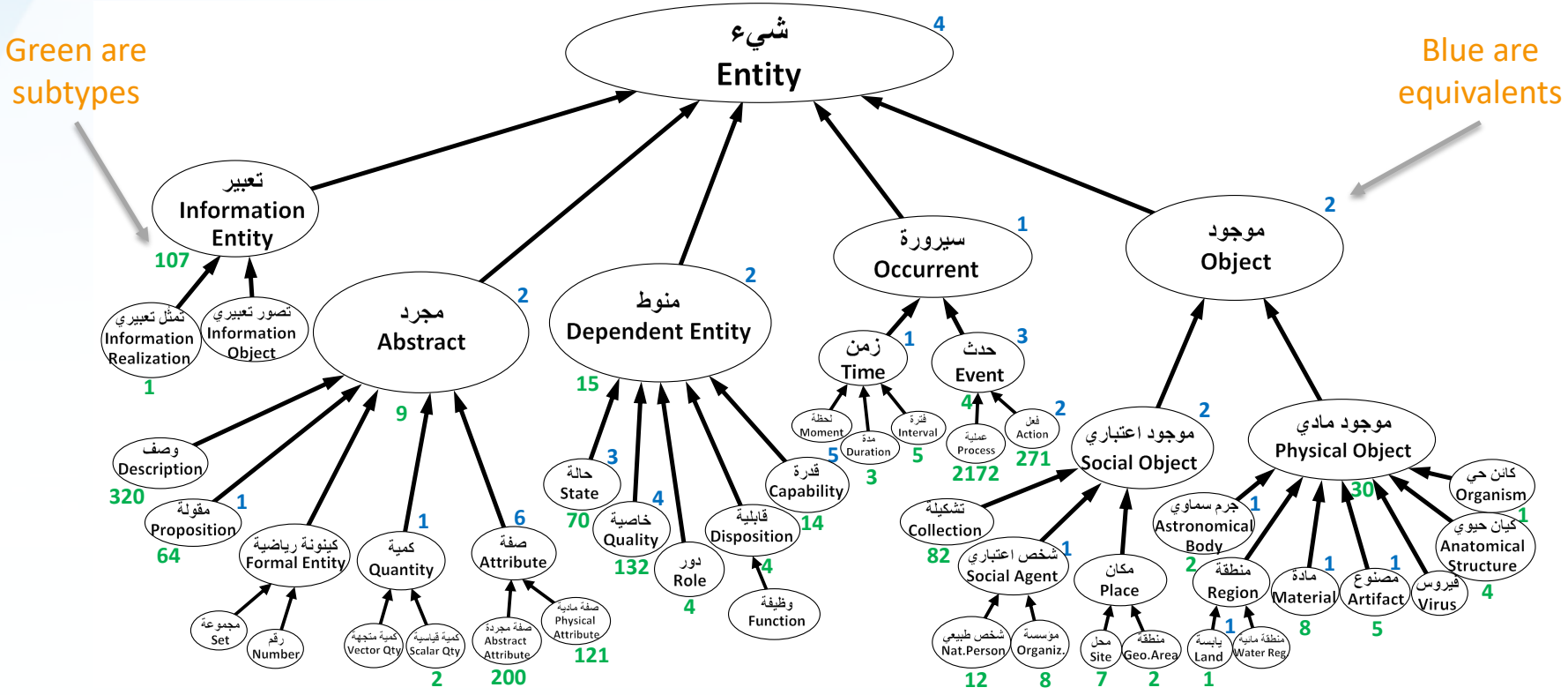
How much is the current version of the ontology able to *top* (i.e., be a supertype of) the concepts of the Arabic terms?



Experiment: classify the 1830 terms found in Al-Jurjānī "Taʿrīfāt" lexicon (1339–1414 AD) under the ontology. This lexicon contains the most abstract notions in Arabic in most domains. 270 concepts are excluded (not understand, can be instantiated , etc).

Ideally: each of the 1830 concepts should be placed either as "equivalent to" a node in the ontology, or as a "*subtype* of" a *leaf node*.

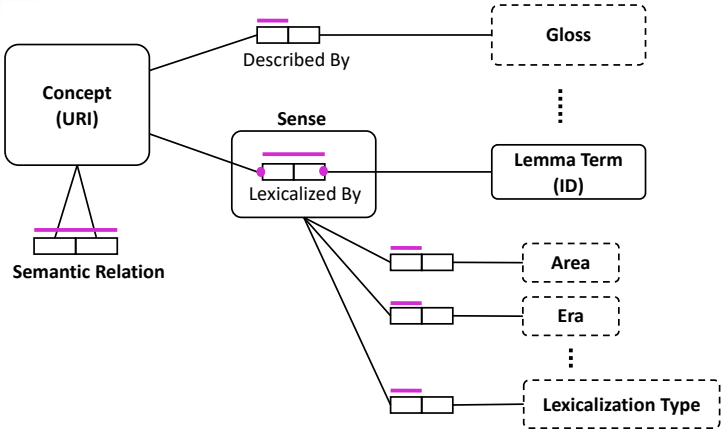
Comprehensiveness Results



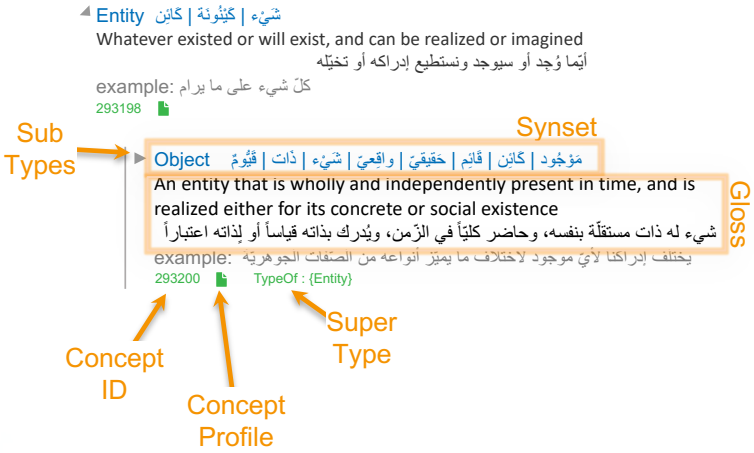
- 1655 (90%) concepts that are correctly placed in the ontology.
- 156 (10%) concepts that are mapped as subclasses of non-leaf nodes illustrate cases of missing top categories in the ontology (we are working on adding them).

Data Representation

Meta Core Data Model



Ontology Portal



➔ Also accessible in the RDF W3C Lemon format

URIs Design

To be linked with other resources in **the Linguistic Linked Open Data Cloud**, the URLs should designed according to the **W3C's Best Practices for Publishing Linked Data**, as the following:

***Concepts*:** each concept is given a URL based on its unique ConceptID:

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

e.g., http://ontology.birzeit.edu/concept/293254

Semantic relations: to allow one to retrieve, e.g. the instances, of a given conceptID, the semantic relations for a given concept can be accessed through URLs:

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

e.g., http://ontology.birzeit.edu/concept/instances/293121
http://ontology.birzeit.edu/concept/parts/293121

***Terms*:** each term is given a URL, which refers to the set of concepts that are lexicalized using a certain term, i.e., that have this *term* among their synsets:

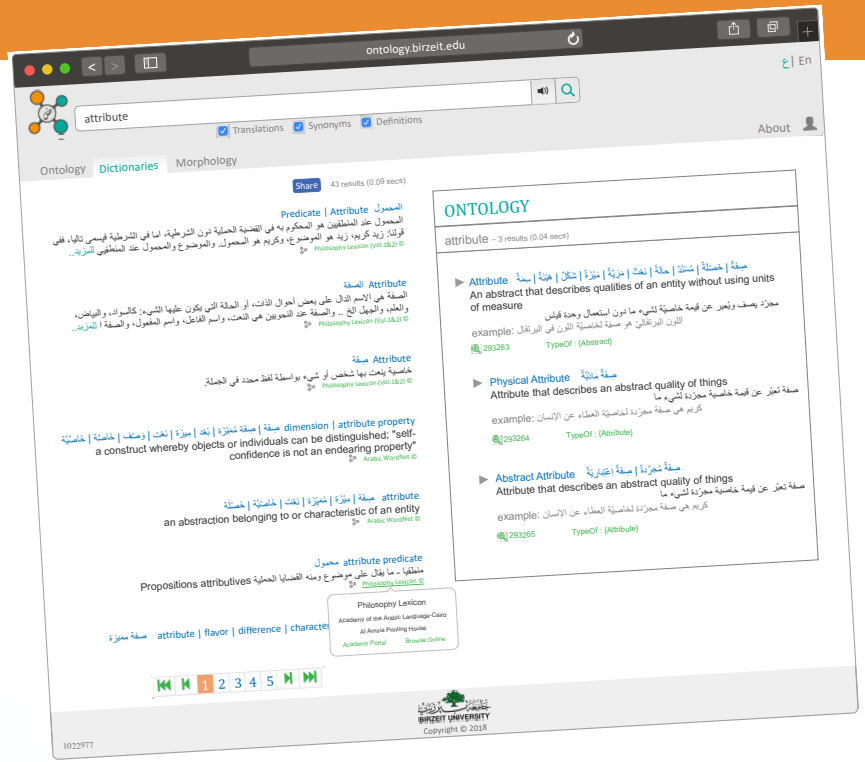
`http://{domain}/concept/{term}`

e.g., http://ontology.birzeit.edu/concept/virus

Overview of the Lexicographic Search Engine

The Lexicographic Database

- **The largest lexicographic Arabic-multilingual database**
- **Contains 150 lexicon, types**: glossaries, thesauri, bi/trilingual dictionaries, morph datasets, Ontology, and more.
- **Covers most domains:** science, technology, law, business, art, philosophy, ...



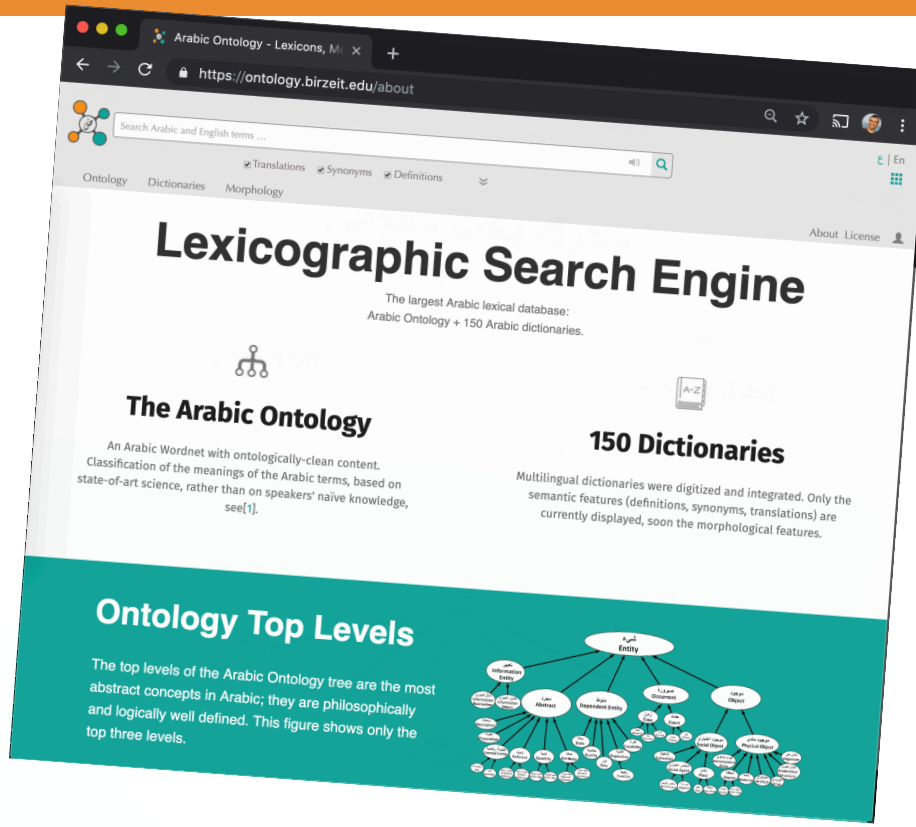
https://ontology.birzeit.edu

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

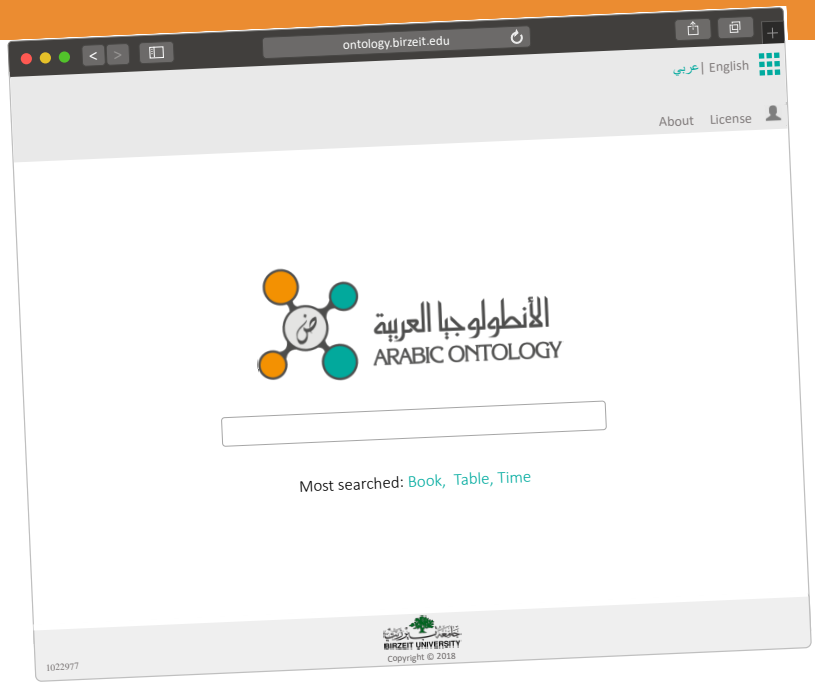
- **Search 150 lexicons** for definitions, synonyms, specialized translations, morphology, ontology [3,4] ...
- **Accurate**! compared with machine translation.
- **The first of its kind**! e.g., there are no similar search engines for English lexicons!



https://ontology.birzeit.edu

Lexicographic Search Engine

- **Free access to people**: students, translators, researchers, Arabic learners …
- **API accessible** for NLP applications.



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.

Some Statistics

Currently!

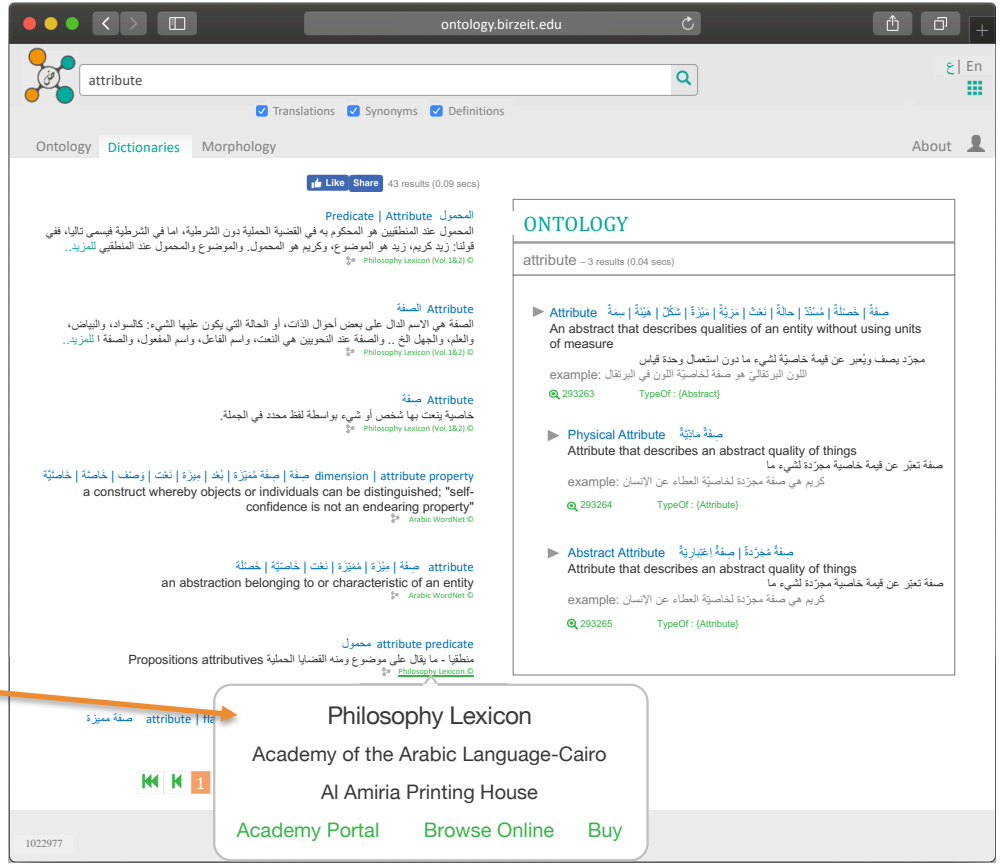
Category	Lexical Concepts	Lexical entries	Synsets	Translations pairs	Glosses	Semantic relations
Total (Millions)	1.1 M	2.4 M	1.8 M	1.5 M	0.7 M	0.5 M
Sub Counts		1,100K Arabic 1,100K English 200K French 3K Others 1,300K Single-word 1,000K Multi-word	800K Arabic 800K English 200K French 50K Others	1,000K 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

Reference:

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

Obtaining Copyrights

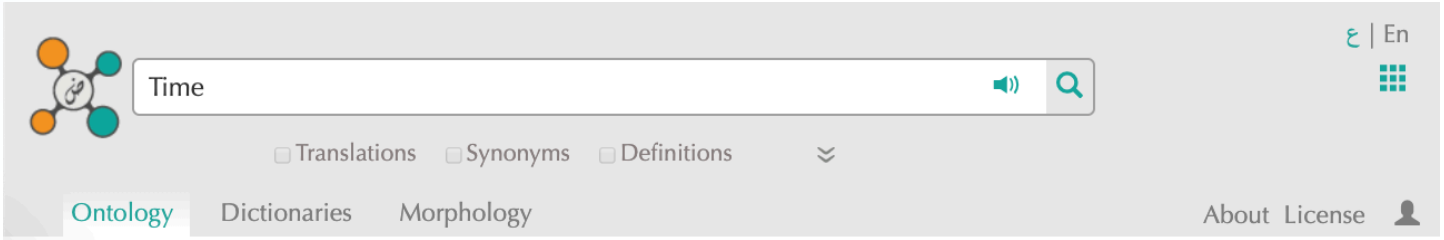
- Obtained permission from each lexicons owner (individually contacted).
- Most accepted!
- Show lexicon name and © copyright symbol beside each result.
- Promote lexicons (click to see lexicon info)



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.

See [4,5] about the Arabic Ontology

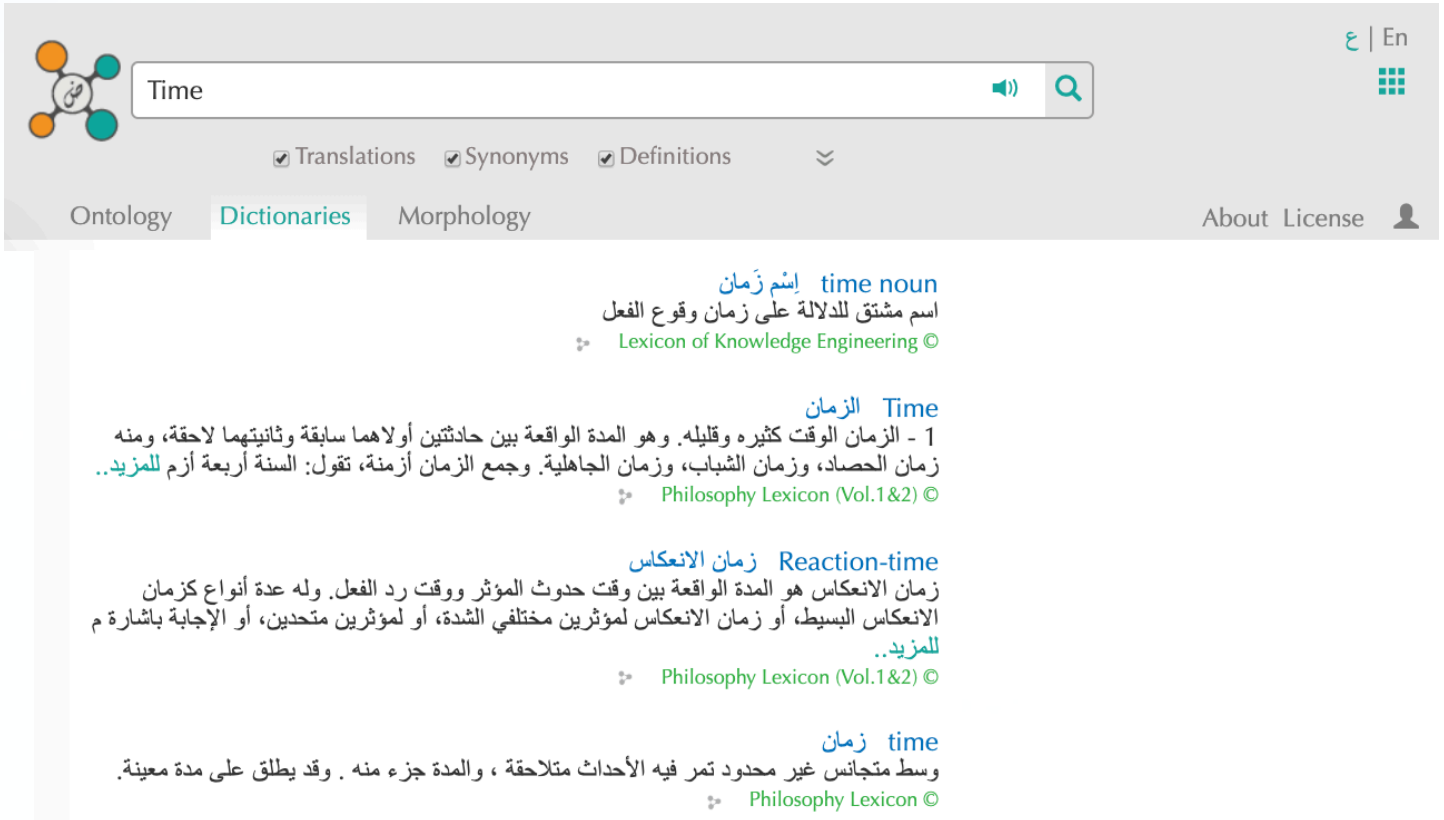


▶ time مُدَّةٌ | زَمَنٌ | فَتْرَةٌ
An occurrent representing a region in the timeline, realized by its starting and ending points, its length represents the temporal dimension of events or objects
مجرّد يمثّل جزءاً من خطّ الزّمن المدرك، تدرك ذاته بنقطة بداية ونقطة نهاية طولها يمثّل زمن أحداث أو موجودات
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}

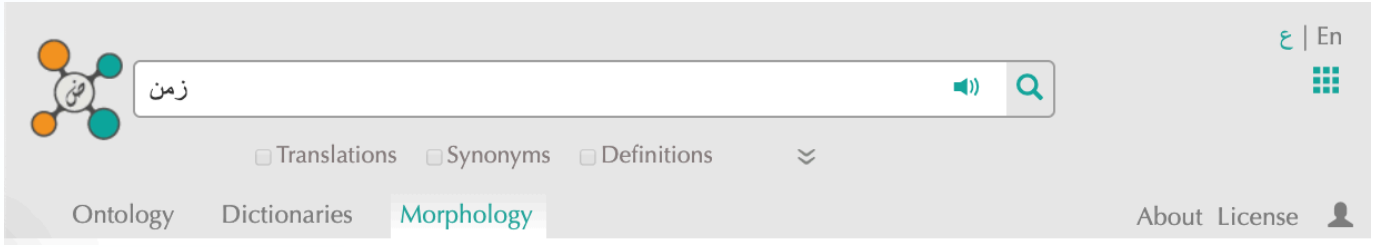
Search Taps

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



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]. (ف: ثلا. لازم). زَمِنْتُ، أَزْمَنُ، مص. زَمَنٌ
المعجم الثاني ©

زَمِنَ
زَمِنَ يَزْمَن، زَمَنًا وزُمْنةً وزَمانَةً، فهو زَمِن وزَمِن
المعجم الأول ©

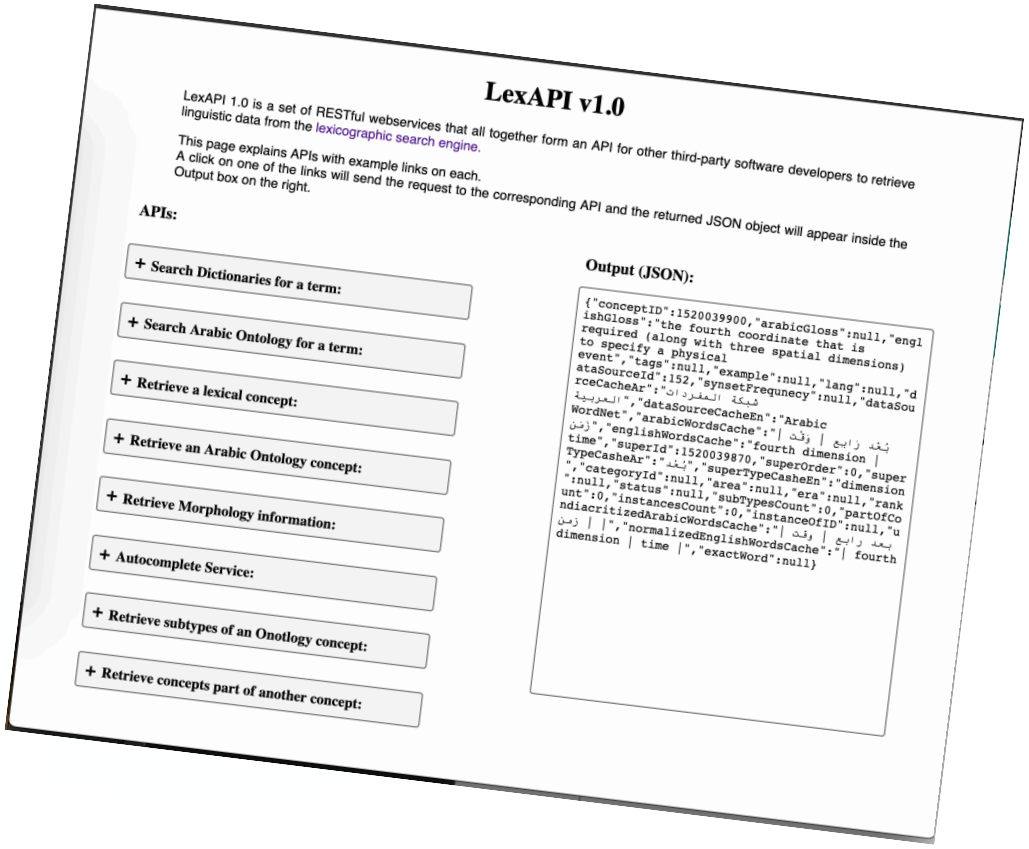
زَمَن
زَمَن [مفرد]: ج أَزْمُن (لغير المصدر) وأَزْمان (لغير المصدر)
المعجم الأول ©

زَمِن

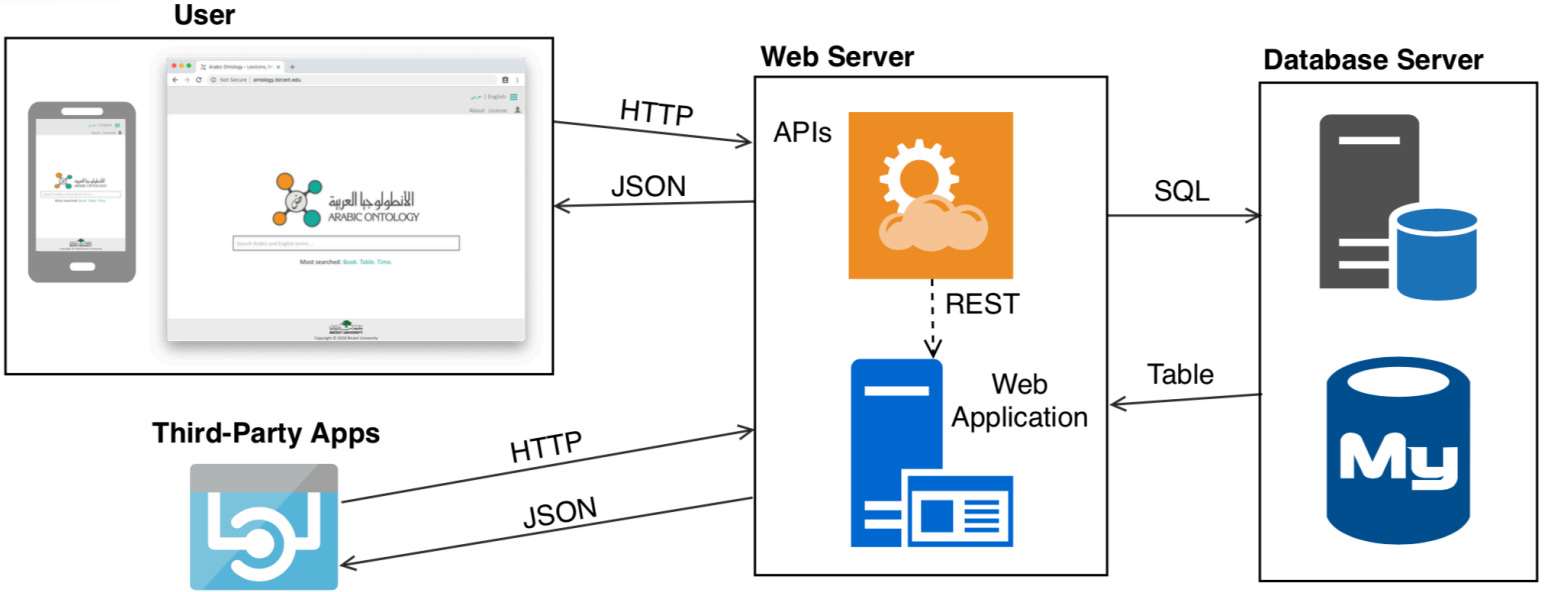
API Access

RESTful web services

Ask us for an API Key!



Search Engine Architecture



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/30400068

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:

Relations	Number of Mappings
SameAs	11400
SubClassOf/SuperClassOf	1050
PartOf/HasPart	100
InstanceOf/Type	770
Similar	125
Total	**13445**

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

Linking Lexicons with the Arabic Ontology

Following the W3C standards:

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

التسوية levelling | grading

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

Hydrology Lexicon ©

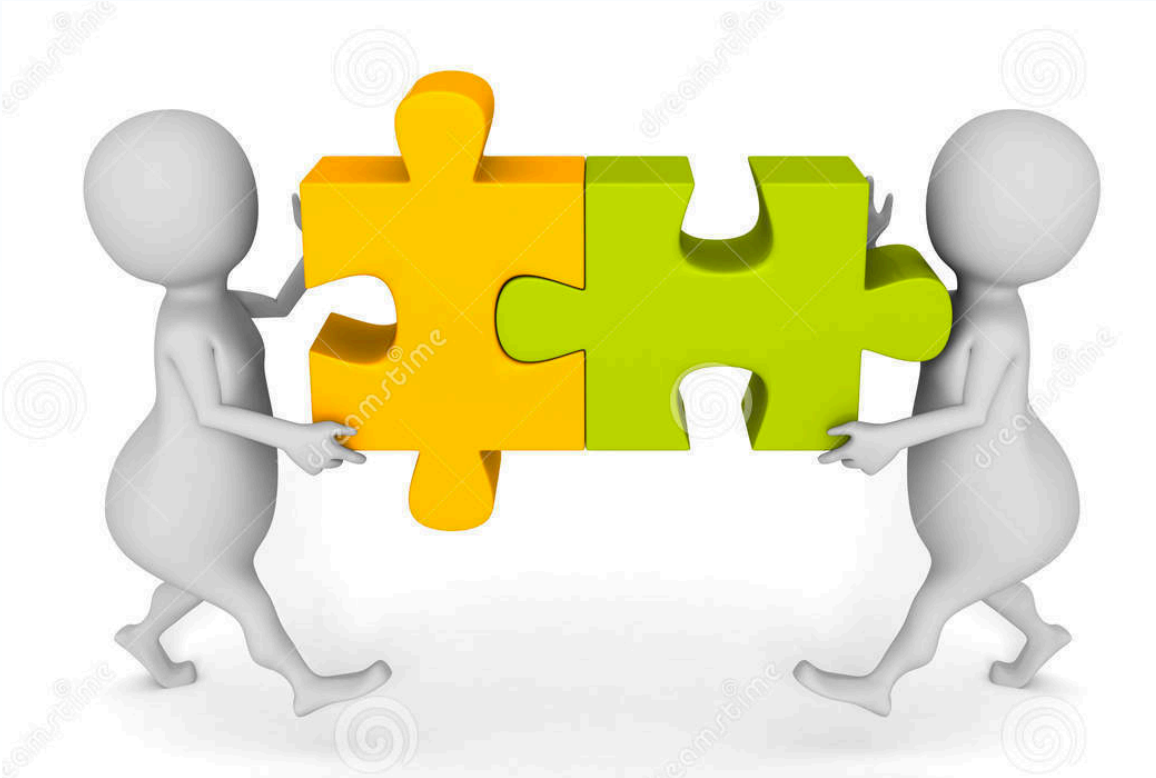
```
...
@prefix aot: <http://ontology.birzeit.edu/term/>.
@prefix aoc: <http://ontology.birzeit.edu/lexicalconcept/>.
@prefix aor: <http://ontology.birzeit.edu/lexicon/>.

<aoc:1623> a ontolex:LexicalConcept;
ontolex:isEvokedBy <aot:Lex-grading>;
ontolex:isEvokedBy <aot:Lex-levelling>;
ontolex:isEvokedBy <aot:Lex-تسوية>;
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-تسوية> a ontolex:LexicalEntry, ontolex:Word;
ontolex:canonicalForm [ontolex:writtenRep "تسوية"@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.



Connecting lexical resources

W3C Lemon RDF Model Standard

Lexicon Model for Ontologies: Community Report, 10 May 2016

Final Community Group Report 10 May 2016

Editors:

Philipp Cimiano (Cognitive Interaction Technology Excellence Center, Bielefeld University)
John P. McCrae (Insight Centre for Data Analytics, National University of Ireland, Galway)
Paul Buitelaar (Insight Centre for Data Analytics, National University of Ireland, Galway)

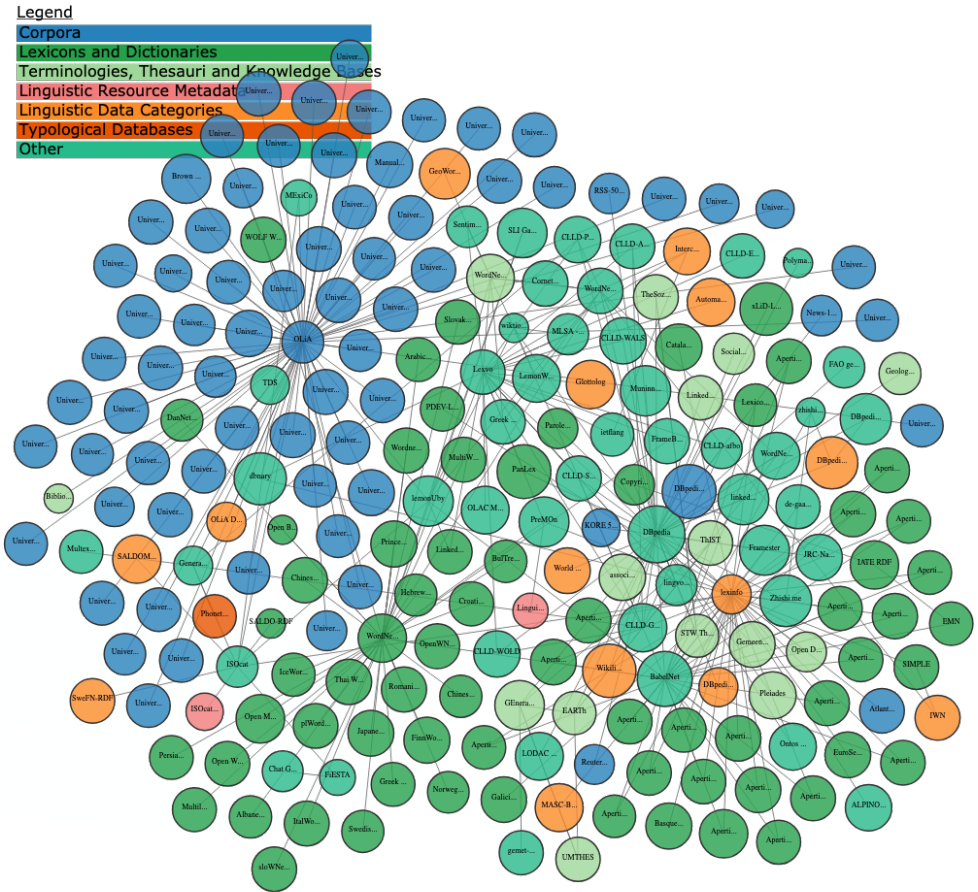
Abstract

This document describes the lexicon model for ontologies (*lemon*) as a main outcome of the work of the Ontology Lexicon (Ontolex) community group.

The Linguistic Linked Open Data Cloud

- A collaborative effort to develop a Linked Open Data (sub-)cloud of linguistic resources.

- Represent (lexical entries, concepts, synsets, and other) using Lemon RDF model, then interlinked.



Summary

Data Semantics + Lexical Semantics =

Building a **Linguistic Big Data Graph**

by linking:

Knowledge Graphs + Ontology + Dictionaries + Corpora +…+ LLODC

Currently we are working on Node embeddings (from semantic networks) for word sense disambiguation, Cultural heritage Knowledge Graphs, Chatbots …

Thank You

Mustafa Jarrar
mjarrar@birzeit.edu

References

Alhafi, D., Deik, D., & Jarrar, M. (2019): Usability Evaluation of Lexicographic e-Services. In Proceedings – 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications, Abu Dhabi (pp.1-7). IEEE. doi:10.1109/AICCSA47632.2019.9035226

Daher, J., & Jarrar, M. (2010). Towards a Methodology for Building Ontologies – Classify by Properties. In Proceedings – 3rd Palestinian International Conference on Computer and Information Technology (PICCIT), Palestine.

Elkateb, S., Black, W., Vossen, P., Farwell, D., Pease A., & Fellbaum, C. (2006). Arabic WordNet and the Challenges of Arabic. In Proceedings – Arabic NLP/MT Conference (pp. 665-670).

Emerson, G. (2020). What are the Goals of Distributional Semantics. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. ACL. (pp. 7436-7453).

Ercan, G., & Haziyev, F. (2019). Synset expansion on translation graph for automatic wordnet construction. Information Processing & Management, 56(1), 130-150.

Helou, M. A., Palmonari, M., & Jarrar, M. (2016). Effectiveness of Automatic Translations for Cross-Lingual Ontology Mapping. Journal of Artificial Intelligence Research, 55, 165-208. doi:10.1613/jair.4789

Helou, M. A., Palmonari, M., & Jarrar, M., Fellbaum, F. (2014). Towards Building Lexical Ontology via Cross-Language Matching. In Proceedings – 7th Conference on Global WordNet. Global WordNet Association. (pp. 346–354). EID: 2-s2.0-84859707947

Jarrar, M., & Meersman, R. (2002). Scalability and Knowledge Reusability in Ontology Modeling. In Proceedings – International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet (SSGRR 2002s). Scuola Superiore G Reiss Romoli. Rome, Italy.

Jarrar, M. (2005). Towards Methodological Principles for Ontology Engineering. Doctoral dissertation, Vrije Universiteit Brussel, Belgium.

Jarrar, M., (2006). Towards the Notion of Gloss, and the Adoption of Linguistic Resources in Formal Ontology Engineering. In Proceedings – 15th international conference on World Wide Web, (pp.497-503). ACM. doi: 10.1145/1135777.1135850

Jarrar, M. (2011): Building A Formal Arabic Ontology (Invited Paper). In Proceedings – Experts Meeting on Arabic Ontologies and Semantic Networks, Tunis. ALECSO, Arab League.

Jarrar, M., Habash, N., Alrimawi, F., Akra, D., & Zalmout, N. (2016). Curras: An Annotated Corpus for the Palestinian Arabic Dialect. Language Resources and Evaluation, 50(219), 1-31. doi:10.1007/S10579-016-9370-7

Jarrar, M., Zaraket, F., Asia, R., & Amayreh, H. (2018). Diacritic-based Matching of Arabic Words. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 18(2), 1-21. doi: 10.1145/3242177

Jarrar, M., & Amayreh, H. (2019). An Arabic-Multilingual Database with a Lexicographic Search Engine. In Proceedings – 24th International Conference on Applications of Natural Language to Information Systems (NLDB 2019). Lecture Notes in Computer Science (vol. 11608, pp. 234-246). Springer. Doi:10.1007/978-3-030-23281-8_19

Jarrar, M., Amayreh, H., & McCrae, J. (2019): Representing Arabic Lexicons in Lemon – a Preliminary Study. In Proceedings – 2nd Conference on Language, Data and Knowledge, Leipzig, Germany. CEUR-WS (vol. 2402, pp. 29-33).

Jarrar, M. (2021). The Arabic Ontology - An Arabic Wordnet with Ontologically Clean Content. Applied Ontology Journal, IOS Press.

Johnson, D. B. (1975). Finding all the elementary circuits of a directed graph. SIAM Journal on Computing, 4(1), 77-84.

Khodak, M., Risteski, A., Fellbaum, C., & Arora, S. (2017). Automated WordNet construction using word embeddings. In Proceedings of the 1st Workshop on Sense, Concept and Entity Representations and their Applications (pp. 12-23).

Lam, K., Tarouti, F., & Kalita J. (2014). Automatically constructing Wordnet synsets. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 106-111).

Miller, J., Beckwith, R., Fellbaum, C., Gross D., & Miller, K. (1990). Introduction to Wordnet: An on-line Lexical Database. International Journal of Lexicography, 3(4), 235-244.

Oliveira, H., & Gomes, P. (2014). ECO and Onto. PT: a flexible approach for creating a Portuguese wordnet automatically. Language resources and evaluation, 48(2), 373-393.

Tarouti, F., & Kalita, J. (2016). Enhancing automatic wordnet construction using word embeddings. In Proceedings – Workshop on Multilingual and Cross-lingual Methods in NLP (pp. 30-34).

Tiziano, F., & Navigli. R. (2012). The CQC algorithm: Cycling in graphs to semantically enrich and enhance a bilingual dictionary. Journal of Artificial Intelligence Research, 43, 135-171.

Torregrosa, D., Mihael, A., Ahmadi, S., & McCrae, J. (2019). TIAD 2019 Shared Task: Leveraging knowledge graphs with neural machine translation for automatic multilingual dictionary generation. Translation Inference Across Dictionaries.

Villegas, M., Melero, M., Gracia J., & Bel, N. (2016). Leveraging RDF graphs for crossing multiple bilingual dictionaries. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) (pp. 868-876).

Wu, H., & Zhow M. (2003). Optimizing synonym extraction using monolingual and bilingual resources. In Proceedings of the second international workshop on Paraphrasing (pp. 72-79).