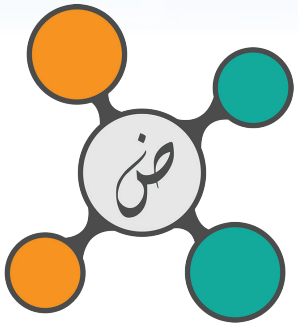
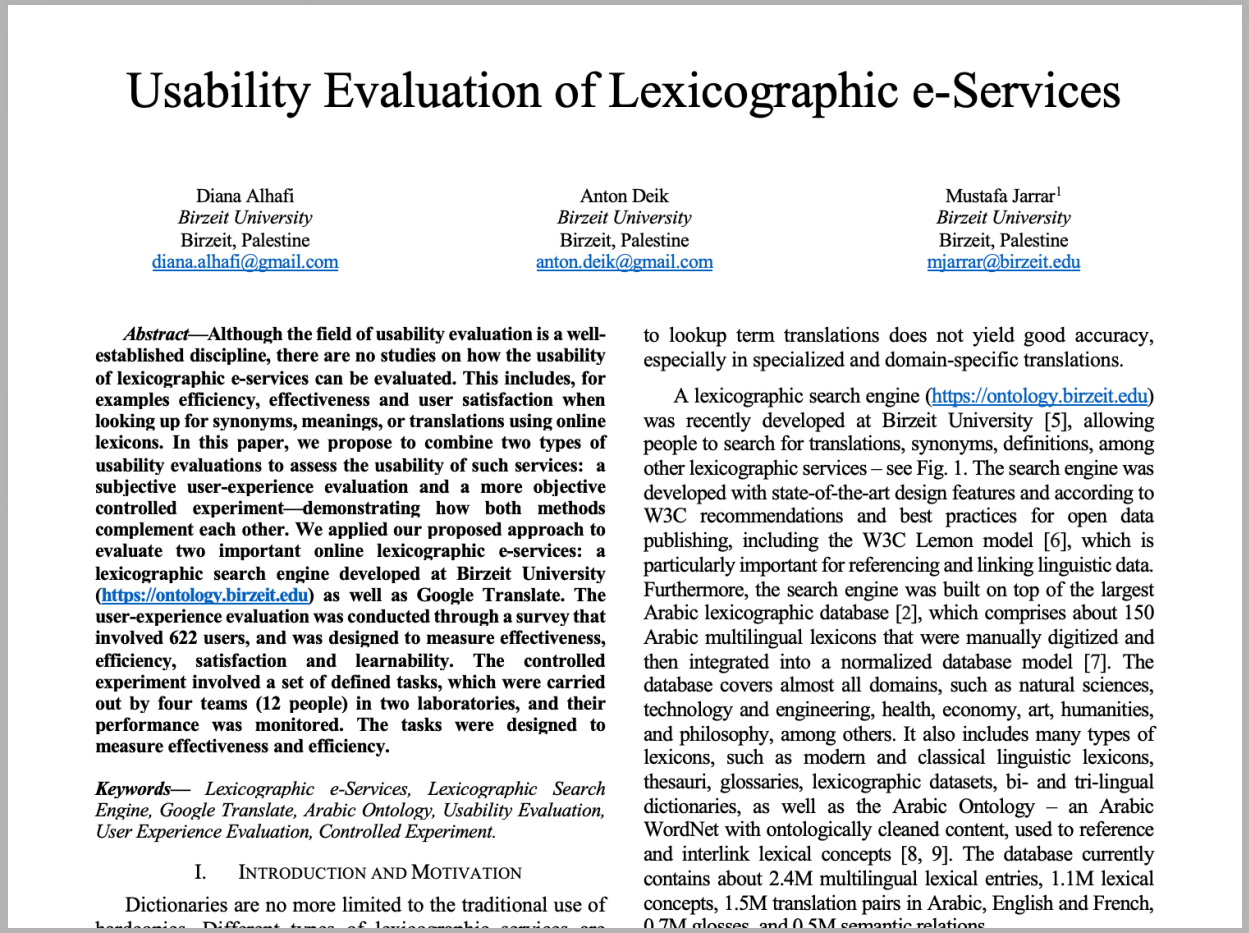
AICCSA 2019, Abu Dhabi
4, November 2019



Usability Evaluation of Lexicographic e-Services

Diana Alhafi **Anton Deik** **Mustafa Jarrar**

Birzeit University
Palestine



Usability Evaluation of Lexicographic e-Services

Diana Alhafi
Birzeit University
Birzeit, Palestine
diana.alhafi@gmail.com

Anton Deik
Birzeit University
Birzeit, Palestine
anton.deik@gmail.com

Mustafa Jarrar[1]
Birzeit University
Birzeit, Palestine
mjarrar@birzeit.edu

***Abstract*—Although the field of usability evaluation is a well-established discipline, there are no studies on how the usability of lexicographic e-services can be evaluated. This includes, for examples efficiency, effectiveness and user satisfaction when looking up for synonyms, meanings, or translations using online lexicons. In this paper, we propose to combine two types of usability evaluations to assess the usability of such services: a subjective user-experience evaluation and a more objective controlled experiment—demonstrating how both methods complement each other. We applied our proposed approach to evaluate two important online lexicographic e-services: a lexicographic search engine developed at Birzeit University (https://ontology.birzeit.edu) as well as Google Translate. The user-experience evaluation was conducted through a survey that involved 622 users, and was designed to measure effectiveness, efficiency, satisfaction and learnability. The controlled experiment involved a set of defined tasks, which were carried out by four teams (12 people) in two laboratories, and their performance was monitored. The tasks were designed to measure effectiveness and efficiency.**

Keywords*— *Lexicographic e-Services, Lexicographic Search Engine, Google Translate, Arabic Ontology, Usability Evaluation, User Experience Evaluation, Controlled Experiment.

I. Introduction and Motivation

Dictionaries are no more limited to the traditional use of [illegible] to lookup term translations does not yield good accuracy, especially in specialized and domain-specific translations.

A lexicographic search engine (https://ontology.birzeit.edu) was recently developed at Birzeit University [5], allowing people to search for translations, synonyms, definitions, among other lexicographic services – see Fig. 1. The search engine was developed with state-of-the-art design features and according to W3C recommendations and best practices for open data publishing, including the W3C Lemon model [6], which is particularly important for referencing and linking linguistic data. Furthermore, the search engine was built on top of the largest Arabic lexicographic database [2], which comprises about 150 Arabic multilingual lexicons that were manually digitized and then integrated into a normalized database model [7]. The database covers almost all domains, such as natural sciences, technology and engineering, health, economy, art, humanities, and philosophy, among others. It also includes many types of lexicons, such as modern and classical linguistic lexicons, thesauri, glossaries, lexicographic datasets, bi- and tri-lingual dictionaries, as well as the Arabic Ontology – an Arabic WordNet with ontologically cleaned content, used to reference and interlink lexical concepts [8, 9]. The database currently contains about 2.4M multilingual lexical entries, 1.1M lexical concepts, 1.5M translation pairs in Arabic, English and French, 0.7M glosses, and 0.5M semantic relations

Diana Alhafi, Anton Deik, Mustafa Jarrar: Usability Evaluation of Lexicographic e-Services. Proceedings of the 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA). Pages(1--7), IEEE. Abu Dhabi, UAE. 2019.

http://www.jarrar.info/publications/ADJ19.pdf

Motivation

- How to measure efficiency, effectiveness and user satisfaction when looking up for synonyms, meanings, or translations using online lexicons?

- Although the field of usability evaluation is a well-established discipline, there are no studies on how the usability of lexicographic e-services can be evaluated!

Our approach

- Combine two types of usability evaluations to assess the usability of e-lexicographic services =

subjective user-experience + **objective** controlled experiment

- We show how both methods complement each other.

- Experiments: evaluate two online lexicographic services: a lexicographic search engine developed at Birzeit University as well as Google Translate.

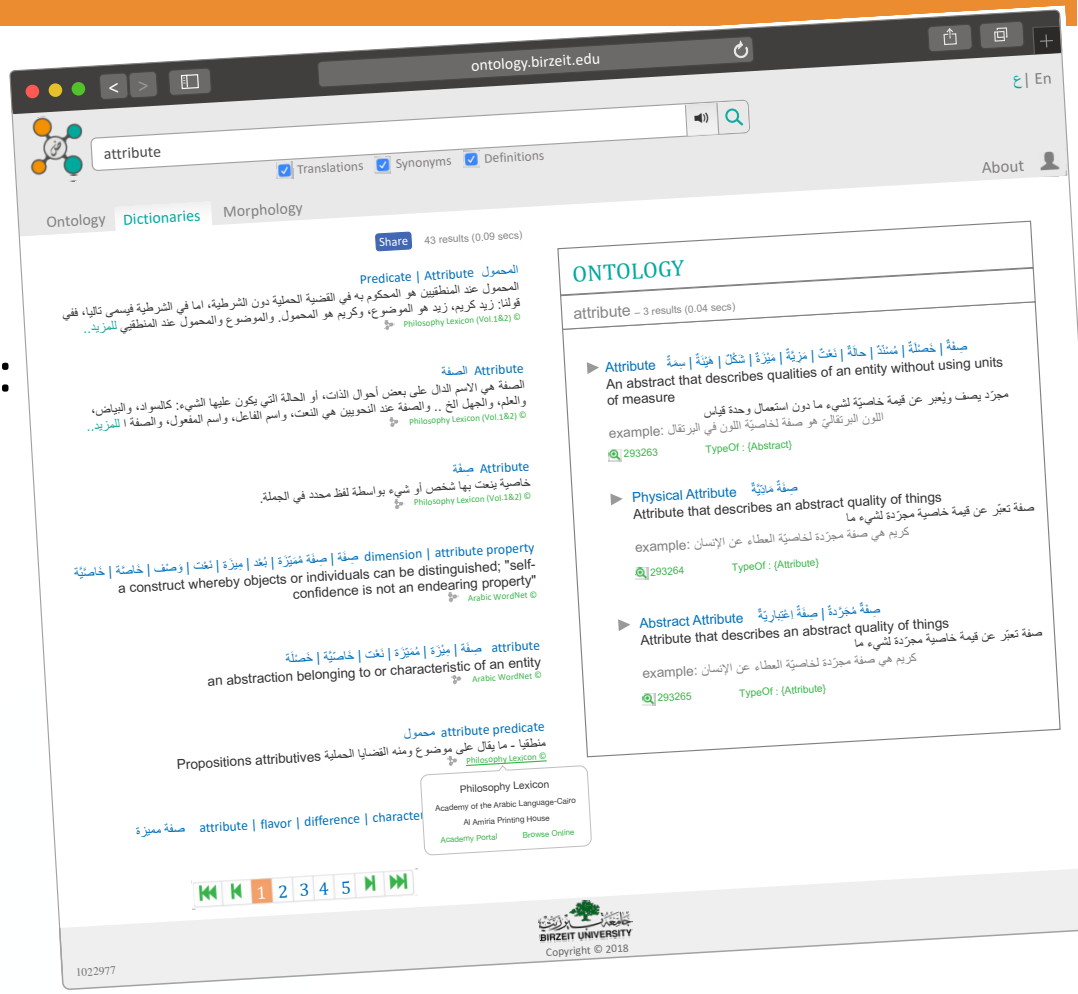
Outline

- Overview of the Lexical Search Engine
- Usability Evaluation of Lexicographic e-Services.

Overview of the Lexicographic Search Engine

The Lexicographic Database

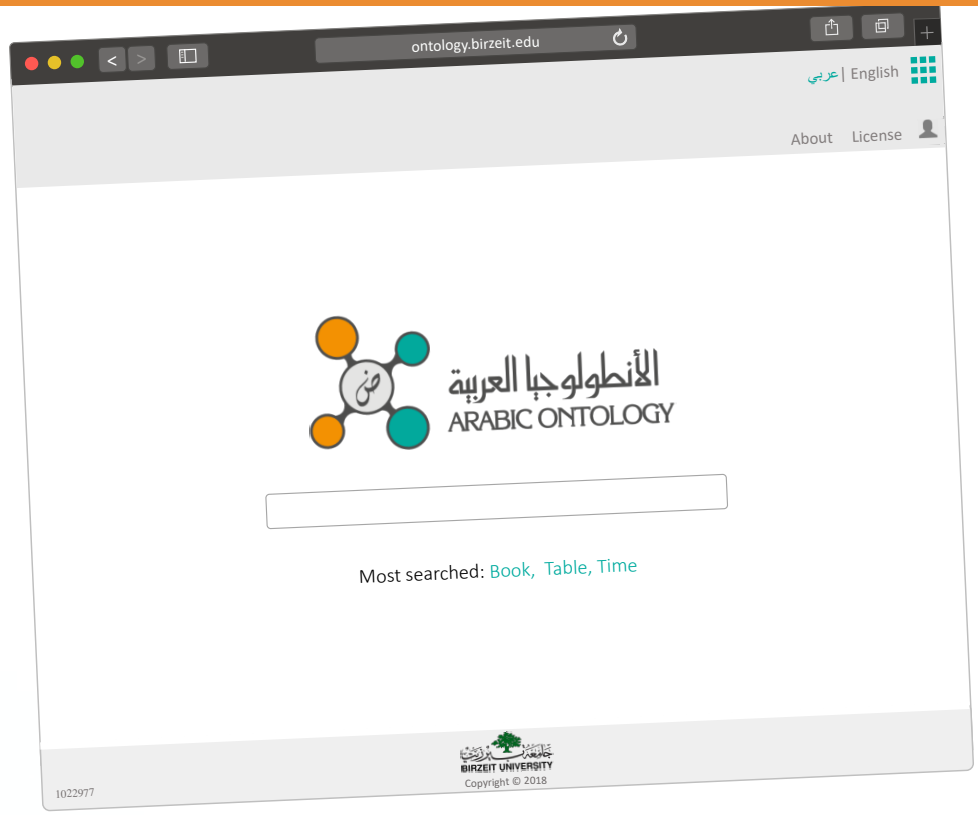
- **The largest lexicographic Arabic database**
- **Contains most lexicon types**: glossaries, thesauri, bi/trilingual dictionaries, morph datasets, **Arabic Ontology**, and more.
- **Covers most domains:** science, technology, law, business, art, philosophy, ...



https://ontology.birzeit.edu

Lexicographic Search Engine

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



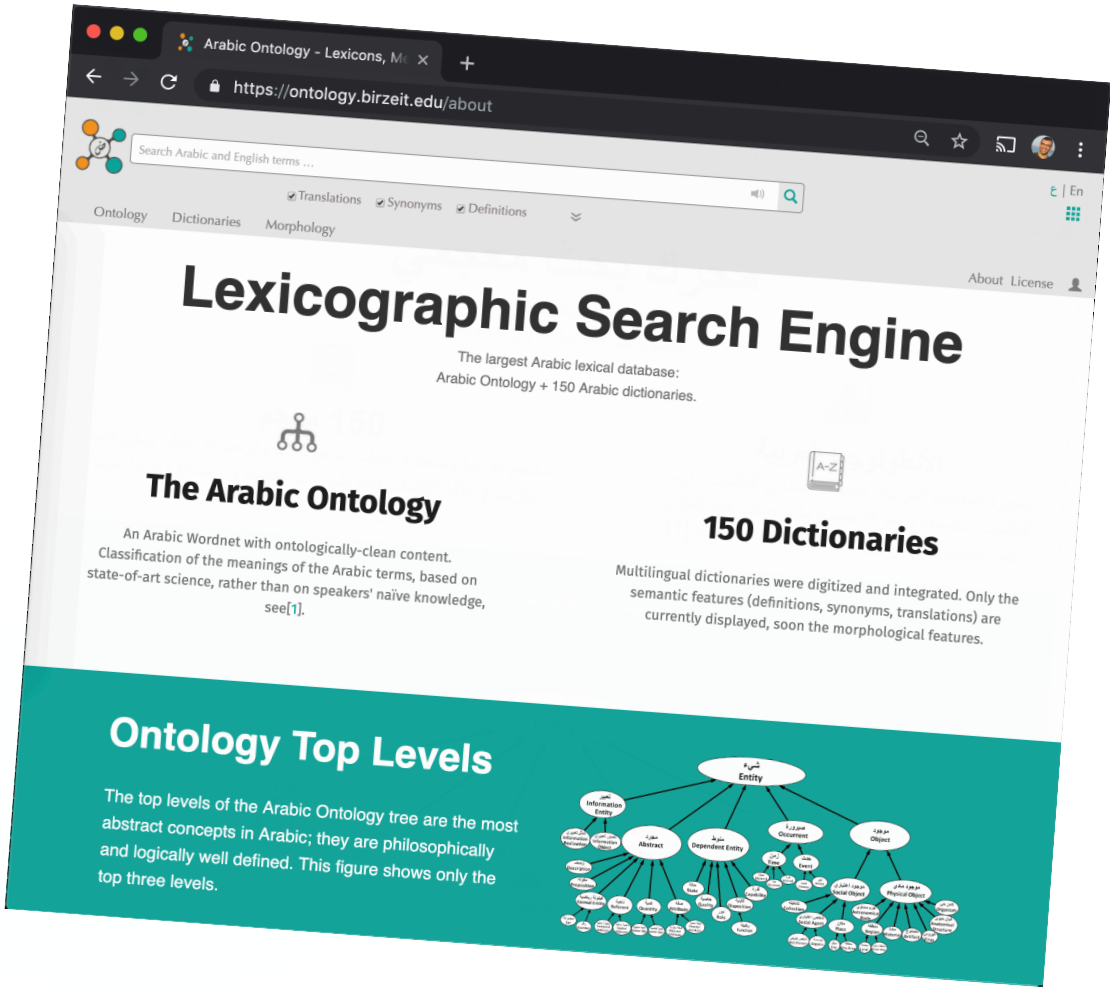
https://ontology.birzeit.edu

Reference:

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

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

Some Statistics

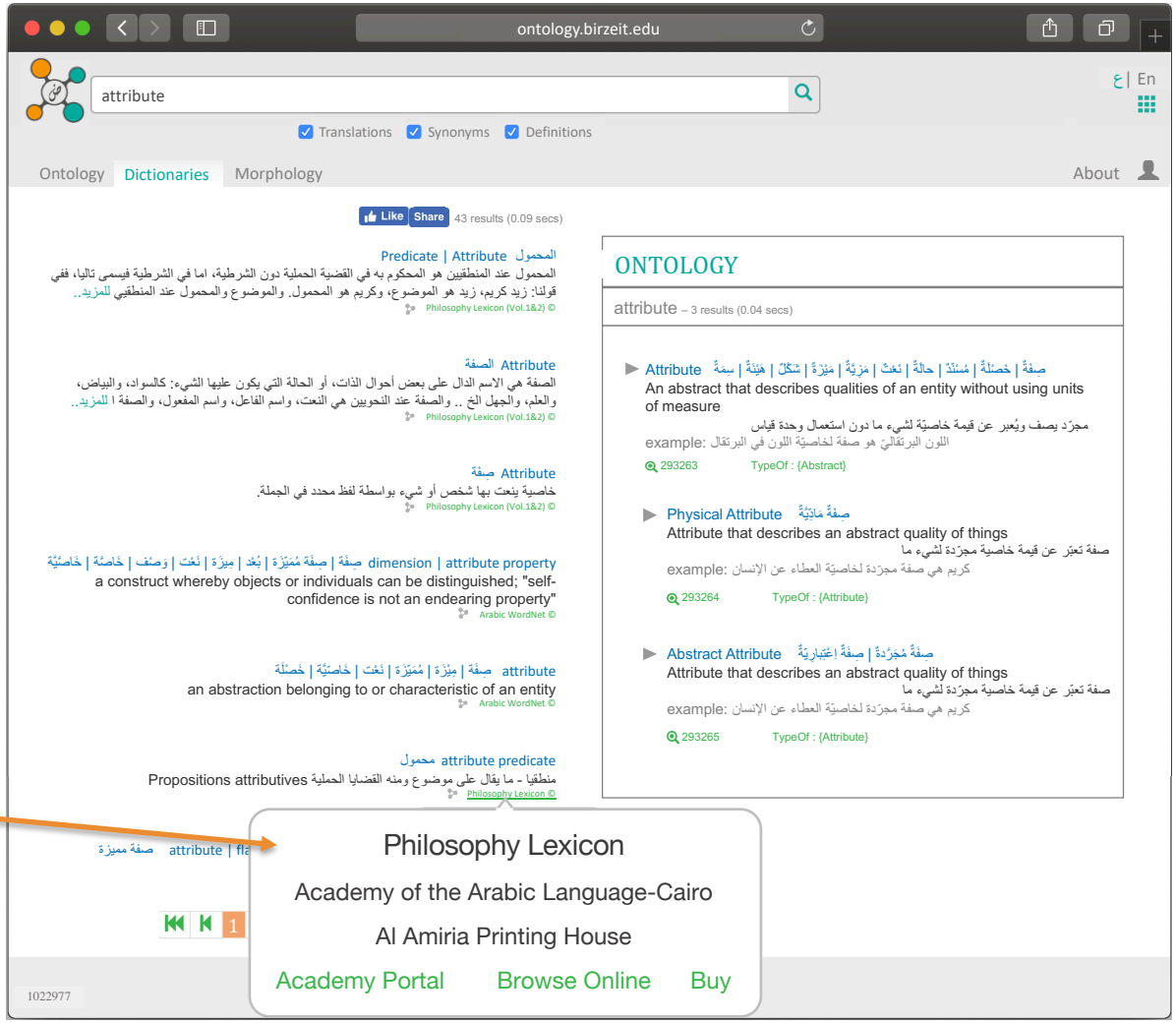
Currently!

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

For more, see [3]

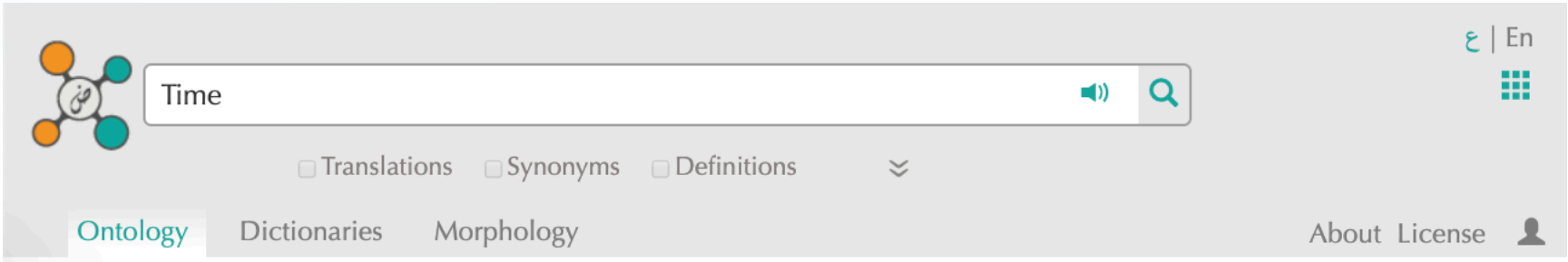
Obtaining Copyrights

- 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

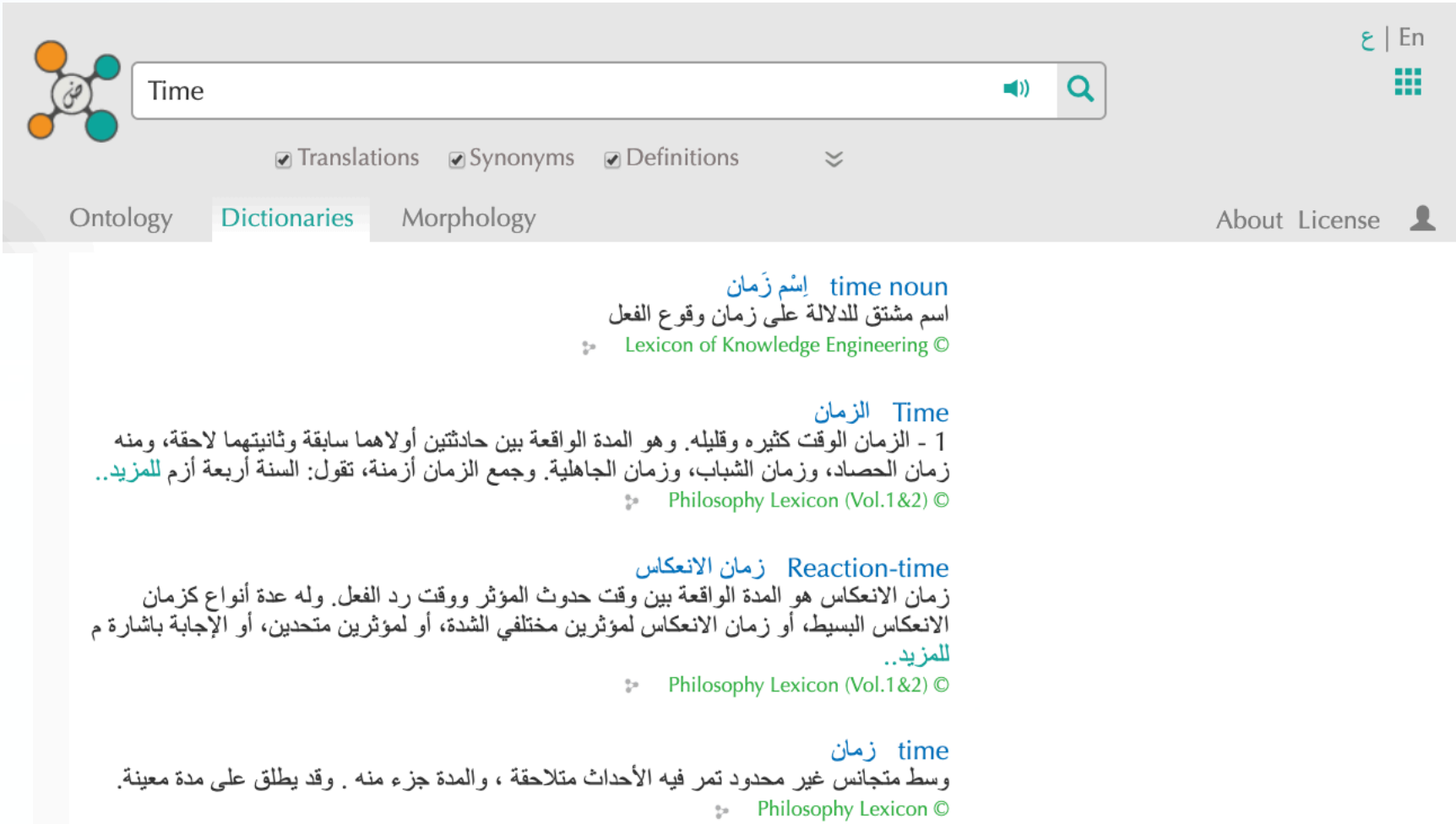


time مُدَّةٌ | زَمَنٌ | فَتْرَةٌ
An occurrent representing a region in the timeline, realized by its starting and ending points, its length represents the temporal dimension of events or objects
مجرّد يمثّل جزءاً من خطّ الزّمن المدرك، تدرك ذاته بنقطة بداية ونقطة نهاية طولها يمثّل زمن أحداث أو موجودات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

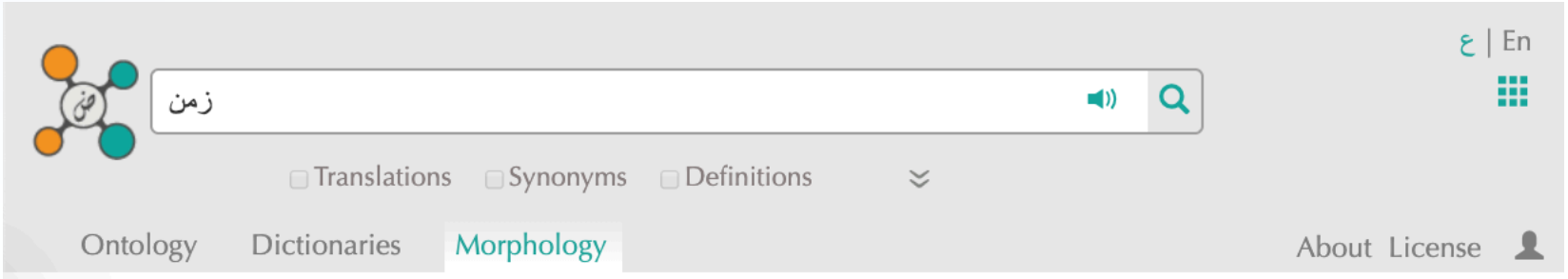
Search Taps

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 ©

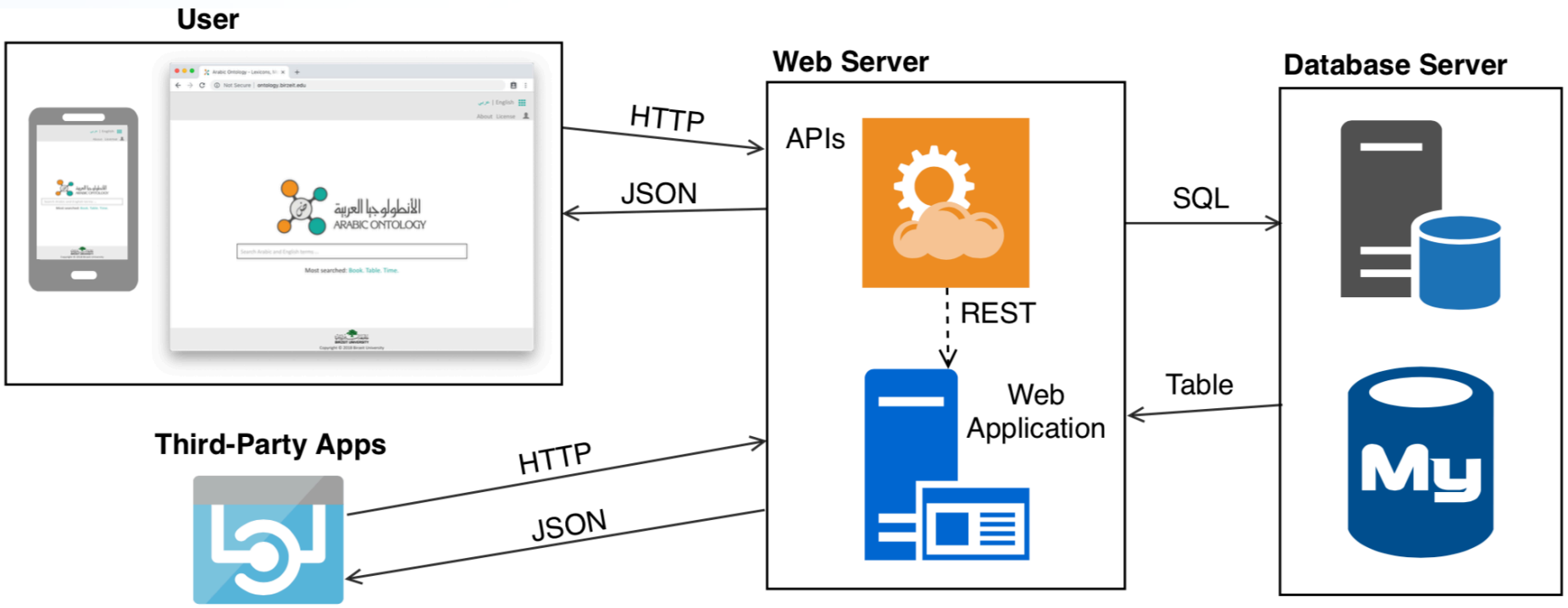
زَمِنَ
[ز م ن]. (ف: ثلا. لازم). زَمِنْتُ، أَزْمَنُ، مص. زَمَنٌ
المعجم الثاني ©

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

levelling | grading التسوية

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.

Conformance with W3C Standards

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

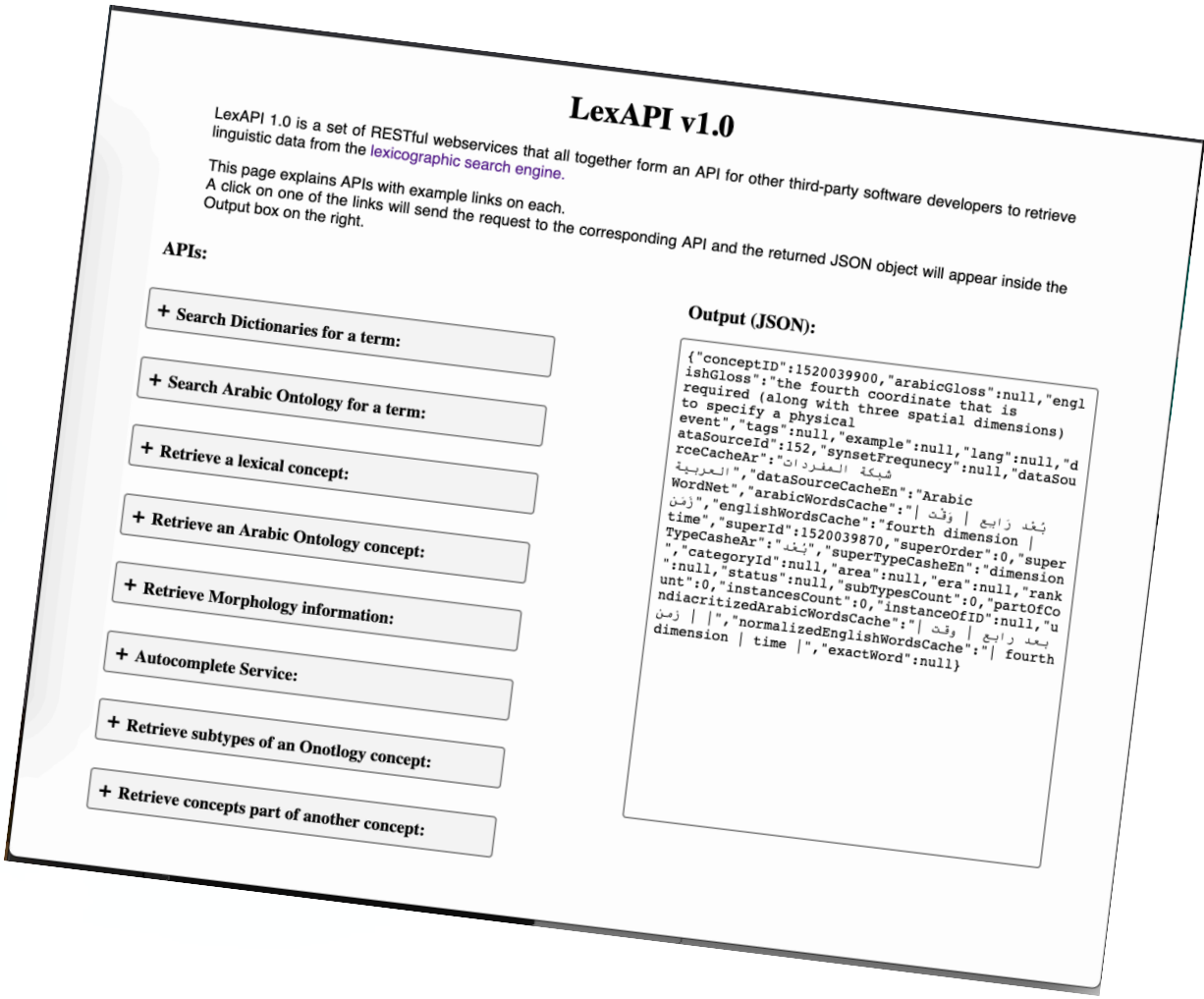
URLs Schema:

- 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/30400068<u>2</u>
- 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,"englishGloss":"the fourth coordinate that is required (along with three spatial dimensions) to specify a physical event","tags":null,"example":null,"lang":null,"dataSourceId":152,"synsetFrequnecy":null,"dataSourceCacheAr":"شبكة المفردات العربية","dataSourceCacheEn":"Arabic WordNet","arabicWordsCache":"بُعد رابع | وقت | زمن","englishWordsCache":"fourth dimension | time","superId":1520039870,"superOrder":0,"superTypeCasheAr":"بُعد","superTypeCasheEn":"dimension","categoryId":null,"area":null,"era":null,"rank":null,"status":null,"subTypesCount":0,"partOfCount":0,"instancesCount":0,"instanceOfID":null,"undiacritizedArabicWordsCache":"بعد رابع | وقت | زمن","normalizedEnglishWordsCache":"| fourth dimension | time |","exactWord":null}

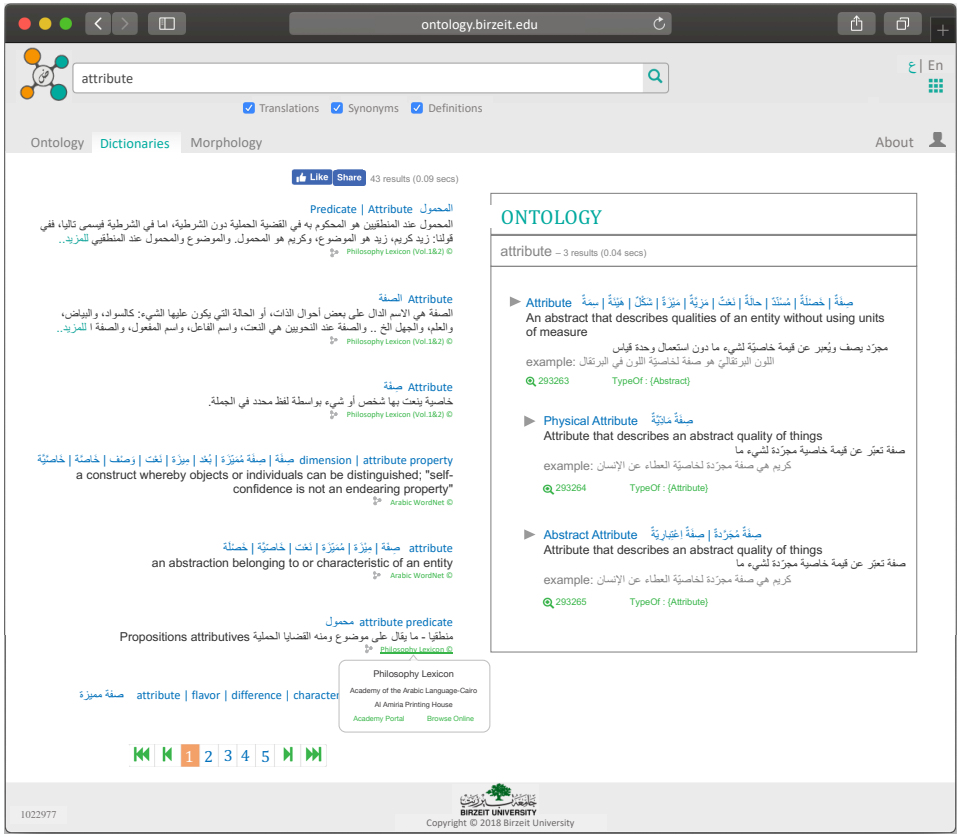
Ranking Metrics

We developed three strategies:

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

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

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



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

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

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

Usability Evaluation of Lexicographic e-Services

1) Subjective user-experience
2) Objective controlled experiment

Subjective User-Experience

Online Survey

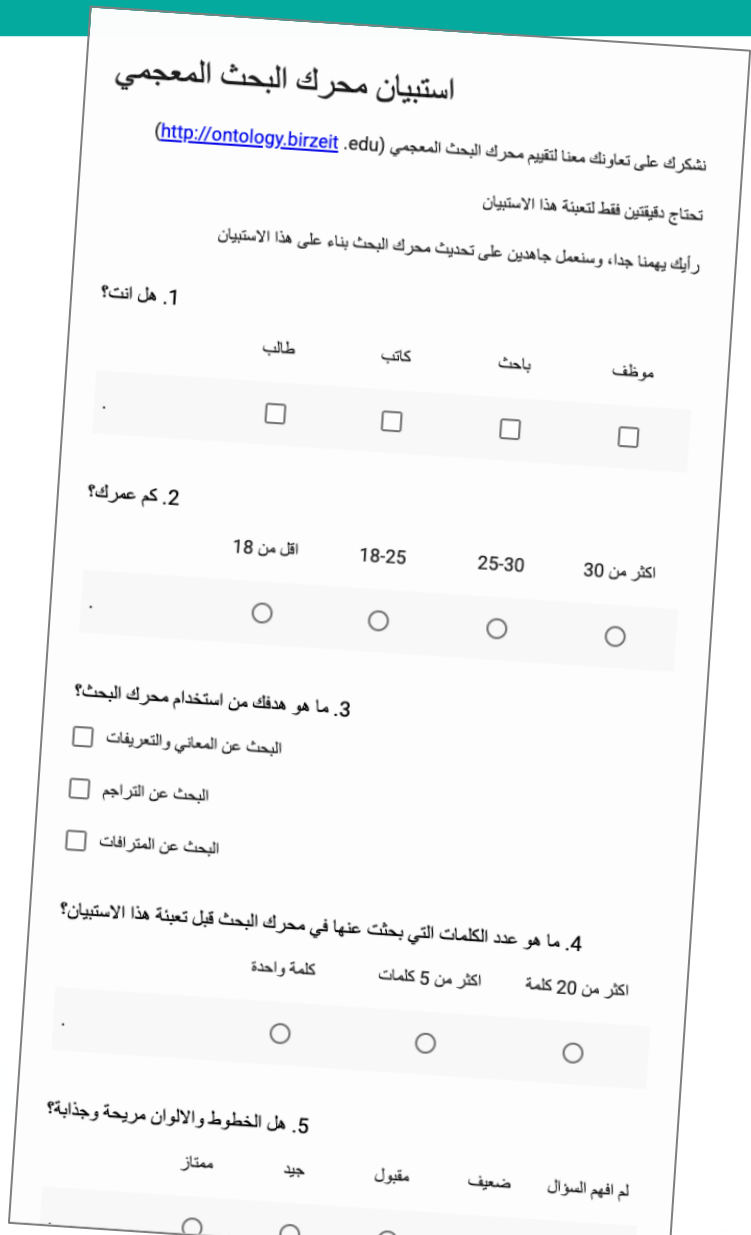
https://ontology.birzeit.edu/s

Survey Design:

- 16 Questions
- Q1-Q4 collected general information
- Q5-Q16 are the core of the survey, targeting Effectiveness, Efficiency, Satisfaction, Learnability

Data Collection:

- 1000 questionnaires were distributed over a period of 2 months.
- 622 valid responses acquired



استبيان محرك البحث المعجمي

نشكرك على تعاونك معنا لتقييم محرك البحث المعجمي (http://ontology.birzeit .edu)

تحتاج دقيقتين فقط لتعبئة هذا الاستبيان

رأيك يهمنا جدا، وسنعمل جاهدين على تحديث محرك البحث بناء على هذا الاستبيان

1. هل انت؟

موظف | باحث | كاتب | طالب

2. كم عمرك؟

اكثر من 30 | 25-30 | 18-25 | اقل من 18

3. ما هو هدفك من استخدام محرك البحث؟

- البحث عن المعاني والتعريفات
- البحث عن التراجم
- البحث عن المترادفات

4. ما هو عدد الكلمات التي بحثت عنها في محرك البحث قبل تعبئة هذا الاستبيان؟

اكثر من 20 كلمة | اكثر من 5 كلمات | كلمة واحدة

5. هل الخطوط والالوان مريحة وجذابة؟

لم افهم السؤال | ضعيف | مقبول | جيد | ممتاز

Summary of Survey Results

Respondents' Backgrounds (Q1-Q2)

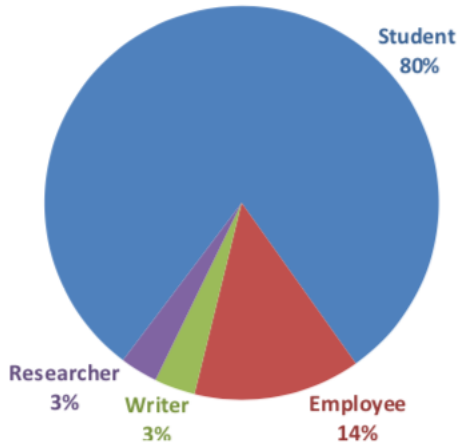


Fig. 2. Respondents' professions

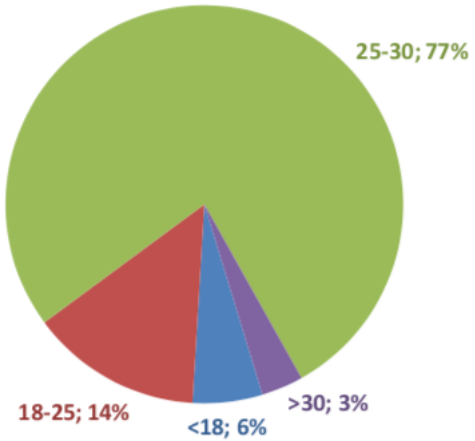


Fig. 3. Respondents' age groups

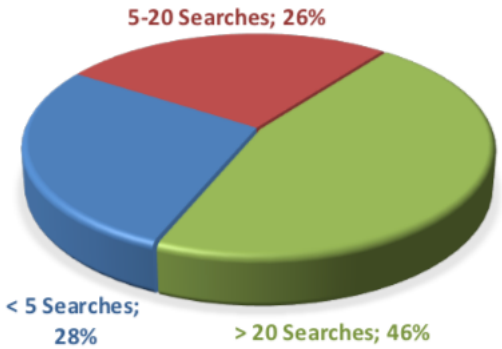


Fig. 5. Frequency of use

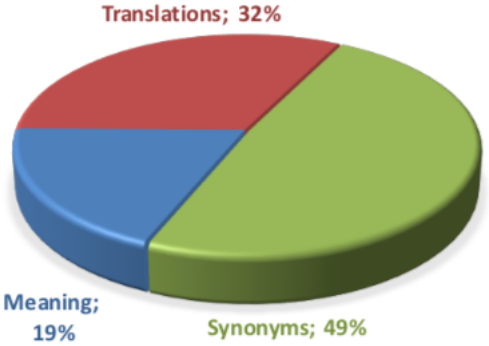
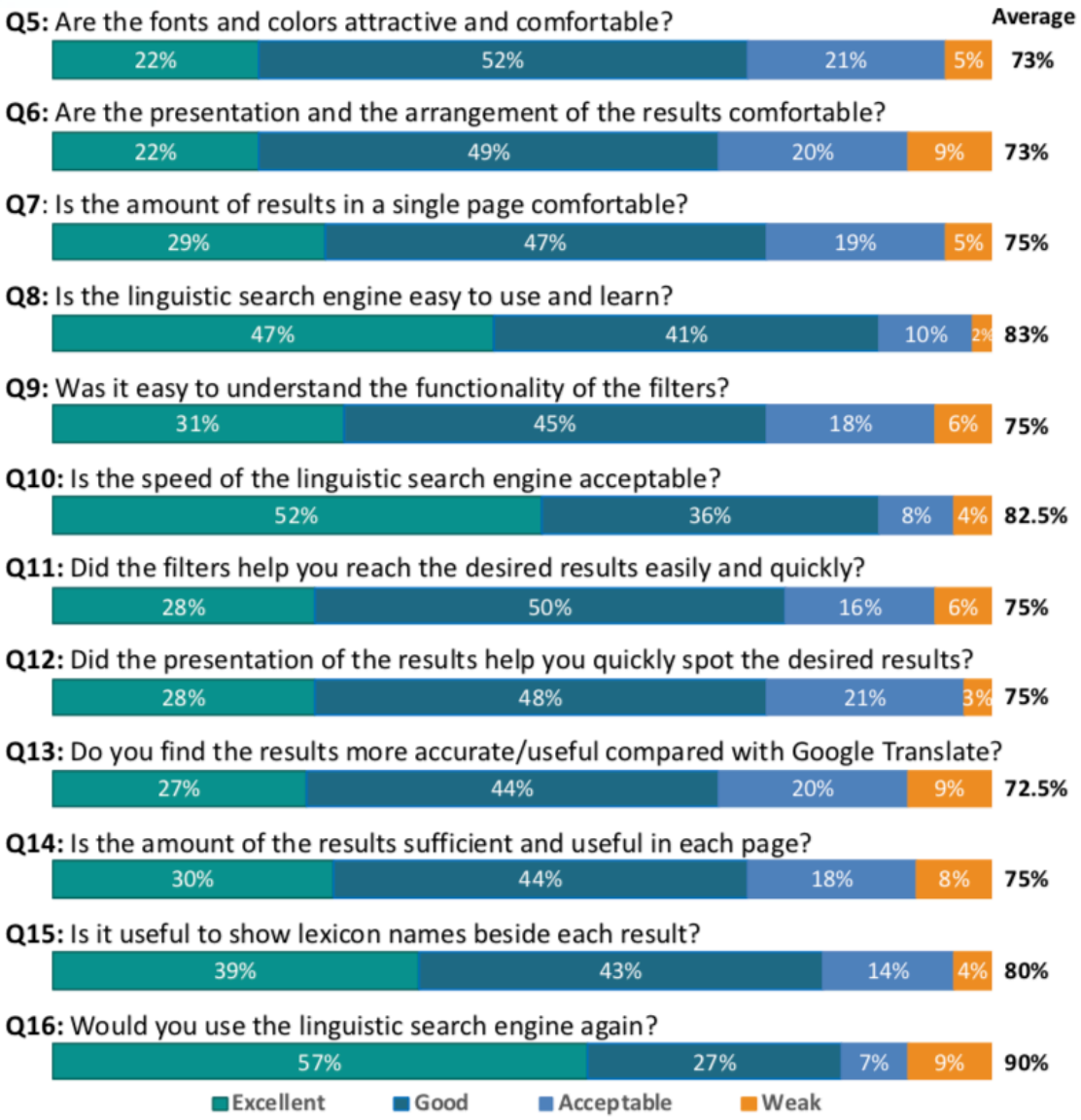


Fig. 4. Purpose of use

Summary of Survey Results



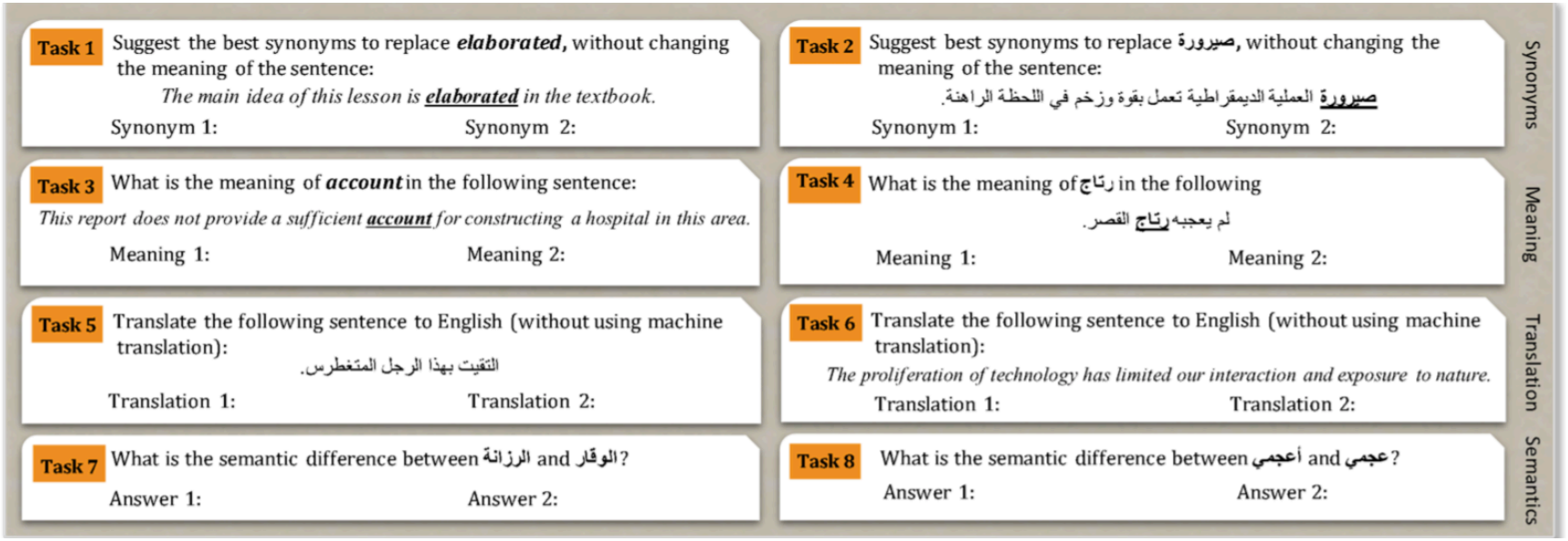
Summary of Survey Results

Subjective user-experience

Criterion	Satisfaction	Learnability	Efficiency	Effectiveness	Average
Questions	Q5-Q7	Q8-Q9	Q10-Q12	Q13-Q16	Q5-Q16
Average	73%	83%	75%	80%	**77.8%**

Control Experiment

- Eight tasks were carried out in a controlled environment, in a lab environment
- Four types of tasks: (*i*) synonyms, (*ii*) meanings, (*iii*) translation, and (*iv*) semantic differences between terms.
- Measure the efficiency and effectiveness of both Google Translate and the lexicographic search engine.
- Two groups (A and B), each consisted of 3 participants (12 in total).
- 10 minutes tutorial to try both tools.



Summary of the Control Experiment Results

Table II. Task Time (in seconds) by participant by task

Participants			Tasks								
			T1	T2	T3	T4	T5	T6	T7	T8	Mean time
Group A: LSE	BZU	P1	58	27	40	20	45	80	116	95	**60**
		P2	43	17	62	30	53	50	116	118	**61**
		P3	30	24	37	51	50	113	115	96	**65**
	PTUK	P4	110	39	60	31	25	195	123	85	**84**
		P5	116	48	77	43	27	190	120	77	**87**
		P6	50	40	67	43	37	160	150	102	**81**
Meantime			**68**	**33**	**57**	**36**	**40**	**131**	**123**	**96**	**73**
Group B: GT	BZU	P7	90	25	50	30	61	85	60	70	**59**
		P8	120	14	120	61	50	128	60	30	**73**
		P9	110	30	80	60	71	220	60	30	**83**
	PTUK	P10	27	26	40	15	18	75	11	12	**28**
		P11	37	27	43	11	22	85	17	30	**34**
		P12	14	20	44	30	30	90	30	11	**34**
Meantime			**66**	**24**	**63**	**35**	**42**	**114**	**40**	**31**	**52**

Table III. Task Score by participant by task

Participants			Tasks								
			T1	T2	T3	T4	T5	T6	T7	T8	Avg
Group A: LSE	BZU	P1	8	10	10	10	10	8	10	10	**9.50**
		P2	6	10	4	8	10	6	6	8	**7.25**
		P3	10	10	10	10	6	8	6	6	**8.25**
	PTUK	P4	10	10	0	10	8	6	8	8	**7.50**
		P5	8	10	0	6	4	10	6	10	**6.75**
		P6	8	10	8	8	8	8	6	8	**8.00**
Average			**8.33**	**10**	**5.66**	**8.66**	**7.66**	**7.66**	**7.00**	**8.33**	**7.91**
Group B: GT	BZU	P7	0	0	4	4	10	8	0	0	**3.25**
		P8	8	0	4	0	6	10	0	0	**3.50**
		P9	8	0	10	0	8	8	0	0	**4.25**
	PTUK	P10	8	0	2	0	6	8	0	0	**3.00**
		P11	4	0	4	4	6	4	0	0	**2.75**
		P12	10	0	4	0	4	8	0	0	**3.25**
Average			**6.33**	**6.67**	**4.66**	**4**	**6.66**	**7.66**	**0**	**0**	**4.5**

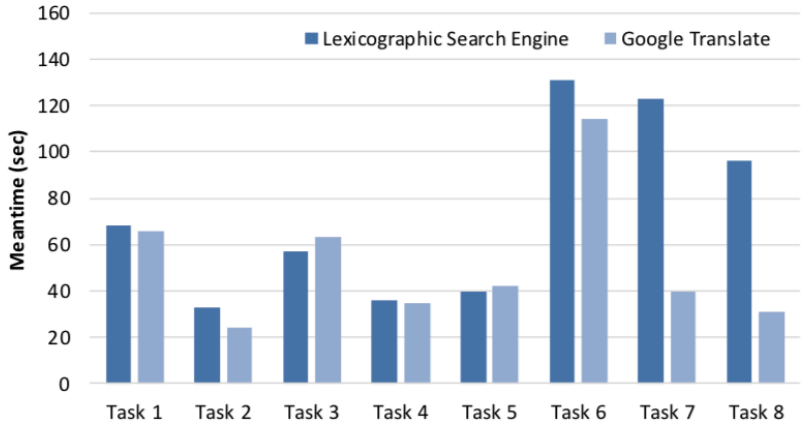


Fig. 8. Meantime, in seconds, for each task (i.e., efficiency)

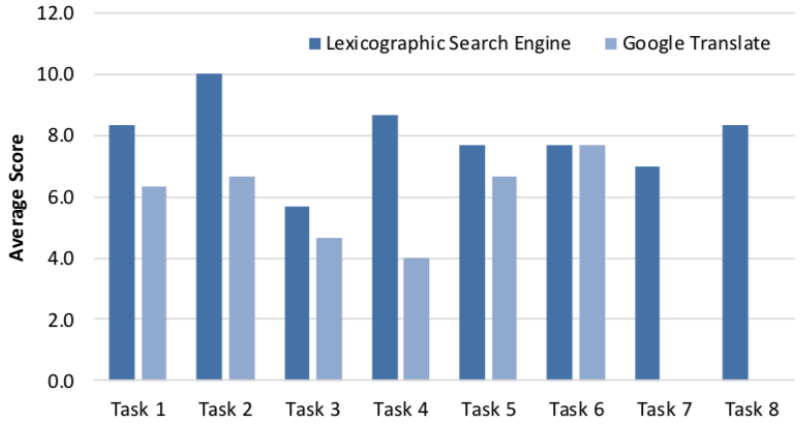


Fig. 9. Average task scores for both tools (i.e., effectiveness)

Comparing both Approaches

subjective user-experience Vs. **objective** controlled experiment

Criterion	Satisfaction	Learnability	Efficiency	Effectiveness	Average
Questions	Q5-Q7	Q8-Q9	Q10-Q12	Q13-Q16	Q5-Q16
Average	73%	83%	75%	80%	**77.8%**

Participants			Tasks								
			T1	T2	T3	T4	T5	T6	T7	T8	Mean time
Group A: LSE	BZU	P1	58	27	40	20	45	80	116	95	**60**
		P2	43	17	62	30	53	50	116	118	**61**
		P3	30	24	37	51	50	113	115	96	**65**
	PTUK	P4	110	39	60	31	25	195	123	85	**84**
		P5	116	48	77	43	27	190	120	77	**87**
		P6	50	40	67	43	37	160	150	102	**81**
Meantime			**68**	**33**	**57**	**36**	**40**	**131**	**123**	**96**	**73**

Participants			Tasks								
			T1	T2	T3	T4	T5	T6	T7	T8	Avg
Group A: LSE	BZU	P1	8	10	10	10	10	8	10	10	**9.50**
		P2	6	10	4	8	10	6	6	8	**7.25**
		P3	10	10	10	10	6	8	6	6	**8.25**
	PTUK	P4	10	10	0	10	8	6	8	8	**7.50**
		P5	8	10	0	6	4	10	6	10	**6.75**
		P6	8	10	8	8	8	8	6	8	**8.00**
Average			**8.33**	**10**	**5.66**	**8.66**	**7.66**	**7.66**	**7.00**	**8.33**	**7.91**

References

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

[2] 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.

[3] H. Amayreh, M. Dwaikat, and M. Jarrar, "**Lexicon Digitization - A Framework for Structuring, Normalizing and Cleaning Lexical Entries**", Technical Report, Birzeit University, 2019.

[4] M. Jarrar, "**Building a Formal Arabic Ontology**," In Proceedings of the Experts Meeting on Arabic Ontologies and Semantic Networks, ALECSO, Arab League, 2011.

[5] M. Jarrar, "**The Arabic Ontology - An Arabic Wordnet with Ontologically Clean Content**," Applied Ontology Journal, 2019 (Forthcoming).

[6] Diana Alhafi, Anton Deik, Mustafa Jarrar: **Usability Evaluation of Lexicographic e-Services**. Proceedings of the 2012 IEEE/ACS 16th International Conference on Computer Systems and Applications. IEEE Computer Society. Abu Dhabi, UAE.