



# Advanced and Collaborative AI Research on Arabic

**Mustafa Jarrar**  
Birzeit University  
Palestine

# Artificial Intelligence

Opportunity  
to invest in  
**Arabic**  
as **industry**  
and  
**sustainable**  
**R&D**

The 4<sup>th</sup>  
Revolution

Opportunity  
for  
developing  
countries to  
compete and  
impact



يرجى ضبط المنبه لـ ٥:١٥  
تشغيل الموسيقى لـ ٦:١٥

كيف سيكون الطقس اليوم لـ ١٥:٠١

أرني المواعيد للأسبوع المقبل لـ ٦:٠١  
انتقل إلى المنزل لـ ٦:٠١

# Huge Market

400  
Million  
users  
(as a first  
language)

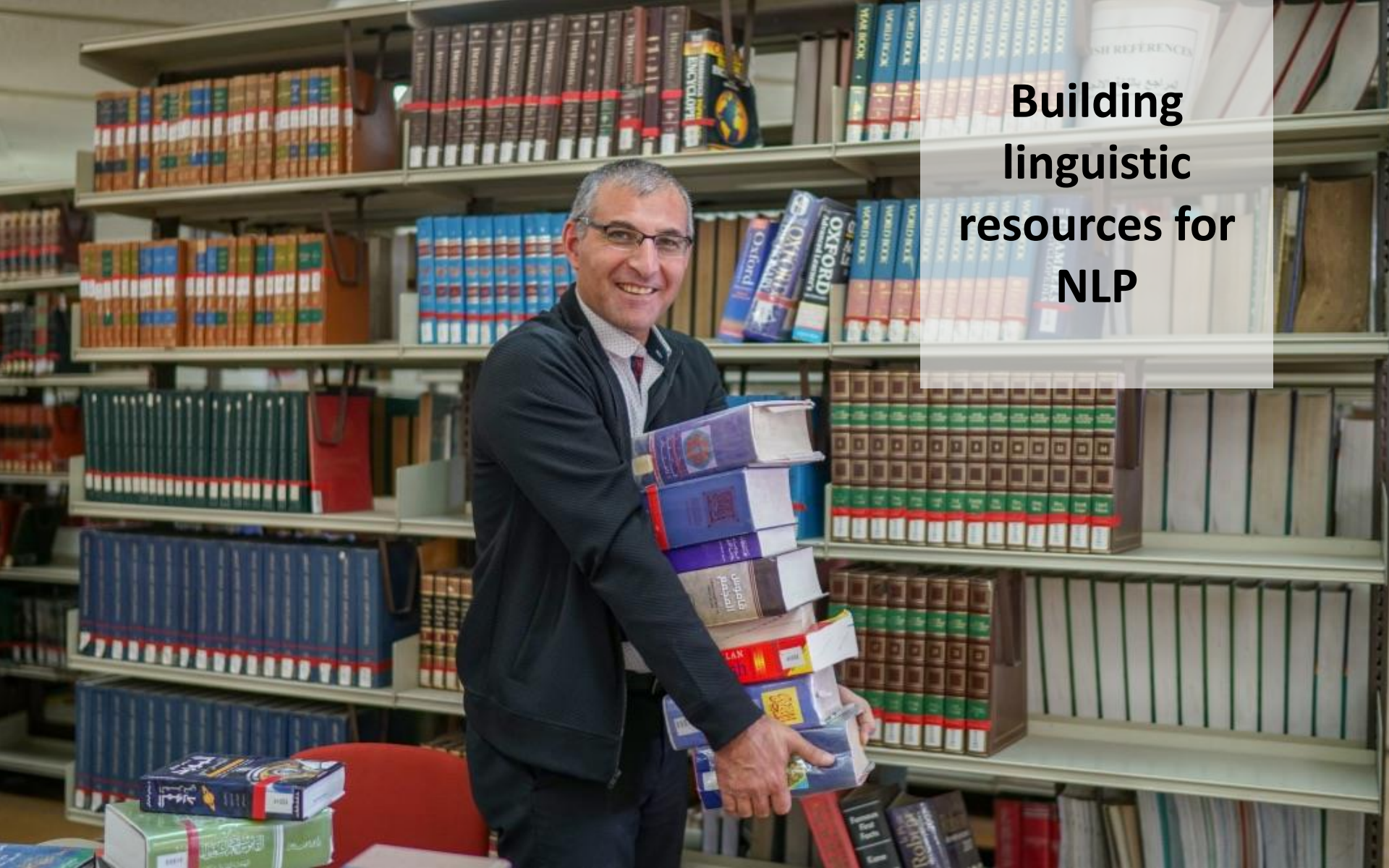
1.6  
Billion  
users  
(liturgical/  
second  
Language)

Other  
users for  
economic  
and security  
...

The background of the slide is a dense, colorful collage of various social media and technology icons. These include recognizable symbols such as a camera, a speech bubble, a hashtag, a Wi-Fi signal, and a document, all rendered in a variety of colors like blue, red, green, and orange. The icons are slightly blurred and overlap each other, creating a vibrant, digital atmosphere.

# Consequences of not supporting Arabic in ICT

On Education, culture, economy, politics, people with special  
needs, .....



**Building  
linguistic  
resources for  
NLP**



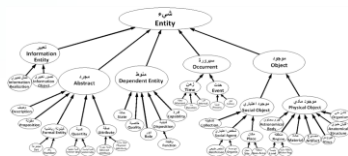
# Lexical Resources at Birzeit University

## Lexicographic Database



150 lexicons  
Largest Arabic lexicographic database

## Arabic Ontology/ Wordnet



Formal Arabic  
Wordnet  
with ontologically clean content

## Annotated Corpora



Arabic Dialects  
Morph. Annotations  
NER, Intents, hate  
....

## NLP library



APIs  
Linguistic Data,  
synonyms, Nested  
NER, intents, ...

WSD 84%

NER 88.4%

## Big Linguistic Data Graph

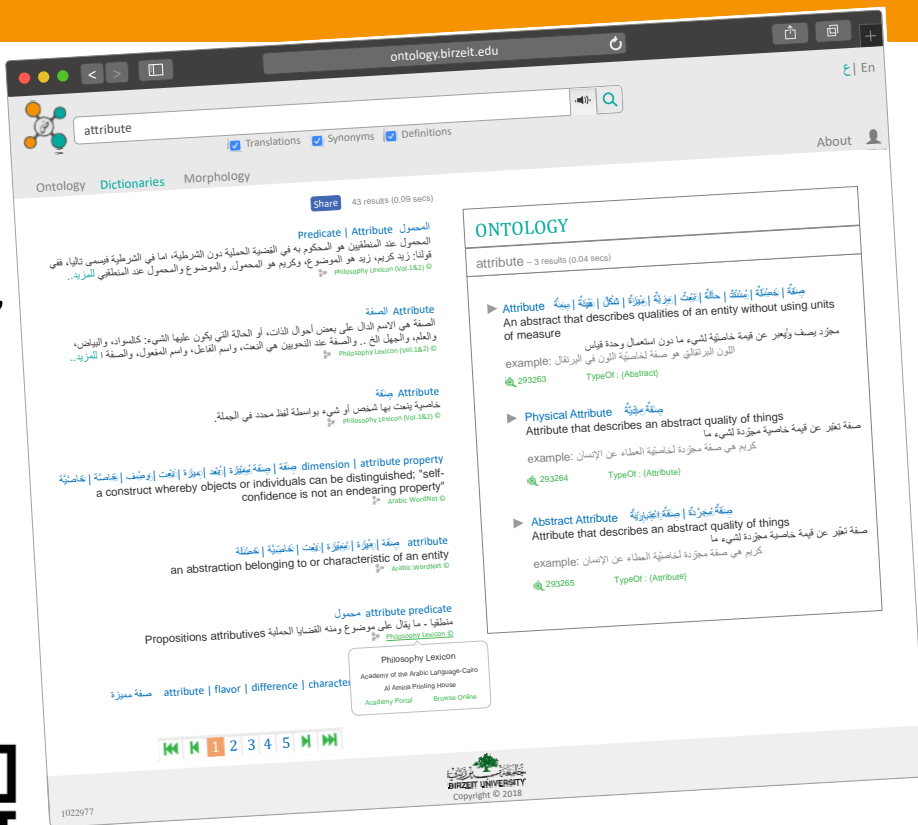
<https://ontology.birzeit.edu>





# The Lexicographic Database

- The largest lexicographic Arabic database (150 Lexicons)
- Contains most lexicon types: glossaries, thesauri, bi/trilingual dictionaries, morph datasets, **Arabic Ontology**, and more.
- Covers most domains: science, technology, law, business, art, philosophy, ...
- Digitized over 10 years.



# Some Statistics

## 150 Lexicons

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

- **API accessible** for NLP applications.



# 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

The screenshot shows the Arabic Ontology website interface. At the top, there is a search bar and navigation links for 'Translations', 'Synonyms', and 'Definitions'. The main content area displays the concept 'Entity' (شَيْءٌ) with its Arabic name and a definition: 'Whatever existed or will exist, and can be realized or imagined'. Below this, there are several related concepts listed with their Arabic names and brief descriptions:

- Object** (مَوْجُودٌ): An entity that is wholly and independently present in time, and is realized either for its concrete or social existence.
- Occurrent** (مَبْرُورٌ): An entity realized by the time of its happening.
- Dependent Entity** (مَعْتَمِدٌ): An entity realized by the time of its happening.
- Abstract** (مُجَرَّدٌ): An entity exists only in mind, cannot be measured or socially realized, and

The website footer includes the Birzeit University logo and the text 'Copyright © 2018'.

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



# Arabic Ontology



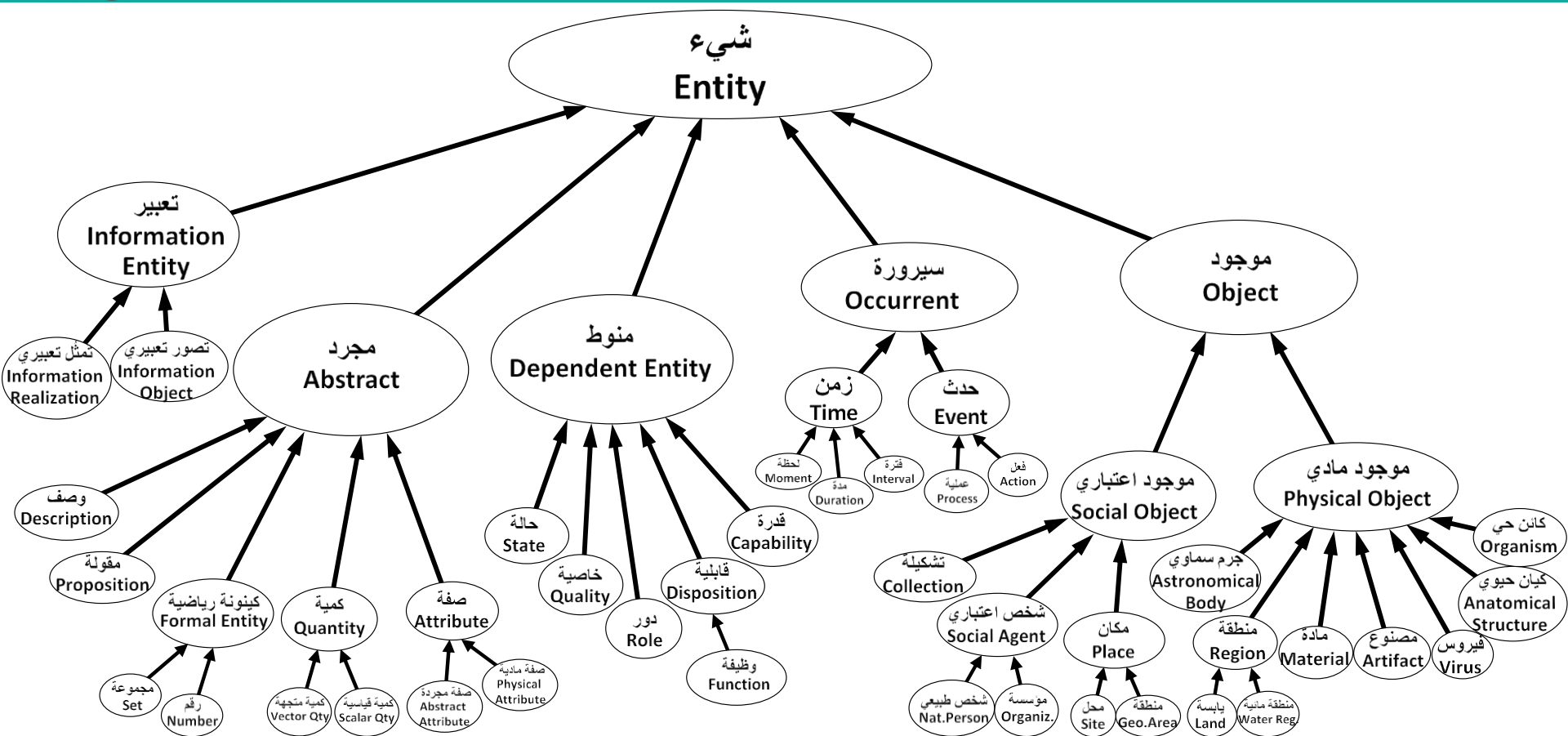
- **Current size** so far (but the numbers are dynamic)
  - 1800** fully-done concepts (mostly top levels)
  - 17K** partially investigated (ready for NLP applications)
- Some branches are elaborated, other not yet.
- **English labels** are not our target - provided for readability and communication.
- **Methodology:** Built top-down and bottom-up at the same time.



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



# Top Levels of the Arabic Ontology



**Curras**, a Palestinian  
morphologically  
annotated corpus (Revised)



**Baladi**, a Lebanese  
morphologically  
annotated corpus



مدونة اللهجات الشامية  
**Levantine Corpus**

65K tokens

El Haff, K., Jarrar, M., Hammouda, T., Zaraket, F., (2022). [Curras + Baladi: Towards a Levantine Corpus](#). In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France.

DA Sentence: بروحي ساكنة ممنون عيونك عيوني وانا عم آخذ المونة رح ظل شوف الدنيا لوما بقيلي عيون

| Token  | CODA   | Prefix(es) | Stem         | Suffix(es)        | POS       | Person | Aspect | Gender | Number | DA Lamma   | MSA Lemma  | Gloss      |
|--------|--------|------------|--------------|-------------------|-----------|--------|--------|--------|--------|------------|------------|------------|
| بروحي  | بروحي  | ب/PREP     | روح/NOUN     | ي/POSS_PRON_1S    | NOUN      | -      | -      | m      | s      | رُوح       | رُوح 1     | spirit     |
| ساكنة  | ساكنة  |            | ساكن/ADJ     | ة/NSUFF_FEM_SG    | ADJ       | -      | -      | f      | s      | ساكن 1     | ساكن 1     | residing   |
| ممنون  | ممنون  |            | ممنون/ADJ    |                   | ADJ       | -      | -      | m      | s      | مَمْنُون 1 | مَمْنُون 1 | grateful   |
| عيونك  | عيونك  |            | عيون/NOUN    | ك/POSS_PRON_2FS   | NOUN      | -      | -      | m      | 1 p    | عَيْن 1    | عَيْن 1    | eye        |
| عيوني  | عيوني  |            | عيون/NOUN    | ي/POSS_PRON_1S    | NOUN      | -      | -      | m      | p      | عَيْن 1    | عَيْن 1    | eye        |
| وانا   | وانا   | و/CONJ     | انا/PRON_1S  |                   | PRON      | 1      | -      | m      | s      | أنا 1      | أنا 1      | I          |
| عم     | عم     |            | عم/PROG_PART |                   | PROG_PART | -      | i      | -      | -      | عَم 0      | ظَلَّ 1    | continue   |
| آخذ    | آخذ    | آ/IV1S     | آخذ/IV       |                   | VERB      | 1      | i      | m      | s      | أَخَذُ 1   | أَخَذُ 1   | take       |
| الموني | المونة | ال/DET     | مون/NOUN     | ة/NSUFF_FEM_SG    | NOUN      | -      | -      | f      | s      | مُؤْنَة 1  | مُؤْنَة 1  | provisions |
| رح     | رح     |            | رح/FUT_PART  |                   | FUT_PART  | -      | -      | -      | -      | رَح 0      | سَوَّف 1   | will       |
| ضل     | ظل     |            | ظل/IV        |                   | VERB      | 1      | i      | m      | s      | ظَلَّ 1    | ظَلَّ 1    | remain     |
| شوف    | شوف    |            | شوف/IV       |                   | VERB      | 1      | i      | m      | s      | شَاف 0     | نَظَرَ 1   | looking    |
| الدني  | الدنيا | ال/DET     | الدنيا/NOUN  |                   | NOUN      | -      | -      | f      | s      | دُنْيَا 2  | دُنْيَا 2  | world      |
| لوما   | لوما   |            | لوما/CONJ    |                   | CONJ      | -      | -      | -      | -      | لُومَا 0   | لُولا 1    | except for |
| بقيلي  | بقيلي  |            | بقي/IV       | لي/IVSUFF_SUBJ:1S | VERB      | 3      | i      | m      | s      | بَقِيَ 1   | بَقِيَ 1   | remain     |
| عيون   | عيون   |            | عيون/NOUN    |                   | NOUN      | -      | -      | m      | p      | عَيْن 1    | عَيْن 1    | eye        |

| الترجمة                                | الترجمة   | الترجمة   | الترجمة                                      | الترجمة                                      | الترجمة   | الترجمة   | الترجمة  | الترجمة   | الترجمة   | الترجمة | الترجمة |
|--|---|---|--|--|---|---|--|---|---|---------|---------|
| ... يجمع السراويل ليس من صدمه - يور... | ... معروف فس نفس . بتطلع الخميس ليش هالقدّة الأزمة بقولك يوم الخميس دايمًا <b>هيك</b> , هذا يوم في الاسبوع طايير... | ... يلم اللي اكلناه , عزمونا الاسبوع الماضي , عشان نردلهم العزيمة , و <b>هيك</b> يدك تحطيلهم تترتسي | رجل 2 : <b>هيك</b> مزبطة بدها <b>هيك</b> ختم | رجل 2 : <b>هيك</b> مزبطة بدها <b>هيك</b> ختم | ... معروف فس نفس . بتطلع الخميس ليش هالقدّة الأزمة بقولك يوم الخميس دايمًا <b>هيك</b> , هذا يوم في الاسبوع طايير .... | ... ما باكل الا لحلال وما بشرب وما بصاحب! الله يستر عنجد , معقول في ناس <b>هيك</b> ما يتكفر بعضها ؟ ! عالم فاضية عفكرة بال... | ... صير تدلني امرار ع اطباق ما فيهن لحمه عشان انو ما كون محدود بخياراتي و <b>هيك</b> . | ... ها شي المكونات , قالتلي انتبه هيدي يمكن فيها كحول و أنت يمكن ما بتاكل <b>هيك</b> شي , و اختارتلي وحدة ما فيها كحول .... | ... ي , ان مش ساعة بيع بانغ انت ... بس في شي تغير , شي مش عارف شو هوي , <b>هيك</b> شي اختفى مثل لما اكلت مش بالبيت وحس... |         |         |
| هيك                                    | هيك   | هيك   | هيك  | هيك  | هيك   | هيك   | هيك  | هيك   | هيك   | هيك     | هيك     |
| هيك                                    | هيك   | هيك   | هيك  | هيك  | هيك   | هيك   | هيك  | هيك   | هيك   | هيك     | هيك     |
| هيك                                    | هيك   | هيك   | هيك  | هيك  | هيك   | هيك   | هيك  | هيك   | هيك   | هيك     | هيك     |
| هيك                                    | هيك   | هيك   | هيك  | هيك  | هيك   | هيك   | هيك  | هيك   | هيك   | هيك     | هيك     |
| هيك                                    | هيك   | هيك   | هيك  | هيك  | هيك   | هيك   | هيك  | هيك   | هيك   | هيك     | هيك     |
| هيك                                    | هيك   | هيك   | هيك  | هيك  | هيك   | هيك   | هيك  | هيك   | هيك   | هيك     | هيك     |
| هيك                                    | هيك   | هيك   | هيك  | هيك  | هيك   | هيك   | هيك  | هيك   | هيك   | هيك     | هيك     |
| هيك                                    | هيك   | هيك   | هيك  | هيك  | هيك   | هيك   | هيك  | هيك   | هيك   | هيك     | هيك     |
| هيك                                    | هيك   | هيك   | هيك  | هيك  | هيك   | هيك   | هيك  | هيك   | هيك   | هيك     | هيك     |

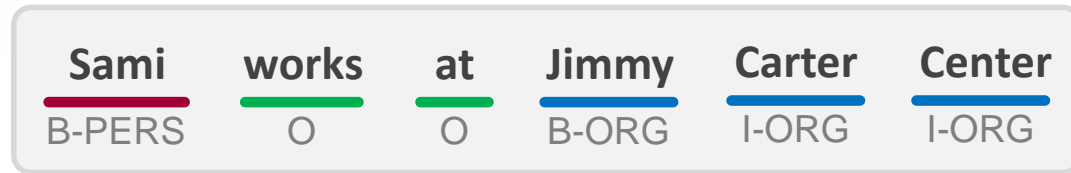




# Wojood

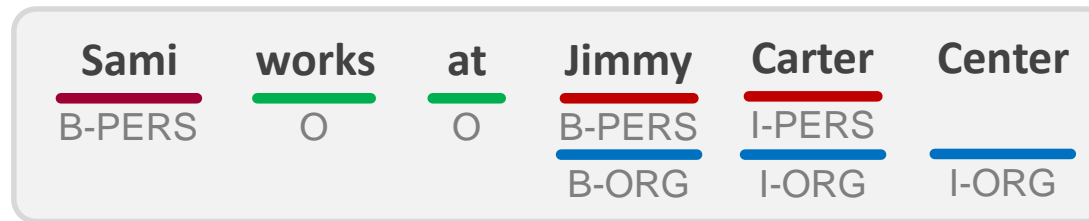
## Nested Arabic Named Entity Corpus

Flat NER



✓mature

Nested NER



!

Jarrar, M., Khalilia, M., Ghanem, S. (2022). [Wojood: Nested Arabic Named Entity Corpus and Recognition using BERT](#). In Proceedings of the International Conference on Language Resources and Evaluation (LREC 2022), Marseille, France.

# Wojood

## ❖ NER Corpus

| Corpus        | Nested? | size (tokens) | No. of entities | Entity classes | Which Arabic  | Domain  |
|---------------|---------|---------------|-----------------|----------------|---------------|---|
| Ontonotes5    | No      | 300k          | 28k             | 18             | MSA           | News  |
| ANERCorp      | No      | 150k          | 11k             | 4              | MSA           | News  |
| CanerCorpus   | No      | 258k          | 72k             | 14             | Classic       | Religion  |
| AQMAR         | No      | 74k           | -               | open           | MSA           | 4 domains   |
| Wojood Corpus | YES     | 550K          | 75K             | 21             | MSA & Dialect | <b>Multi domains</b><br>Media, History, Culture, Health, Finance, ICT, Law, Elections, Politics, Migration, Terrorism |

## ❖ NER BERT

- Multi-task learning (nested entities)
- 88.4% F1-score

# Annotation Guidelines

## 21 entity classes

**PERS** People names

فيروز، عادل إمام، ابن احمد، الملك عبدالله، النبي محمد

**NORP** Group of people

العرب، المسيحيين، سكان القدس، وزراء الخارجية العرب

**OCC** Occupation or professional title

رئيس جامعة بيرزيت، مدير بنك فلسطين، قائد الجيش

**ORG** Legal/social body

بنك القاهرة، ريال مدريد، داعش، الجيش المصري،

**GPE** Geopolitical: country, city, state

ليبيا، مدينة القدس، الجمهورية اللبنانية، روسيا الاتحادية

**LOC** Geographical location (non-GPE)

البحر الميت، قناة السويس، آسيا، الوطن العربي

**FAC** Places: landmark, road, building..

مطار صنعاء، سجن ابو غريب، المسجد الأقصى

**EVENT** Events of general interest

حرب 1973، القمة العربية 2005، عيد الفطر ، يوم الأرض

**DATE** Specific/relative date (>day)

13 يونيو، 2019-2020، الفترة العثمانية

**TIME** Specific/relative time (<day)

الساعة ١٢ ، من الساعة 5 حتى 7 مساء، خلال ساعتين

**LANGUAGE** Human language or dialect

اللغة العربية، الفصحى، الدارجة المغربية، اللغة الفرنسية

**WEBSITE** Website or specific URL

موقع فيسبوك، يوتيوب، schema.org

**LAW** Geographical location (non-GPE)

قانون الاستثمار ، المادة 114 من قانون العقوبات 2005

**PRODUCT** Vehicle, weapon, food, ...

مرسيدس سي ١٨٠، ايفون ١٣، دبابه مركابا، تروفين

**CARDINAL** Numerals in digits/words

150، 30، 1.5، صفر، اثنان ، أربعة وعشرون ، مليون

**ORDINAL** does not refer to a quantity

٣ كيلومتر، مئة قدم ، 3 طن ، 50 غرام ، 25 سم مربع

**PERCENT** Word/symbol refers to a percent

5 بالمئة ، 10% ، 9 من كل الف

**QUANTITY** Value measured by units

٣ كيلومتر، مئة قدم ، 3 طن ، 50 غرام ، 25 سم مربع

**UNIT** Name/symbol of a unit

ميل ، كيلو ، كيلومتر ، إنش ، كيلوغرام ، هكتار ، مل

**MONEY** Monetary quantity, incl.

مئة وخمسون درهم اماراتي ، اثنان وثلاثون يورو ، 8 دولار

**CURR** Name/symbol of currency

دولار ، جنيه مصري ، دينار ، فرنك ، ريال سعودي € ،

# Corpus Collection

| Source - Topics   | Sentences     | Tokens         |
|---|---------------|----------------|
| Web Articles <sup>1</sup> (MSA)<br>Health, Finance, ICT, Law, Elections,<br>Politics, Migration and Terrorism | 9,053         | 258,102        |
| Archive <sup>2</sup> (MSA)<br>History and Culture   | 12,271        | 227,020        |
| Social Media <sup>3</sup> (Dialect)<br>General topics   | 5,653         | 65,342         |
| <b>Total</b>  | <b>26,977</b> | <b>550,464</b> |

<sup>1</sup> un.org, hrw.org, msf.org, who.org, mipa.institute, elections.ps, sa.usembassy.gov, diplo- matie.ma, quora.com ...

<sup>2</sup> Awraq, Birzeit University Digital Palestinian Archive

<sup>3</sup> Palestinian and Lebanese dialect corpora

# Nested NER Results

| Tag      | Precision | Recall | F1-Score |
|----------|-----------|--------|----------|
| PERS     | 0.9135    | 0.9122 | 0.9129   |
| NORP     | 0.6828    | 0.7037 | 0.6931   |
| OCC      | 0.7993    | 0.8402 | 0.8193   |
| ORG      | 0.8924    | 0.9072 | 0.8997   |
| GPE      | 0.9424    | 0.9516 | 0.9470   |
| LOC      | 0.8054    | 0.7059 | 0.7524   |
| FAC      | 0.7366    | 0.6481 | 0.6895   |
| PRODUCT  | 0.3333    | 0.2500 | 0.2857   |
| EVENT    | 0.6364    | 0.6488 | 0.6425   |
| DATE     | 0.9253    | 0.9394 | 0.9323   |
| TIME     | 0.6000    | 0.5122 | 0.5526   |
| LANGUAGE | 0.9310    | 0.7105 | 0.8060   |
| WEBSITE  | 0.4496    | 0.5472 | 0.4936   |
| LAW      | 0.8525    | 0.9123 | 0.8814   |
| CARDINAL | 0.8437    | 0.8575 | 0.8505   |
| ORDINAL  | 0.9411    | 0.9448 | 0.9430   |
| PERCENT  | 0.2903    | 0.9310 | 0.4426   |
| QUANTITY | 0.2500    | 0.1667 | 0.2000   |
| UNIT     | 0.5000    | 0.1667 | 0.2500   |
| MONEY    | 0.9143    | 0.8205 | 0.8649   |
| CURR     | 0.8810    | 0.9487 | 0.9136   |
| Overall  | 0.8772    | 0.8909 | 0.8840   |

Entities of the same time  
are not supported yet

# Downloads and Demo

<https://ontology.birzeit.edu/wojood>

بيريت وبالتعاون مع اادوارد في تنظيم مهرجان  
في الشعبي، سيبدأ المهرجان الربط عصرا، بتاريخ  
16/5/2022، وذلك في بنك فلسطين بقيمة خمسة الاف دولار.

output format : highlighted ▼

جامعة بيرزيت GPE وبالتعاون مع مؤسسة ادوارد سعيد PERS ORG تنظم مهرجان للفن الشعبي، EVENT

سيبدأ المهرجان الساعة الرابعة عصرا، TIME بتاريخ 16/5/2022، DATE وذلك برعاية من بنك فلسطين GPE

بمبلغ خمسة الاف دولار CURRENCY MONEY ORG

# ArabGlossBERT

## Arabic Word-Sense Disambiguation

### The Word Sense Disambiguation (WSD) Task

Given a word in a context, which sense (i.e. meaning) this word denotes?

قصيدة من عيون الشعر

#### Set of senses

1. عُضُو الإِبْصَار فِي الْإِنْسَانِ وَالْحَيَوَانَ: لَهُ عَيْنَانِ كَعَيْنَيْ الصَّقْرِ - أَلَا إِنَّمَا الْعَيْنَانِ لِلْقَلْبِ رَائِدٌ ...
2. جَاسُوسٌ، "كَانَ عَيْنًا لِدَوْلَةٍ أُجْنَبِيَّةٍ . بِنْتُ الْعَيُونِ : تَجَسَّسٌ، رَاقِبٌ - فَلَانٌ عَيْنٌ عَلَى فَلَانٍ : نَاطِرٌ عَلَيْهِ
3. أَجُودُ كُلِّ شَيْءٍ وَأَحْسَنُهُ وَنَفِيسُهُ: عَيُونُ الْفَنِّ.
4. حَارِسٌ: فَلَانٌ عَيْنٌ عَلَى الْمَكَانِ.
5. الْحَاضِرُ مِنْ كُلِّ شَيْءٍ أَصْبَحَ أَثَرًا بَعْدَ عَيْنٍ ...
6. عَيْنُ الْمَاءِ:- يَنْبُوعُهُ، تُحَلِّقُ الطَّيُورُ فَوْقَ عَيُونِ الْمَاءِ
7. عَيْنُ الشَّيْءِ:- نَفْسُهُ، ذَاتُهُ (تَسْتَعْمَلُ لِلتَّوَكِيدِ): جَاءَ الْقَوْمُ أَعْيُنَهُمْ - كُنَّا فِي الْمَكَانِ عَيْنَهُ.
8. عَيْنُ الْعَقْلِ:- قُدْرَةُ ذَهْنِيَّةٍ مَوْرُوثَةٌ عَلَى التَّخْيُّلِ وَتَذَكُّرِ الْأَحْدَاثِ.
9. ....

# ArabGlossBERT

## Arabic Word-Sense Disambiguation

### ❖ Arabic context-gloss pairs Dataset (167k)

- Extracted from Birzeit University's Lexicographic database
- Annotated target words in context;

| Gloss                                 | Context                         | Label |
|---------------------------------------|---------------------------------|-------|
| [SEP] أجود كل شيء وأحسنه ونفيسه [SEP] | [CLS] قصيدة من عيون الشعر [SEP] | True  |
| [SEP] ذاته ( تستعمل للتوكيد ) [SEP]   | [CLS] قصيدة من عيون الشعر [SEP] | False |
| [SEP] ذاته ( تستعمل للتوكيد ) [SEP]   | [CLS] جاء القوم أعينهم [SEP]    | True  |
| [SEP] أجود كل شيء وأحسنه ونفيسه [SEP] | [CLS] جاء القوم أعينهم [SEP]    | False |

### ❖ Three Fine-tuned BERT Models

- WSD into **binary sequence-pair classification task**
- **Accuracy 84%**
- 4 types of signals to emphasize target words in context



## ❖ Fine-tuned three Arabic pre-trained BERT models

The WSD task is converted into **binary sequence-pair classification task**

[CLS] قصيدة من عيون الشعر [SEP] أجود كل شيء وأحسنه ونفيسه [SEP] True  
 [CLS] قصيدة من عيون الشعر [SEP] عين الشيء : نفسه ، ذاته ( تستعمل للتوكيد ) [SEP] False  
 [CLS] جاء القوم أعينهم [SEP] عين الشيء : نفسه ، ذاته ( تستعمل للتوكيد ) [SEP] True  
 [CLS] جاء القوم أعينهم [SEP] أجود كل شيء وأحسنه ونفيسه [SEP] False

## ❖ Results

| Model      |           | True | False | Accuracy |
|------------|-----------|------|-------|----------|
| AraBERTv02 | Precision | 81   | 85    | 84       |
|            | Recall    | 66   | 93    |          |
|            | F1-score  | 72   | 89    |          |
| CAMELBERT  | Precision | 77   | 83    | 82       |
|            | Recall    | 60   | 92    |          |
|            | F1-score  | 67   | 87    |          |
| QARiB      | Precision | 73   | 82    | 80       |
|            | Recall    | 58   | 90    |          |
|            | F1-score  | 65   | 86    |          |

# Constructing a dataset of context-gloss pairs

## Statistics:

|                               | <b>count</b> |
|-------------------------------|--------------|
| Unique Lemmas (undiacritized) | 26169        |
| Avg glosses per Lemmas        | 1.25         |
| Unique Glosses                | 32839        |
| Unique Contexts               | 60272        |
| Avg context per gloss         | 1.83         |
| True context-gloss pairs      | 60323        |
| False context-gloss pairs     | 106884       |
| Total True and False pairs    | 167207       |

## Training and Test Datasets

- every context selected in the test set should not be selected in the training set;
- every gloss should be selected in both the training and the test sets.

| <b>Datasets</b> | <b>Pairs</b> | <b>Count</b>   | <b>Total</b> |
|-----------------|--------------|----------------|--------------|
| Training        | True pairs   | 55,585         | 152,035      |
|                 | False pairs  | 96,450         |              |
| Test            | True pairs   | 4,738          | 15,172       |
|                 | False pairs  | 10,434         |              |
|                 | <b>Total</b> | <b>167,207</b> |              |

Download: <https://ontology.birzeit.edu/downloads>

# ArabGlossBERT

## Arabic Word-Sense Disambiguation



Arabic Language Understanding Tools

..... Confirm

قصيدة من عيون الشعر

Tokenize Lemmatize NER WSD

◀ قصيدة (قصيدَة 1 303044571): مجموعة من الأبيات الشعرية متحدة في الوزن والقافية والرؤي وهي تتكون من سبعة أبيات فأكثر "قصيدة غزلية".  
يبتئ القصيدة: البيت المتضمن غاية الشاعر، أو أنفاس أبياتها، أو مثل يضرب في تفضيل بعض الشيء على كله - مطلع القصيدة: أول بيت منها

◀ من (من 1)

◀ عيون (عَيْن 2 303038477): أجود كل شيء وأحسنه ونقيسه "قصيدة من عيون الشعر - عيون الفن".

◀ الشعر (شِعْر 1 303029103): كلام موزون مقفى قصداً يعتمد على التخيل والتأثير؛ ليوحي بإحساسات مؤثرة وصور خيالية "شعر صافي الديباجة - نظم الشعْر - ما الشعْر إلا شعور المرء يُرسله ... عفو البديهة عن صدق وإيمان - إنَّ مِنَ الشعْرِ لِحِكْمَةٌ [حديث] - حَوْماً عَلَّمْنَاهُ الشعْرَ وَمَا يَنْبَغِي لَهُ <يس/". أنشدته الشعْر: قرأه عليه - أو ابد الشعْر: ما لا تماثل جودته أو قوافيه الشاردة - ربَّ الشعْر: إلهة الشعْر عند الوثنيين - شطرا بيت الشعْر: الصدر

جميع الحقوق محفوظة © 2022 جامعة بيرزيت

BIRZEIT UNIVERSITY



# Visit Our NLP Library



<https://ontology.birzeit.edu/lexapi>



Search Arabic and English terms ...



Translations  Synonyms  Definitions



[Ontology](#) [Dictionaries](#) [Morphology](#)

[About](#) [License](#)

## NLP Library

Download and access NLP data, corpora, tools and services

We developed hundreds of RESTful web services (APIs) for other third-party software developers to directly download and access our lexicographic databases, corpora, and tools. Some are listed on this page. Please get in touch with us for details.

The "sampleKey" API token allows you to use all web services but with a limited number of requests (enough to try it!), if you need more requests, you need to [Request API Token](#).

### APIs:

+ Search 150 dictionaries (synonyms, translations, glosses)

+ Arabic Ontology

+ Dialect Corpora

+ Word Sense Disambiguation

+ Arabic Named Entity Recognition (Wojood)

+ Lemmatizer

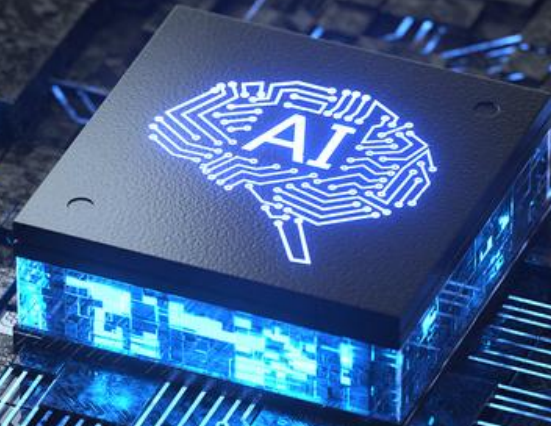
+ Retrieve Morphology Information (MSA and Dialects)

+ Arabic Processing Tools

+ Autocomplete Service (Arabic and English)

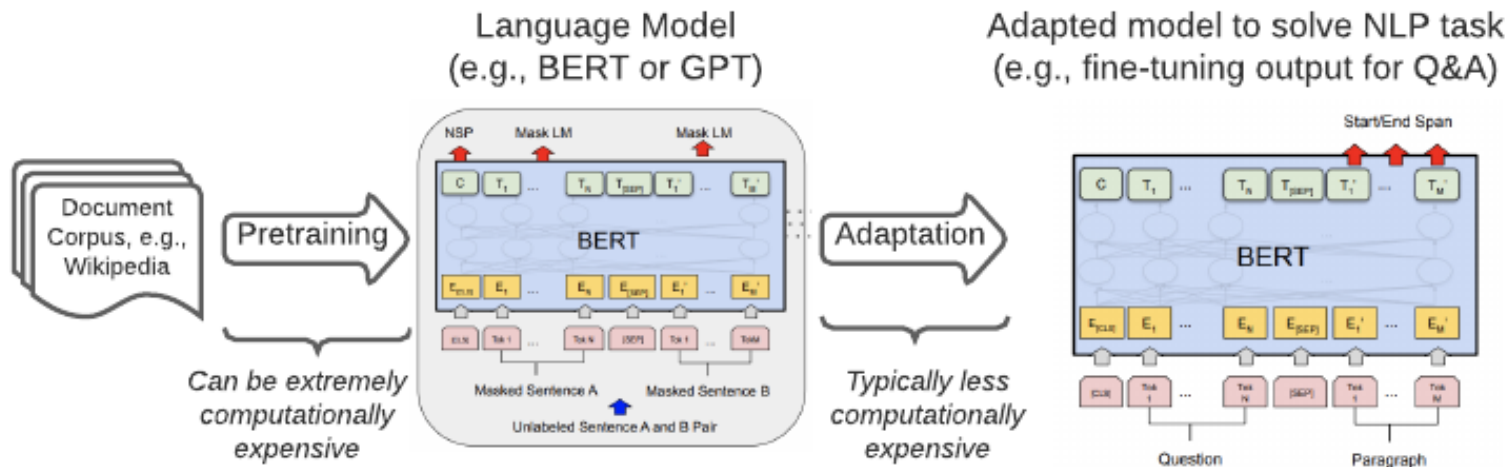


AI research  
needs  
GPUs  
(at least)



# Computing Resources

## Deep Learning



AI research Language is not possible without **fast GPUs**

Each experiment takes **many, many hours/days**

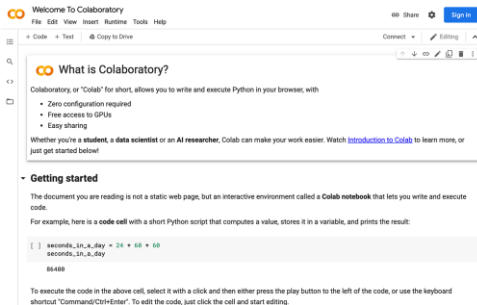
# Cloud is expensive of students



## Google Cloud

Currently, each student have to pay **50\$/month to Google**

Google also offer Colab



# GPUs are not good enough!

CPU



X

GPU



✓

Takes hours/days  
to train a language  
model

TPU

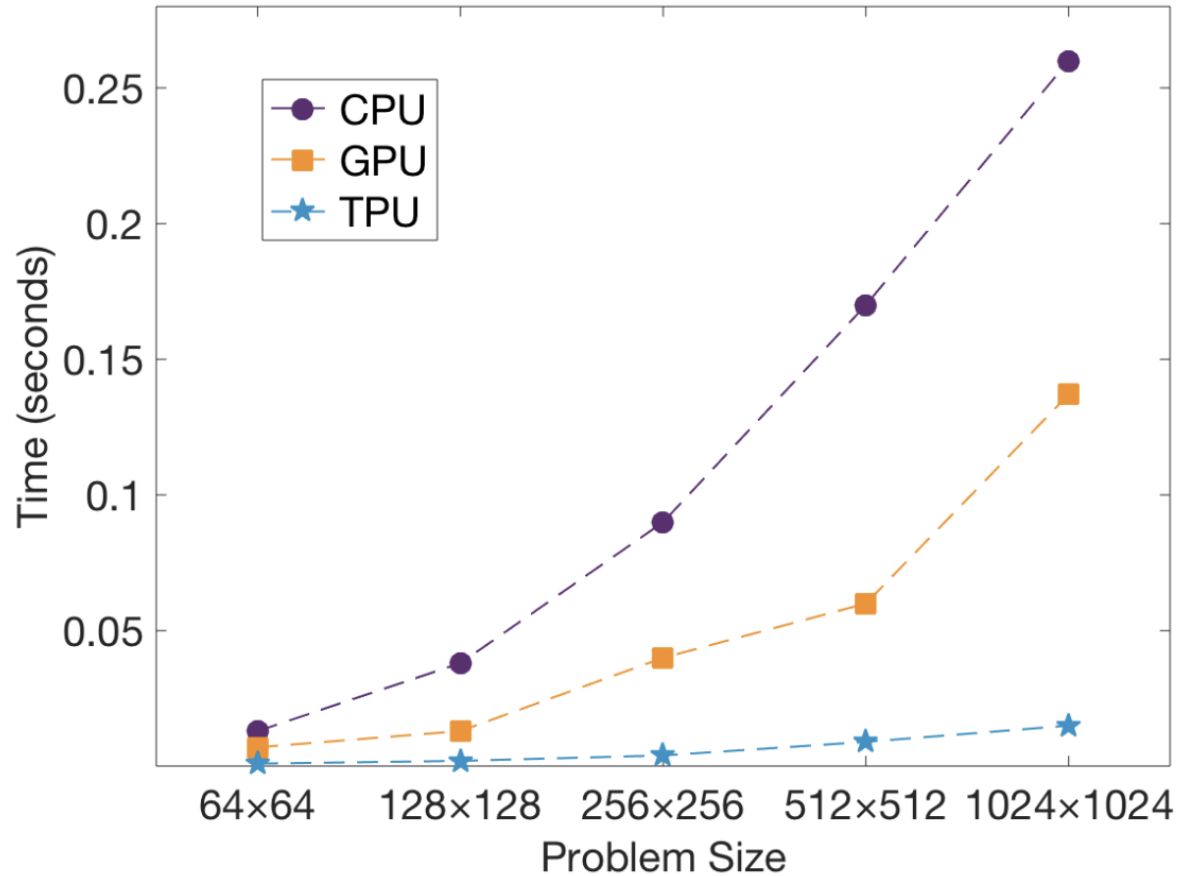


✓

Really fast!



# CPU vs GPU vs TPU



Source: <https://arxiv.org/pdf/2103.11927.pdf>

# Recommendations

## **Build own TPU Cloud**

**For top 20 AI centers  
in the Arab countries**

## **Negotiate Google/Amazon**

**For affordable prices  
for more researchers**

# Thank You

Mustafa Jarrar  
mjarrar@birzeit.edu