# NAMED ENTITY RECOGNITION IN PRE-MODERN ARABIC BIOGRAPHICAL TEXTS

Yuri Ishida (Okayama University, Japan), Kensuke Baba (Fukuoka Institute of Technology, Japan), Takahiro Baba (Kurume Institute of Technology, Japan)

#### What is NER?

Picking out proper nouns (names of persons, places, organizations, etc.) from the text.

#### **Annotation Method: IOB2 tagging**

### Yuri belongs to Okayama University in Japan

B-PERS O

)

C

B-ORG

5

I-ORG

O B-GPE

B = the beginning of the entity mention

I = the inside token

O = Outside entity

PERS = People names, including first, middle, last and nicknames.

ORG = Legal or social bodies like institutions, companies, agencies, teams, parties, armies, and governments.

GPE = Geopolitical like countries, cities, and states.

(See, Mustafa Jarrar et al., "Wojood: Nested Arabic Named Entity Corpus and Recognition using BERT," arXiv, 2022, https://doi.org/10.48550/arXiv.2205.09651)

#### Why Arabic?

- ✓ One of the six official languages of the United Nations.
- ✓ Native population of approximately 500 million.
- The grammatical features make it difficult for NLP.

#### Research Target: Pre-modern (17-18C) Arabic

Classical Arabic (7-11C): Quran, Hadith

Pre-Modern Arabic (17-18C): Biography, History, Philosophy

→Rich data for names of Persons & Places

Modern Standard Arabic (MSA, 19C -): Newspapers, SNS

Goal: Tracking of Muslim Scholars' location during the formative age of Islamic Reformism (17C-)

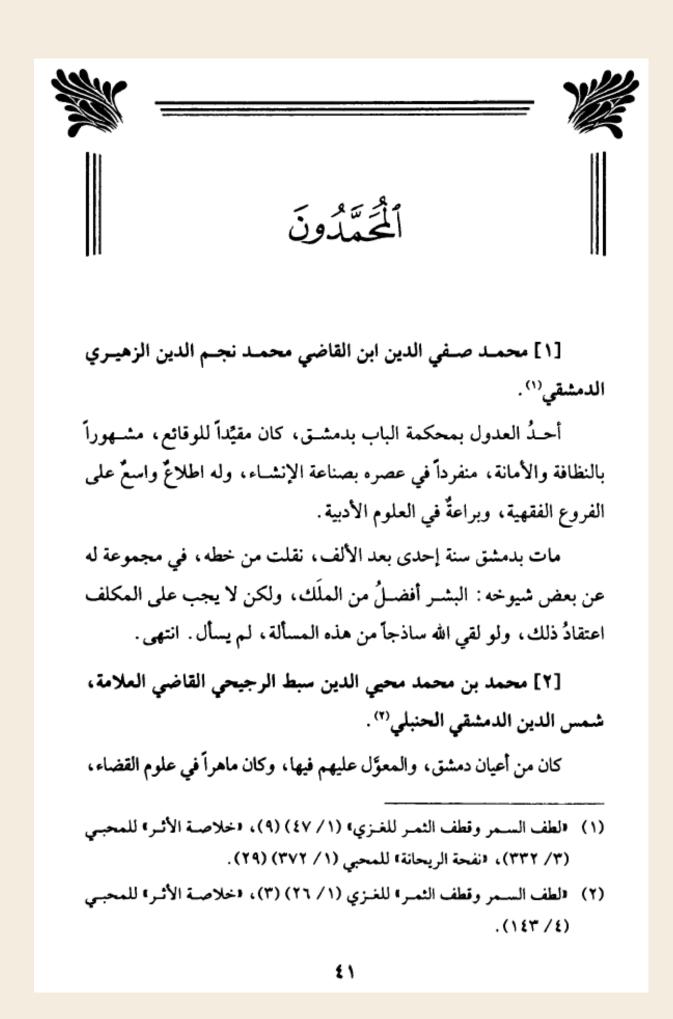
It is said that Scholars are gathered in Mecca and Medina and form Islamic Reformism → SUBSTANTIATE

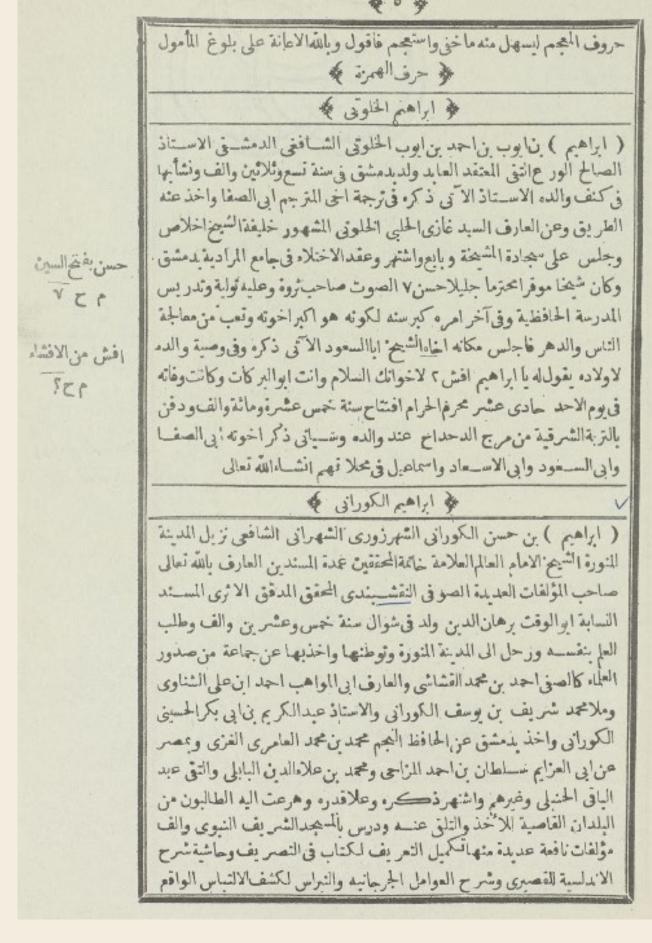
#### Sources

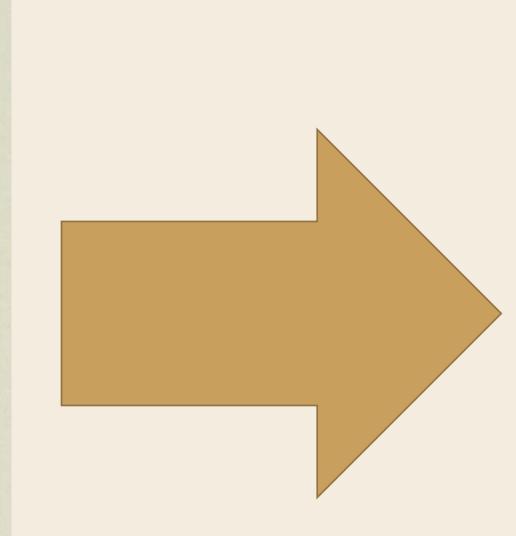
- 1. Khulāṣat al-athār by Muḥammad Amīn ibn Faḍl Allāh al-Muḥibbī (1651-1699): Text data available by OpenITI
- 2. Fawā'id al-irtiḥāl by Muṣṭafā Ibn-Fatḥallāh al-Ḥamawī (1633/34–1711-12)
- 3. Silk al-durar by Muḥammad Khalīl al-Murādī (d. 1791): Text data available by OpenITI

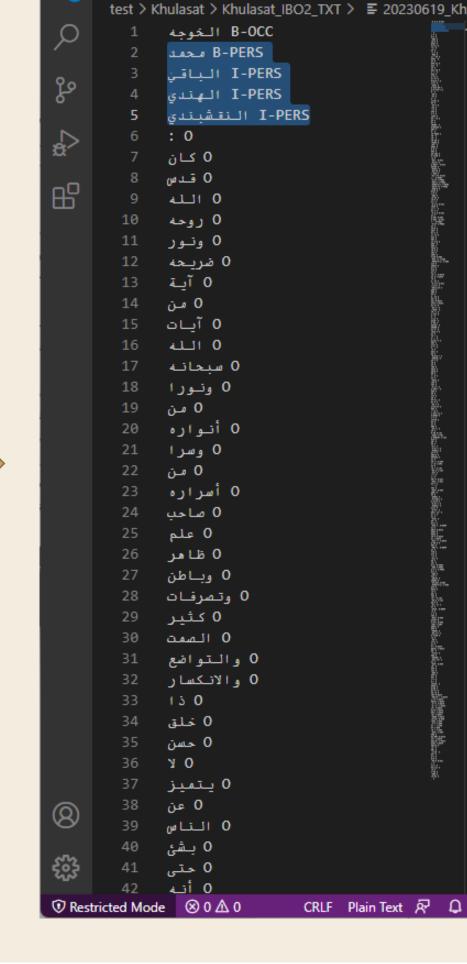
🍙 حرف الهمزة 🖪 من اسمه : آدم (١) - آدم الرومي الأنطالي الحنفي: الأستاذ الشهير، أحد خلفاء طريقة العارف بالله تعالى جلال الدين الرومي المعروف بـ « منلا خداوند كار » وكان شيخ زاويتهم المعروفة بمدينة الغلطة ، وليها في سنة إحدى وأربعين وألف، وكان له الحظوة التامة عند أركان دولة بني عشمان، سلاطين زماننا، نصرهم الله تعالى، لا يزال مجلسه غاصًا بأعيانهم، وهو من بيت كبير بأنطالية، على وزن : أنطاكية، بلدة كبيرة بأراضي قرمان على ساحل البحر الرومي، وطاؤها في نطَّق العوام تبدل ضاداً، ويحذفون نونها فيقولون : إضالية، ولبيتهم فيها أملاك وتعلقات جمة، وكان مائلاً إلى الترفه والاحتشام الزائد، وكان إذا ركب مشي في ركابه ما يقارب المائة رجل، من حفدته ومريديه، وكان للناس عليه إِقبال زائد، ومع ذلك كان ملازمًا على العبادة والوعظ، وكان يحل المثنوي حلاً جيدًا، وكان في أوائل أمره مفرط السخاء لا تكاد عطيته تنقص عن مائة دينار . وحكى بعض الأفاضل ممن يعرفه أنه كان في عهد السلطان مراد ظهر شخص يتقن ضرب الطنبور، فشغف به السلطان وطلبه ليلة فوجده عند آدم هذا، فأتوا به، فقال له : كم كانت جائزتك ؟ فقال : ها هي بيدي، وكانت مائة دينار، وكان لمشايخ الغلطة في ذلك العهد ميقات في داخل حرم السلطنة في كل شهر ليلة يقيمون فيها السماع بحضرة السلطان، ولهم تعايين، فحضر آدم ليلة ومعه جماعته، وأقاموا السماع فأمر السلطان بأن ينقص معلومهم بمسمع من آدم، وقال لجماعته : قولوا له : العطايا مهما كثرت لا تبلغ عطيته، فكف من ذلك العهد كفه عن الإفراط، واقتصر على ما هو متعارف عند الدولة، وسافر آخر أمره إلى القاهرة من طريق البحر بنية الحج في جمادي الأخرة سنة ثلاث وستين وألف، فمرض بمصر مدة وتوفي بها، وكانت وفاته في شهر رمضان سنة ثلاث وستين وألف، رحمه الله تعالى .

خلاصة الأثر في أعيان القرن الحادي عشر/الجزء الأول -









₹ 20230619\_Khulasat\_1112\_Baqi Billah\_IBO2.tx/ 🔲

# Making NER

- ✓ Rule Based & Machine Learning
- ✓ Training Data

Classical Arabic (CANERCorpus), Pre-Modern (NEW!), MSA (AQMAR & ANERCorp)

✓ Precision, Recall, Accuracy

## Conclusions

The Impact of Damascus Not Moroccan & Maliki?

Hometown	N=1284	%
Damascus	130	10
Egypt (Misr)	76	6
Yemen	75	6
Turkey (Rum)	36	3
Mecca	34	3
Jersalem (Quds)	25	2
Morroco (Maghrib)	19	1
Medina	14	1