

Multilingual Event Linking to Wikidata

Adithya Pratapa, Rishubh Gupta, Teruko Mitamura

Multilingual Information Access (MIA) @ NAACL 2022



Carnegie Mellon University

Language Technologies Institute

Overview

Overview

- We present the task of linking *event* references to a KB

Overview

- We present the task of linking *event* references to a KB
- For this task, we provide
 - a large scale multilingual dataset
 - two evaluation settings, multilingual and crosslingual
 - retrieve+rank model that improves upon a BM25 baseline

Background: Linking

Background: Linking

- Grounding (/linking) of (textual) concepts to a context (Chandu et al., 2021)
 - concepts: entities & events
 - context: a knowledge base (Wikipedia, Wikidata, Freebase etc.,)

Background: Linking

- Grounding (/linking) of (textual) concepts to a context (Chandu et al., 2021)
 - concepts: entities & events
 - context: a knowledge base (Wikipedia, Wikidata, Freebase etc.,)
- Linking differs from typing (person, infection, organization, attack etc.,)

Grounding 'Grounding' in NLP (Chandu et al., Findings 2021)

Background: Event Linking

Background: Event Linking

- Linking complements coreference
 - Nothman et al., 2012 proposed linking event mentions to first-report articles from a news archive
 - Linking helps avoid the notion of partial identity in event coreference (Nothman et al., 2012, Pratapa et al., 2021)

[Event Linking: Grounding Event Reference in a News Archive](#) (Nothman et al., ACL 2012)

[Cross-document Event Identity via Dense Annotation](#) (Pratapa et al., CoNLL 2021)

Background: Event Linking

- Linking complements coreference
 - Nothman et al., 2012 proposed linking event mentions to first-report articles from a news archive
 - Linking helps avoid the notion of partial identity in event coreference (Nothman et al., 2012, Pratapa et al., 2021)
- We propose linking mentions to entries in Wikidata

[Event Linking: Grounding Event Reference in a News Archive](#) (Nothman et al., ACL 2012)

[Cross-document Event Identity via Dense Annotation](#) (Pratapa et al., CoNLL 2021)

Background: Event Linking

- Linking complements coreference
 - Nothman et al., 2012 proposed linking event mentions to first-report articles from a news archive
 - Linking helps avoid the notion of partial identity in event coreference (Nothman et al., 2012, Pratapa et al., 2021)
- We propose linking mentions to entries in Wikidata
- Our work is closely related to Botha et al., 2020

[Event Linking: Grounding Event Reference in a News Archive](#) (Nothman et al., ACL 2012)

[Cross-document Event Identity via Dense Annotation](#) (Pratapa et al., CoNLL 2021)

[Entity Linking in 100 Languages](#) (Botha et al., EMNLP 2020)

Structure in Wikidata

Structure in Wikidata

Q734020

**Men's 100m
@ 2012 Summer
Olympics**

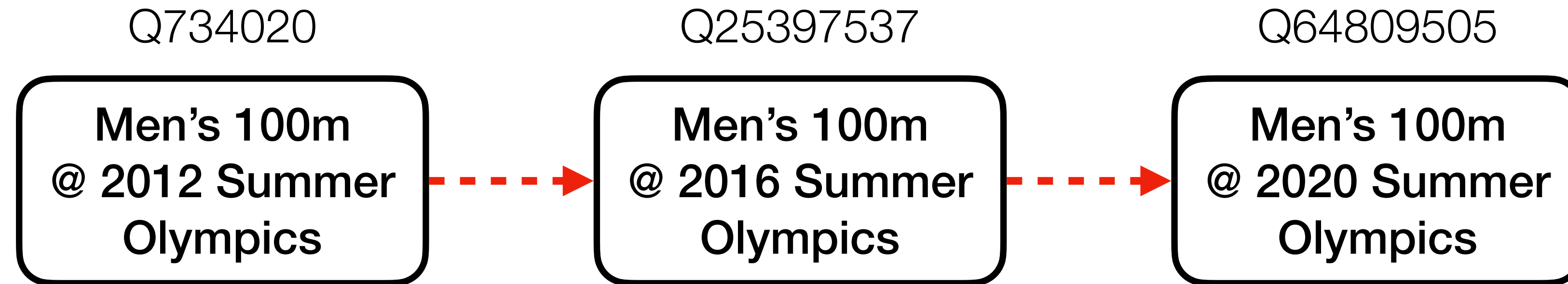
Q25397537

**Men's 100m
@ 2016 Summer
Olympics**

Q64809505

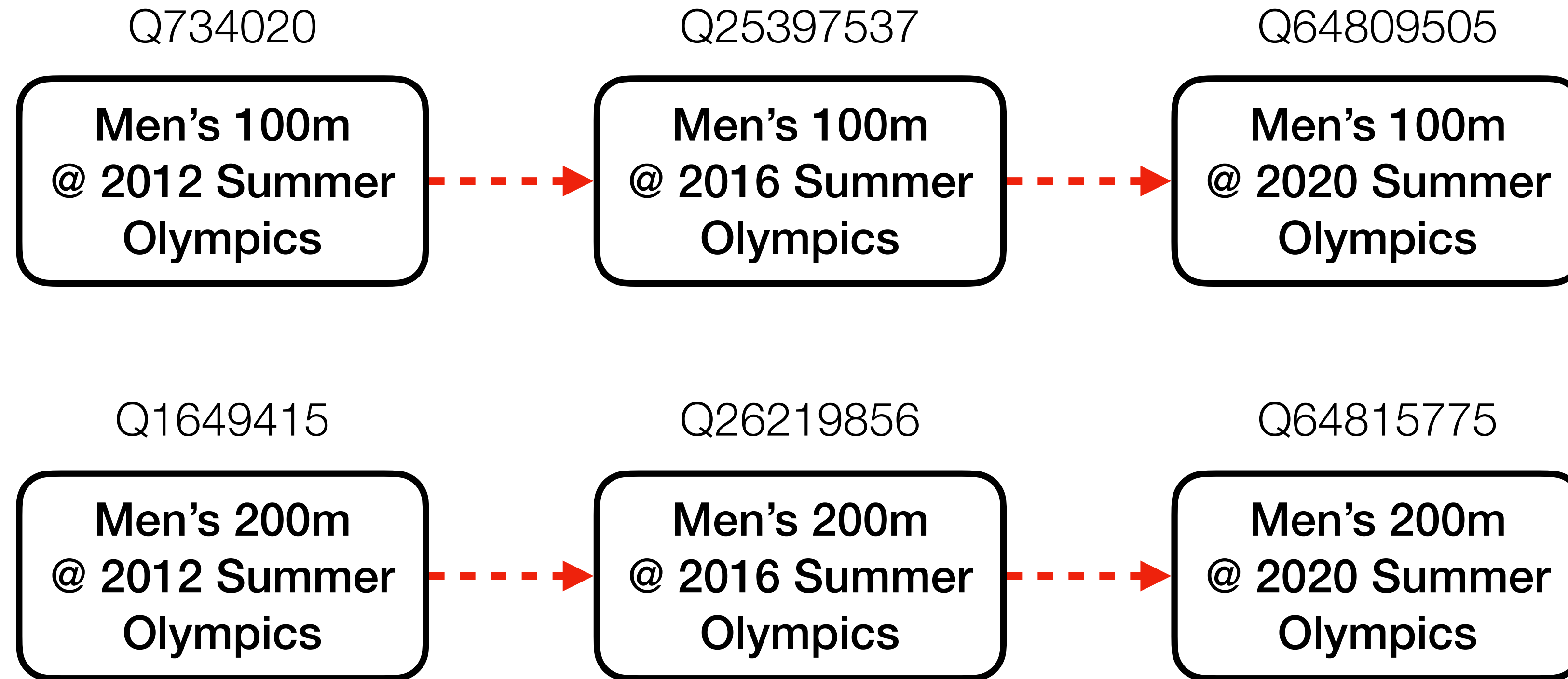
**Men's 100m
@ 2020 Summer
Olympics**

Structure in Wikidata



An illustration of temporal sequences of Wikidata Events

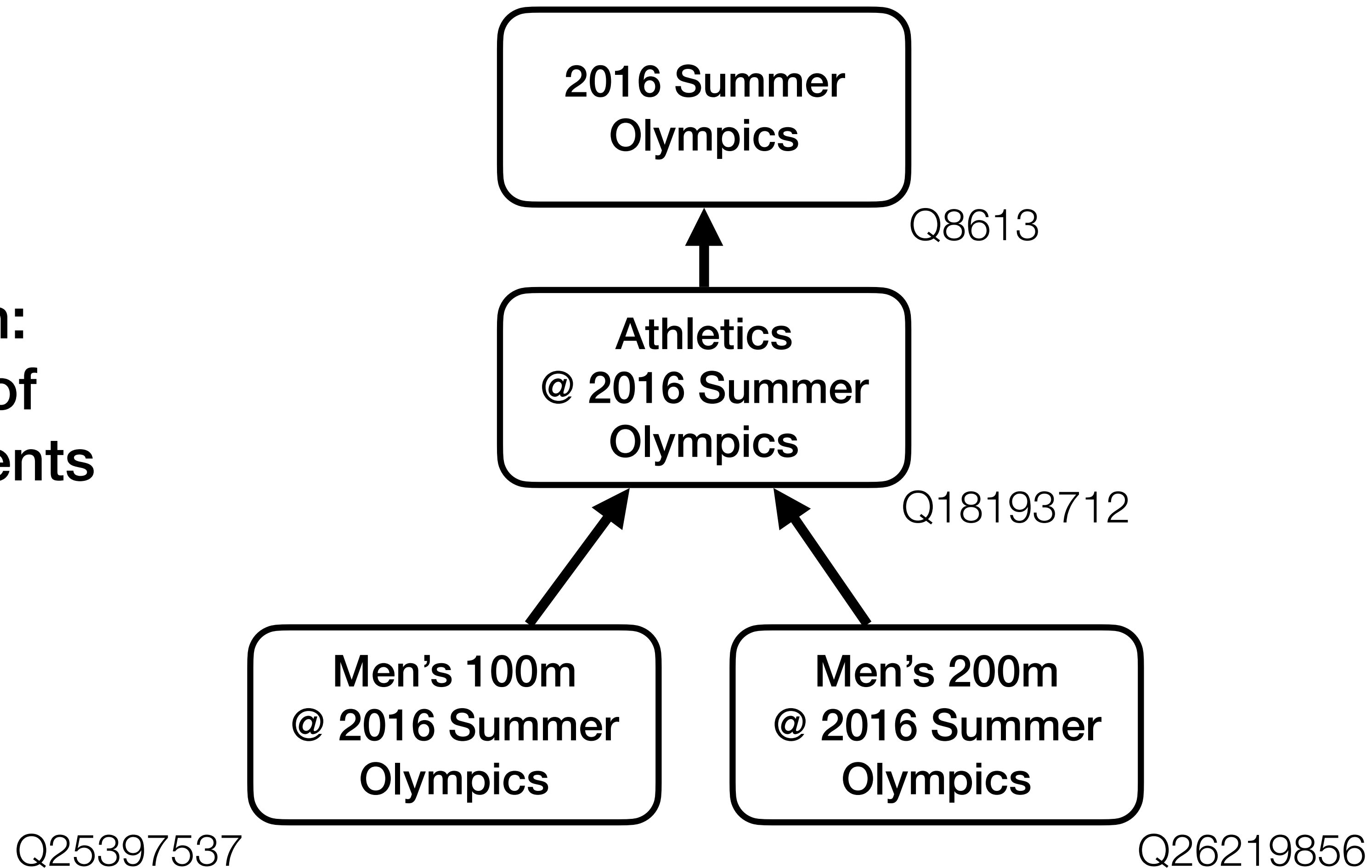
Structure in Wikidata



An illustration of temporal sequences of Wikidata Events

Structure in Wikidata

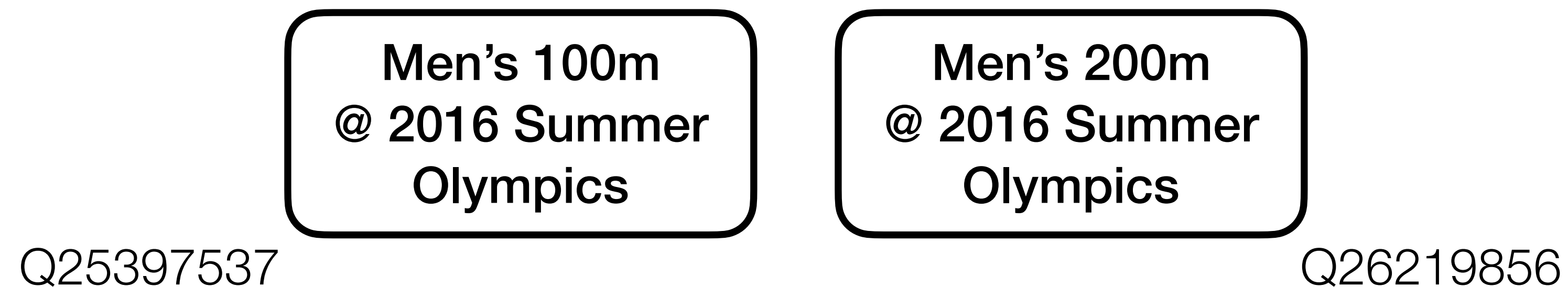
Illustration:
hierarchy of
Wikidata Events



Structure in Wikidata

**Illustration:
hierarchy of
Wikidata Events**

**For simplicity, we
focus only on
atomic events**



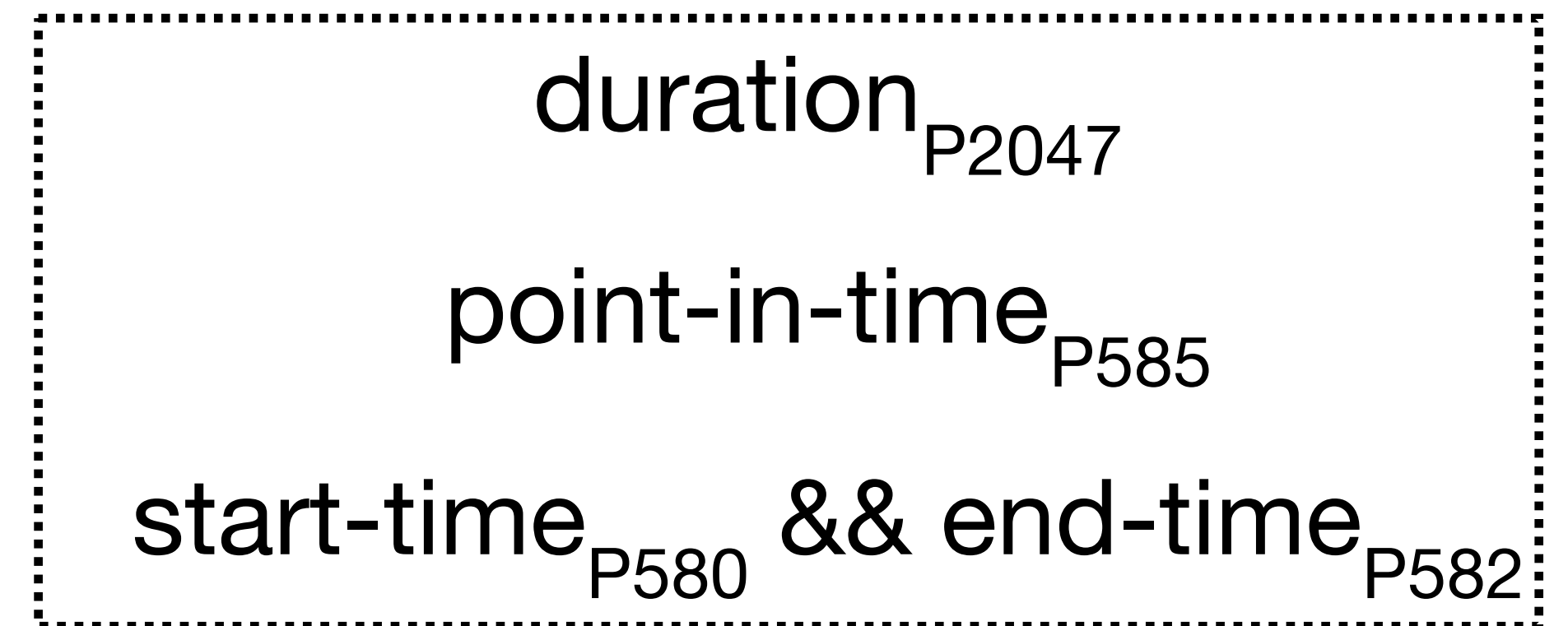
Event Dictionary

Event Dictionary

- Wikidata items with temporal and spatial properties are *potential* events

Event Dictionary

- Wikidata items with temporal and spatial properties are *potential* events



Temporal properties

Event Dictionary

- Wikidata items with temporal and spatial properties are *potential* events

duration_{P2047}
point-in-time_{P585}
start-time_{P580} && end-time_{P582}

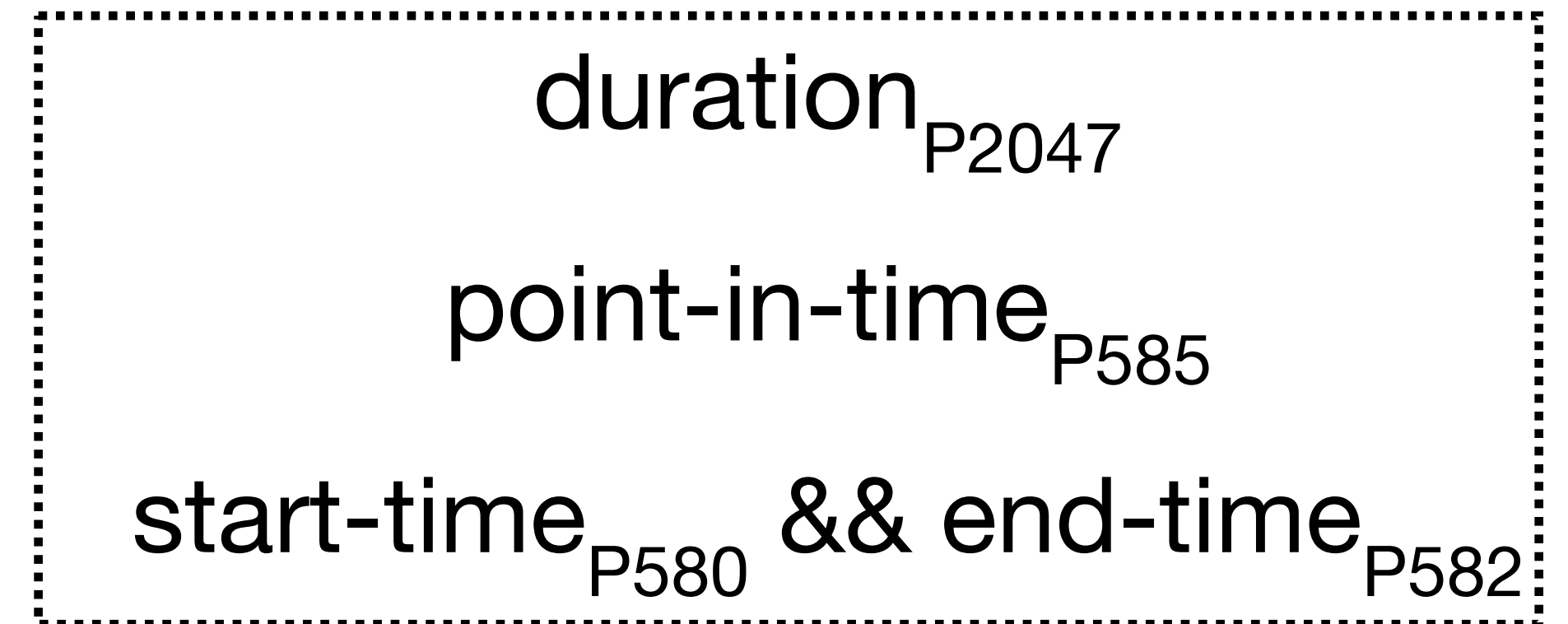
Temporal properties

location_{P276}
coordinate-location_{P625}

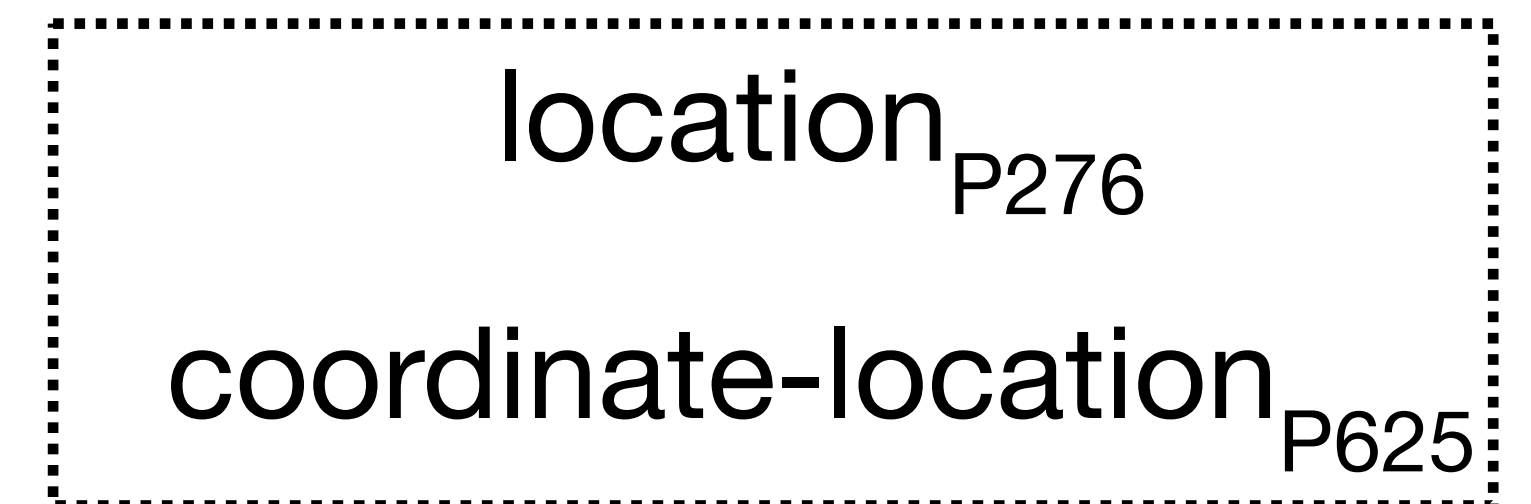
Spatial properties

Event Dictionary

- Wikidata items with temporal and spatial properties are *potential* events
- Correspond to state changes and grounded in spatio-temporal contexts



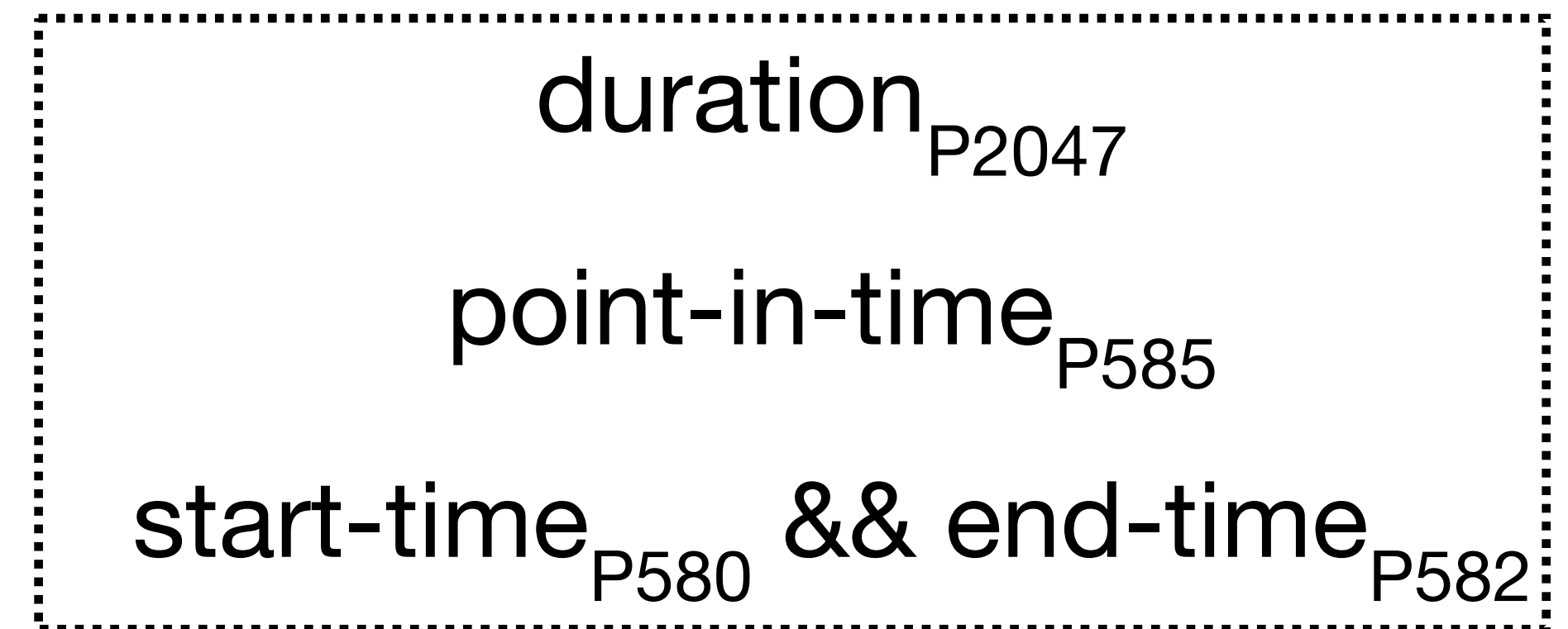
Temporal properties



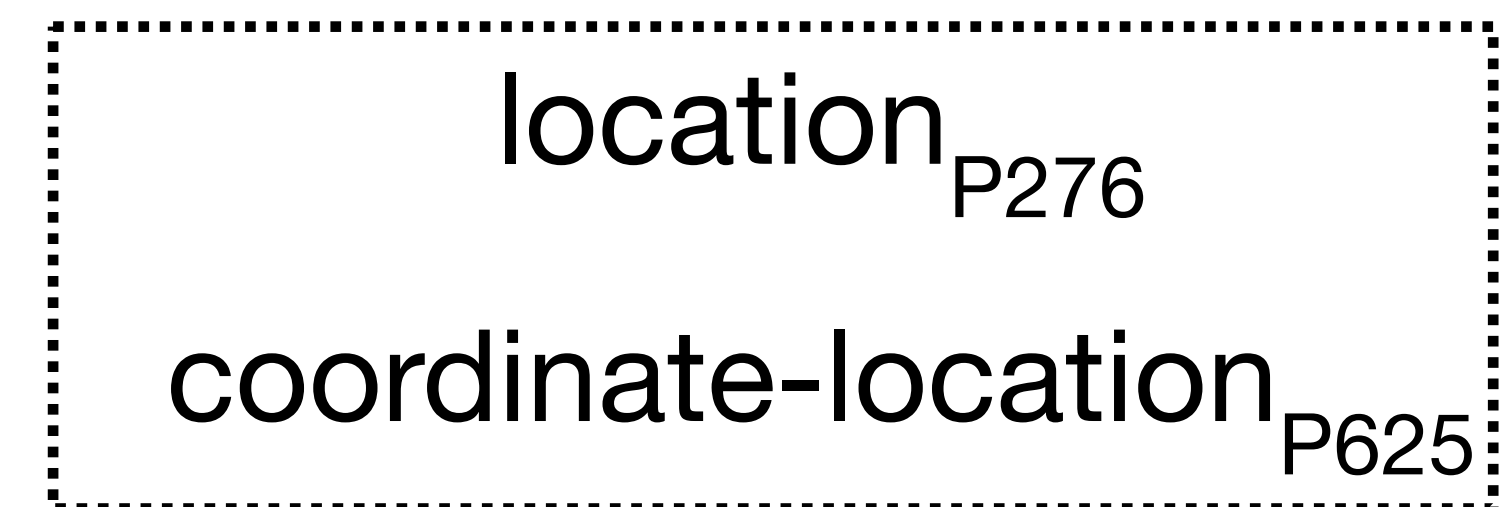
Spatial properties

Event Dictionary

- Wikidata items with temporal and spatial properties are *potential* events
- Correspond to state changes and grounded in spatio-temporal contexts
- Categorized as eventive nouns (Weischedel et al. 2013)



Temporal properties



Spatial properties

Mentions

Mentions

- For Wikidata events, we collect mentions from Wikipedia and Wikinews

Mentions

- For Wikidata events, we collect mentions from Wikipedia and Wikinews
- Two steps,

Mentions

- For Wikidata events, we collect mentions from Wikipedia and Wikinews
- Two steps,
 1. Identify corresponding language Wikipedia event pages

Mentions

- For Wikidata events, we collect mentions from Wikipedia and Wikinews
- Two steps,
 1. Identify corresponding language Wikipedia event pages
 2. Collect inlinks from other Wikipedia/Wikinews pages to event pages

Event Linking: Example

Mention from language Wikipedia

(frwiki) Aliaksandra Herasimenia

Aliaksandra Herasimenia est une nageuse biélorusse en activité spécialiste des épreuves de sprint en nage libre et en dos. ... Multiple médaillée au niveau planétaire et continental, elle décroche en 2010 son premier titre international majeur lors des **Championnats d'Europe** de Budapest, sur dos.

(enwiki) Viktor Minibaev

Minibaev's first major international medal came in the men's synchronized 10 metre platform event at the **2010 European Championships**.

(dewiki) Nóra Barta

Bei Schwimmeuropameisterschaften gewann sie insgesamt drei Medaillen. 2006 und **2010** gewann sie in ihrer Heimatstadt Budapest jeweils Bronze vom 3 m-Brett, 2008 holte sie in Eindhoven Silber vom 1 m-Brett.

Event Linking: Example

Mention from language Wikipedia

(frwiki) Aliaksandra Herasimenia

Aliaksandra Herasimenia est une nageuse biélorusse en activité spécialiste des épreuves de sprint en nage libre et en dos. ... Multiple médaillée au niveau planétaire et continental, elle décroche en 2010 son premier titre international majeur lors des **Championnats d'Europe** de Budapest, sur dos.

(enwiki) Viktor Minibaev

Minibaev's first major international medal came in the men's synchronized 10 metre platform event at the **2010 European Championships**.

(dewiki) Nóra Barta

Bei Schwimmeuropameisterschaften gewann sie insgesamt drei Medaillen. 2006 und **2010** gewann sie in ihrer Heimatstadt Budapest jeweils Bronze vom 3 m-Brett, 2008 holte sie in Eindhoven Silber vom 1 m-Brett.

Q830917



2010 European Aquatics Championships (Q830917)

Swimming event of LEN in Budapest (Hungary)

Event Linking: Example

Mention from language Wikipedia

(frwiki) Aliaksandra Herasimienia

Aliaksandra Herasimienia est une nageuse biélorusse en activité spécialiste des épreuves de sprint en nage libre et en dos. ... Multiple médaillée au niveau planétaire et continental, elle décroche en 2010 son premier titre international majeur lors des **Championnats d'Europe** de Budapest, sur dos.

(enwiki) Viktor Minibaev

Minibaev's first major international medal came in the men's synchronized 10 metre platform event at the **2010 European Championships**.

(dewiki) Nóra Barta

Bei Schwimmeuropameisterschaften gewann sie insgesamt drei Medaillen. 2006 und **2010** gewann sie in ihrer Heimatstadt Budapest jeweils Bronze vom 3 m-Brett, 2008 holte sie in Eindhoven Silber vom 1 m-Brett.

Event description from language Wikipedia

(frwiki) Championnats d'Europe de natation 2010

La des Championnats d'Europe de natation se tient du 4 au à Budapest en Hongrie. C'est la quatrième fois que la capitale hongroise accueille l'événement bisannuel organisé par la Ligue européenne de natation après les éditions 1926, 1958 et 2006.

(enwiki) 2010 European Aquatics Championships

The 2010 European Aquatics Championships were held from 4–15 August 2010 in Budapest and Balatonfüred, Hungary. It was the fourth time that the city of Budapest hosts this event after 1926, 1958 and 2006. Events in swimming, diving, synchronised swimming (synchro) and open water swimming were scheduled.

(dewiki) Schwimmeuropameisterschaften 2010

Bei Schwimmeuropameisterschaften gewann sie insgesamt drei Medaillen. 2006 und **2010** gewann sie in ihrer Heimatstadt Budapest jeweils Bronze vom 3 m-Brett, 2008 holte sie in Eindhoven Silber vom 1 m-Brett.

Q830917



Event Linking: Example

Mention from language Wikipedia

(frwiki) Aliaksandra Herasimienia

Aliaksandra Herasimienia est une nageuse biélorusse en activité spécialiste des épreuves de sprint en nage libre et en dos. ... Multiple médaillée au niveau planétaire et continental, elle décroche en 2010 son premier titre international majeur lors des **Championnats d'Europe** de Budapest, sur dos.

(enwiki) Viktor Minibaev

Minibaev's first major international medal came in the men's synchronized 10 metre platform event at the **2010 European Championships**.

(dewiki) Nóra Barta

Bei Schwimmeuropameisterschaften gewann sie insgesamt drei Medaillen. 2006 und **2010** gewann sie in ihrer Heimatstadt Budapest jeweils Bronze vom 3 m-Brett, 2008 holte sie in Eindhoven Silber vom 1 m-Brett.

Event description from language Wikipedia

(frwiki) Championnats d'Europe de natation 2010

La des Championnats d'Europe de natation se tient du 4 au à Budapest en Hongrie. C'est la quatrième fois que la capitale hongroise accueille l'événement bisannuel organisé par la Ligue européenne de natation après les éditions 1926, 1958 et 2006.

(enwiki) 2010 European Aquatics Championships

The 2010 European Aquatics Championships were held from 4–15 August 2010 in Budapest and Balatonfüred, Hungary. It was the fourth time that the city of Budapest hosts this event after 1926, 1958 and 2006. Events in swimming, diving, synchronised swimming (synchro) and open water swimming were scheduled.

(dewiki) Schwimmeuropameisterschaften 2010

Bei Schwimmeuropameisterschaften gewann sie insgesamt drei Medaillen. 2006 und **2010** gewann sie in ihrer Heimatstadt Budapest jeweils Bronze vom 3 m-Brett, 2008 holte sie in Eindhoven Silber vom 1 m-Brett.

Q830917



Multilingual task

Event Linking: Example

Mention from language Wikipedia

(frwiki) Aliaksandra Herasimienia

Aliaksandra Herasimienia est une nageuse biélorusse en activité spécialiste des épreuves de sprint en nage libre et en dos. ... Multiple médaillée au niveau planétaire et continental, elle décroche en 2010 son premier titre international majeur lors des **Championnats d'Europe** de Budapest, sur dos.

(enwiki) Viktor Minibaev

Minibaev's first major international medal came in the men's synchronized 10 metre platform event at the **2010 European Championships**.

(dewiki) Nóra Barta

Bei Schwimmeuropameisterschaften gewann sie insgesamt drei Medaillen. 2006 und **2010** gewann sie in ihrer Heimatstadt Budapest jeweils Bronze vom 3 m-Brett, 2008 holte sie in Eindhoven Silber vom 1 m-Brett.

Event description from language Wikipedia

(frwiki) Championnats d'Europe de natation 2010

La des Championnats d'Europe de natation se tient du 4 au à Budapest en Hongrie. C'est la quatrième fois que la capitale hongroise accueille l'événement bisannuel organisé par la Ligue européenne de natation après les éditions 1926, 1958 et 2006.

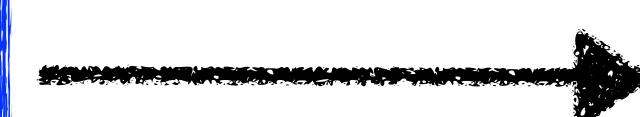
(enwiki) 2010 European Aquatics Championships

The 2010 European Aquatics Championships were held from 4–15 August 2010 in Budapest and Balatonfüred, Hungary. It was the fourth time that the city of Budapest hosts this event after 1926, 1958 and 2006. Events in swimming, diving, synchronised swimming (synchro) and open water swimming were scheduled.

(dewiki) Schwimmeuropameisterschaften 2010

Bei Schwimmeuropameisterschaften gewann sie insgesamt drei Medaillen. 2006 und **2010** gewann sie in ihrer Heimatstadt Budapest jeweils Bronze vom 3 m-Brett, 2008 holte sie in Eindhoven Silber vom 1 m-Brett.

Q830917



Crosslingual task

Dataset Stats: XLEL-WD

Dataset Stats: XLEL-WD

- Disjoint event sequences in train/eval

<i>Wikipedia</i>	Train	Dev	Test	Total
Events	8653	1090	1204	10947
Event Sequences	6758	844	846	8448
Mentions	1.44M	165K	190K	1.8M
Languages	44	44	44	44

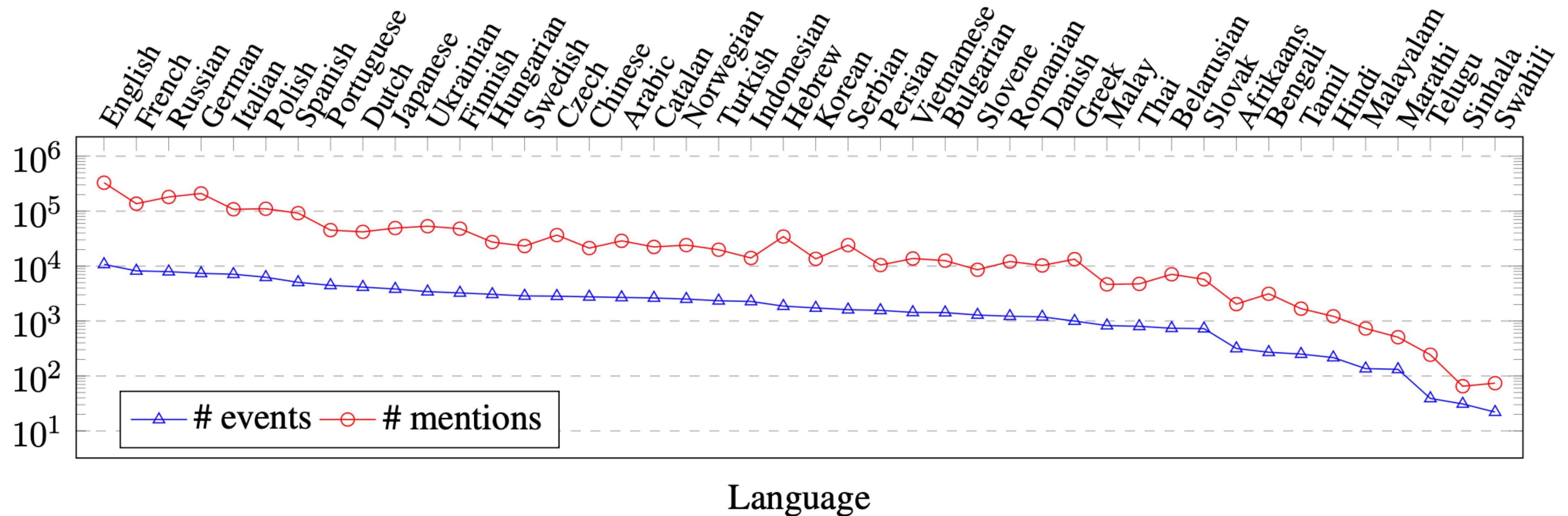
Dataset Stats: XLEL-WD

- Disjoint event sequences in train/eval
- Wikinews-based evaluation sets
 - Cross-domain (unseen domain)
 - Zero-shot (unseen domain, events)

<i>Wikipedia</i>	Train	Dev	Test	Total
Events	8653	1090	1204	10947
Event Sequences	6758	844	846	8448
Mentions	1.44M	165K	190K	1.8M
Languages	44	44	44	44

<i>Wikinews</i>	Cross-domain	Zero-shot
Events	802	149
Mentions	2562	437
Languages	27	21

Language Distribution



Event Linker: Retrieve+Rank

Event Linker: Retrieve+Rank

Retrieve

- BM25+ (Lv and Zhai, 2011)
- (mBERT, XLM-RoBERTa)
biencoder (BLINK; Wu et al., 2020)

[Lower-bounding term frequency normalization](#) (Lv and Zhai, CIKM 2011)

[Scalable Zero-shot Entity Linking with Dense Entity Retrieval](#) (Wu et al., EMNLP 2020)

Event Linker: Retrieve+Rank

Retrieve

- BM25+ (Lv and Zhai, 2011)
- (mBERT, XLM-RoBERTa) biencoder (BLINK; Wu et al., 2020)

Rank

- (mBERT, XLM-RoBERTa) crossencoder (BLINK; Wu et al., 2020)

[Lower-bounding term frequency normalization](#) (Lv and Zhai, CIKM 2011)

[Scalable Zero-shot Entity Linking with Dense Entity Retrieval](#) (Wu et al., EMNLP 2020)

Biencoder: Retrieval

- Given an input context, retrieve top-k event candidates

Biencoder: Retrieval

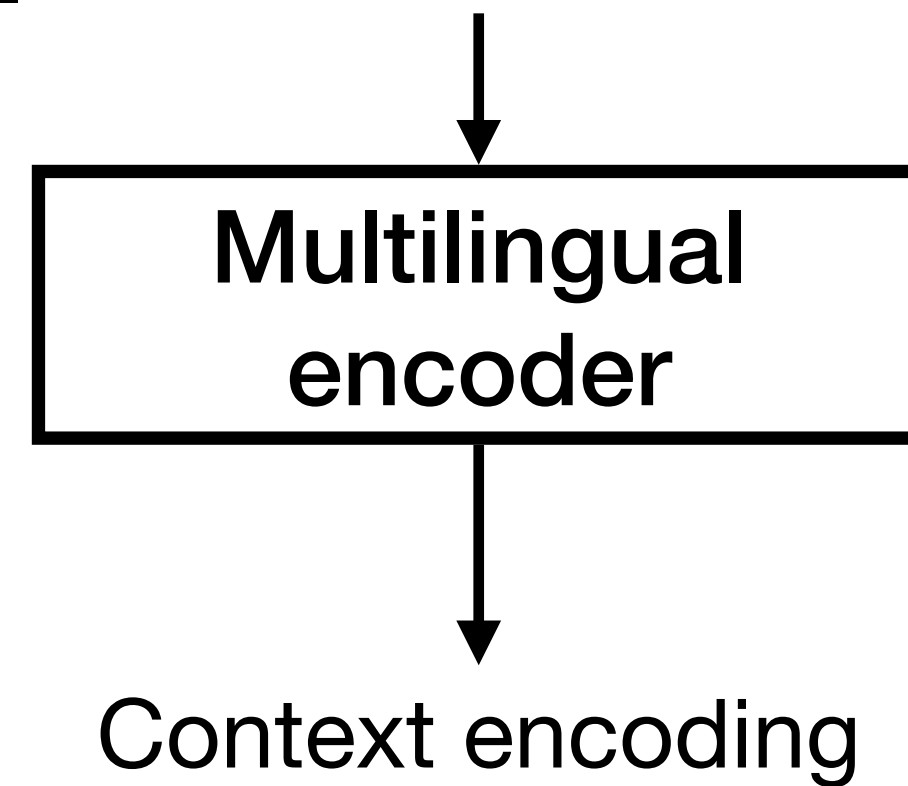
- Given an input context, retrieve top-k event candidates

[CLS] left context [MENTION_START] mention [MENTION_END] right context [SEP]

Biencoder: Retrieval

- Given an input context, retrieve top-k event candidates

[CLS] left context [MENTION_START] mention [MENTION_END] right context [SEP]

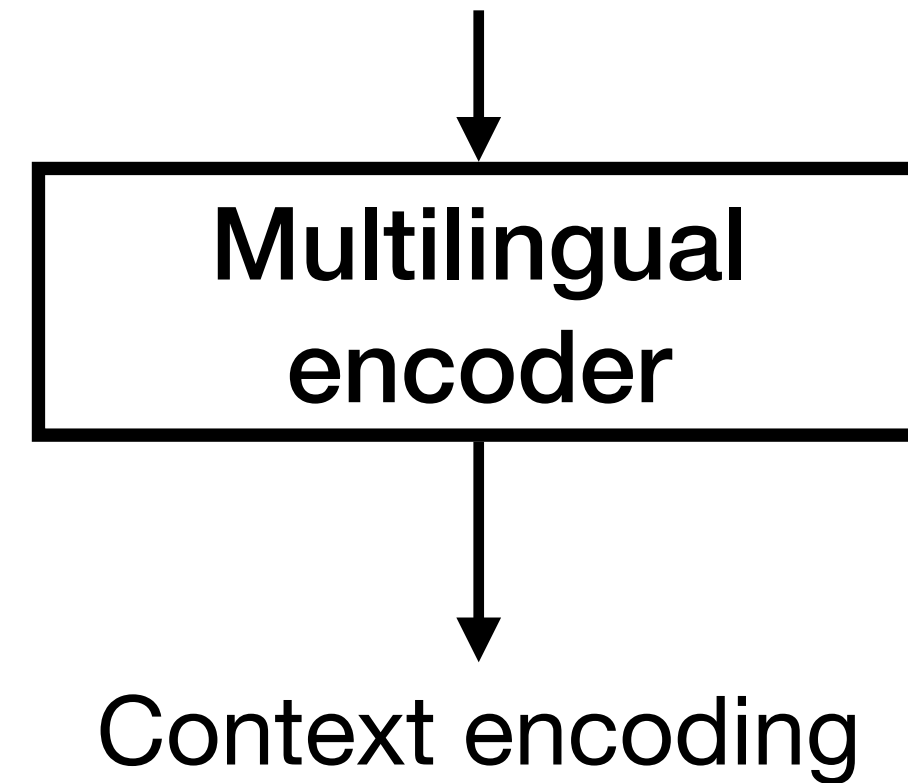


Biencoder: Retrieval

- Given an input context, retrieve top-k event candidates

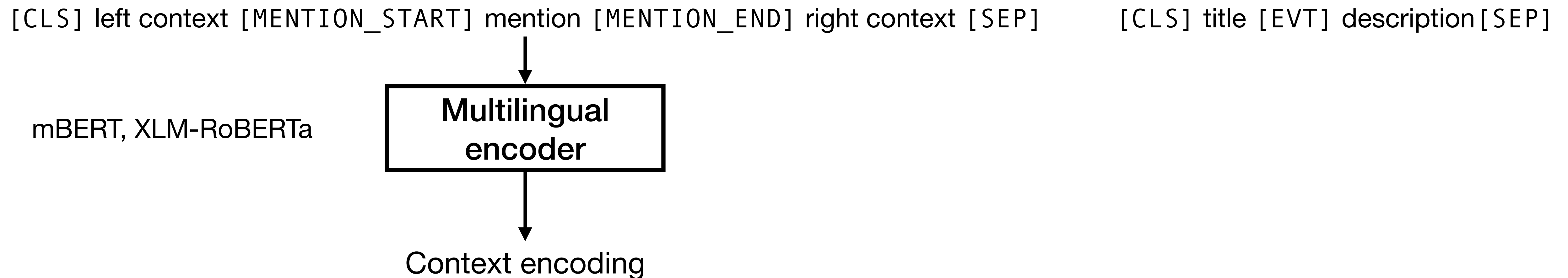
[CLS] left context [MENTION_START] mention [MENTION_END] right context [SEP]

mBERT, XLM-RoBERTa



Biencoder: Retrieval

- Given an input context, retrieve top-k event candidates

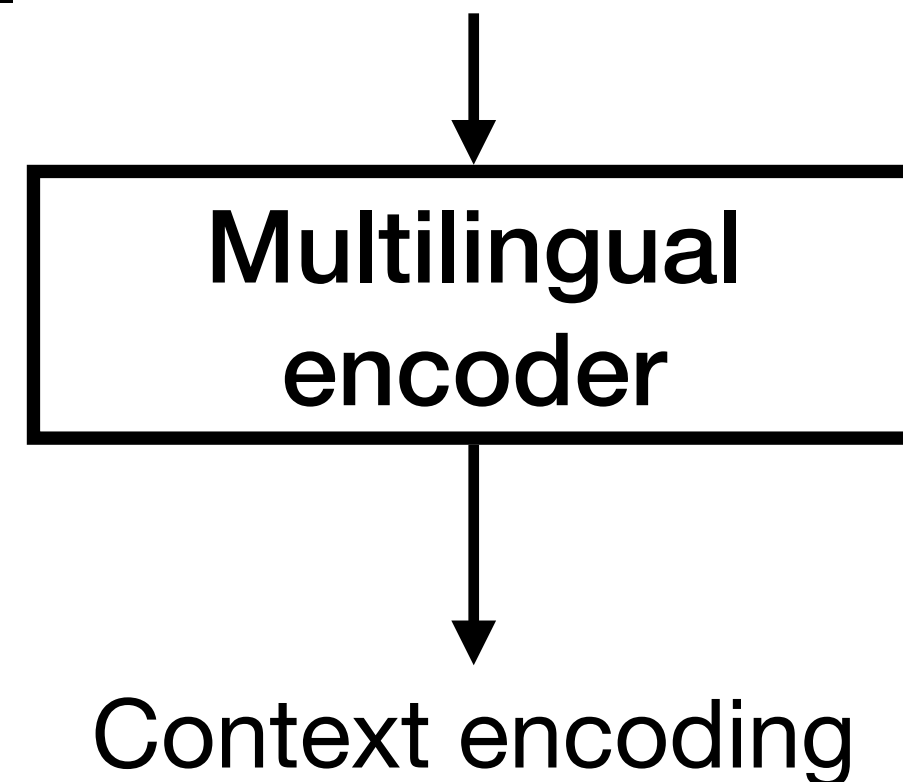


Biencoder: Retrieval

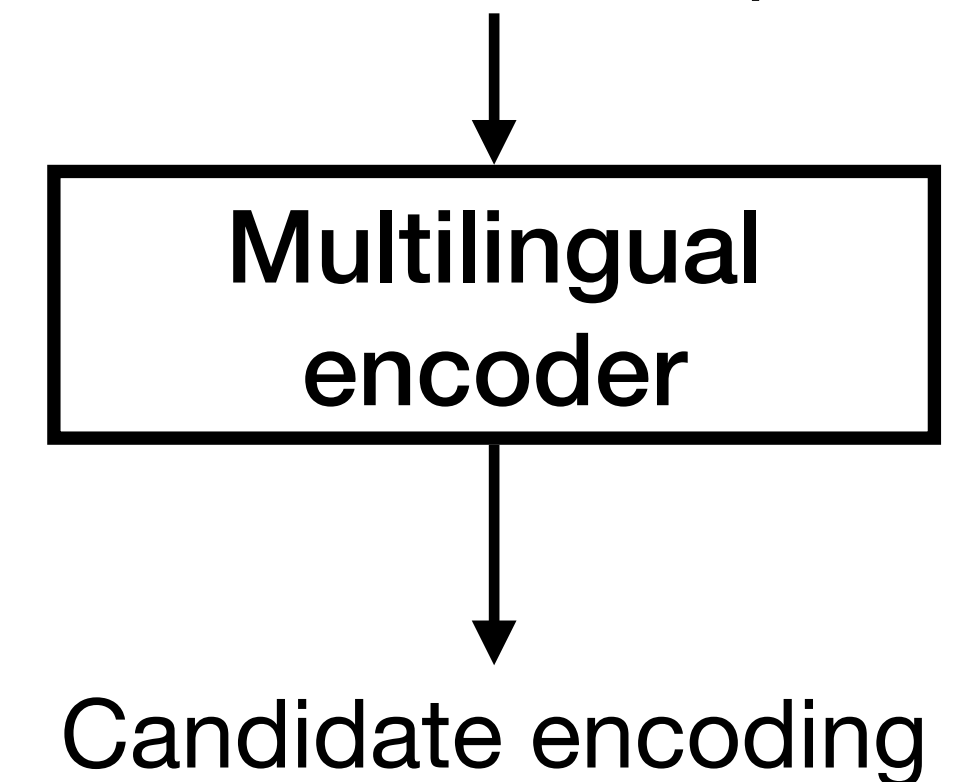
- Given an input context, retrieve top-k event candidates

[CLS] left context [MENTION_START] mention [MENTION_END] right context [SEP]

mBERT, XLM-RoBERTa

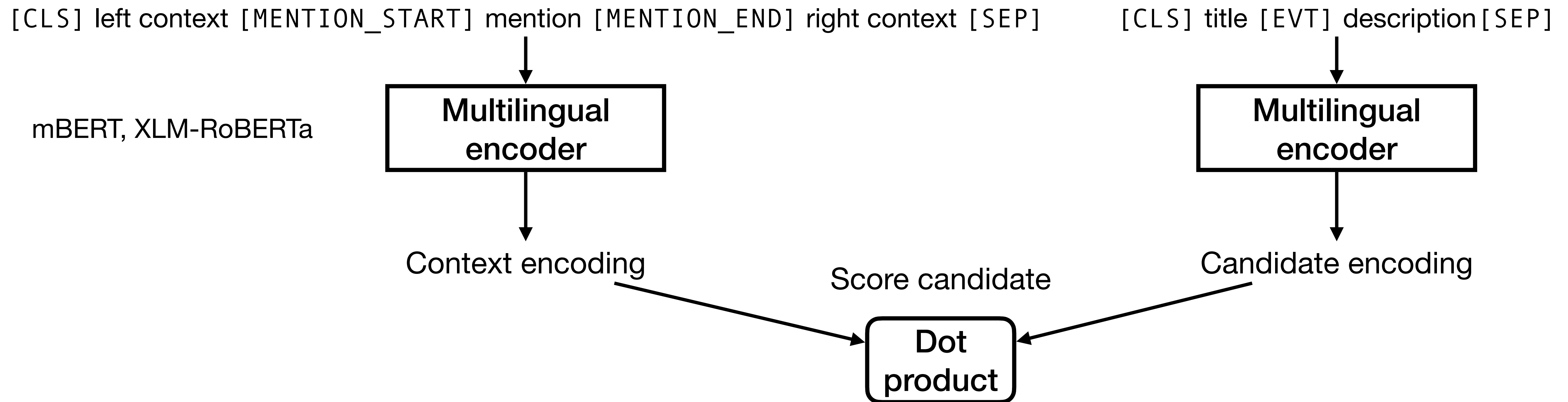


[CLS] title [EVT] description [SEP]



Biencoder: Retrieval

- Given an input context, retrieve top-k event candidates



Crossencoder: Rank

- Given an input context and top-k retrieved event candidates, identify gold label

Crossencoder: Rank

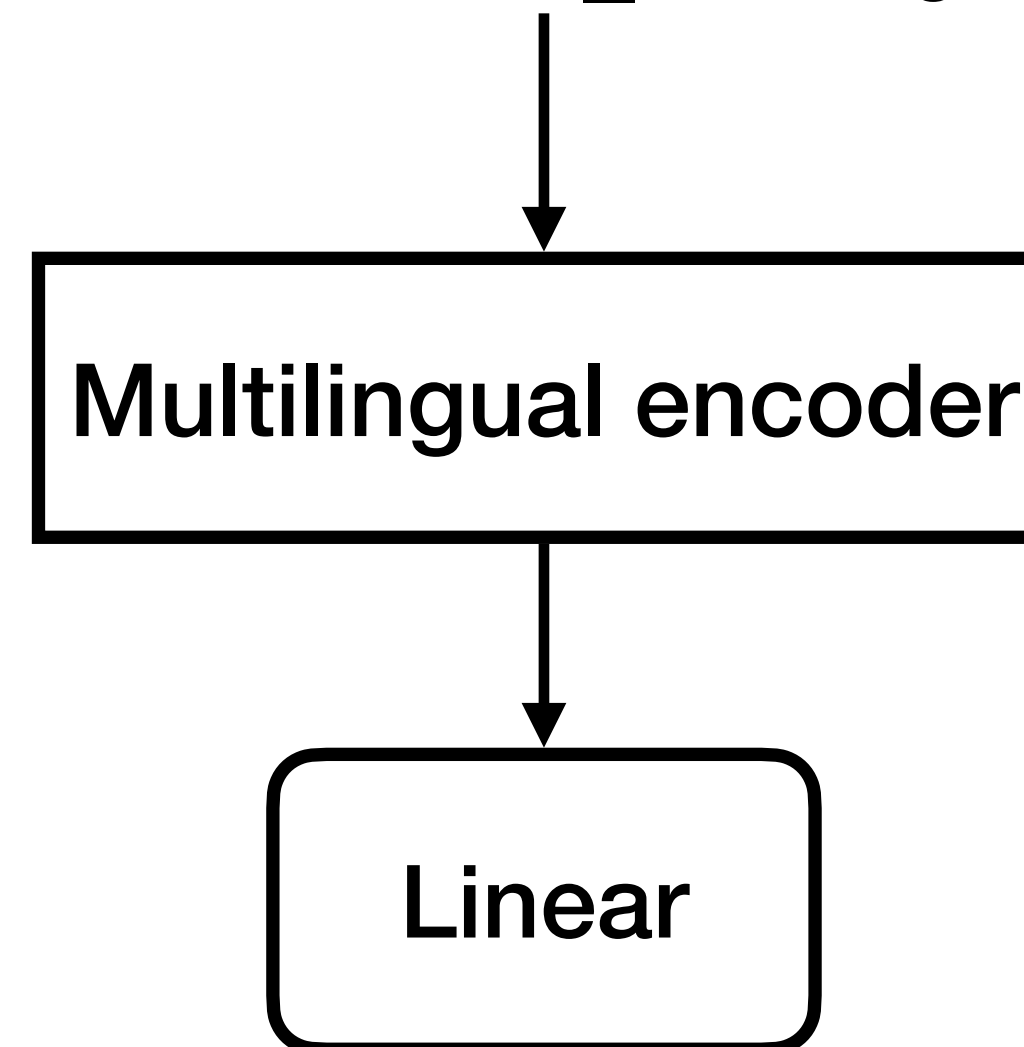
- Given an input context and top-k retrieved event candidates, identify gold label

[CLS] left context [MENTION_START] mention [MENTION_END] right context [SEP] title [EVT] description[SEP]

Crossencoder: Rank

- Given an input context and top-k retrieved event candidates, identify gold label

[CLS] left context [MENTION_START] mention [MENTION_END] right context [SEP] title [EVT] description [SEP]



Results: Wikipedia

- Crosslingual task is harder than multilingual
- Bi- and cross-encoders outperform BM25+

<i>Wikipedia</i>	Multilingual		Crosslingual	
Model	Dev	Test	Dev	Test
BM25+	53.4	50.1	-	-
mBERT-bi	84.7	84.6	83.2	83.9
XLM-R-bi	84.5	84.3	79.3	79.1
mBERT-cross	89.8	89.3	81.3	73.9
XLM-R-cross	88.8	87.3	81.0	75.6

Results: Wikinews

- Domain transfer is challenging
- Zero-shot transfer further lowers accuracy
- Including meta-information (date, title) improves accuracy by 4-12%

Model	Multilingual		Crosslingual	
	Cross-domain	Zero-shot	Cross-domain	Zero-shot
BM25+	53.5	58.6	-	-
mBERT-bi	81.2	76.7	85.4	78.0
XLM-R-bi	82.2	76.7	82.6	76.4
mBERT-cross	90.1	84.4	89.3	76.2
XLM-R-cross	89.7	84.4	88.9	76.0

Analysis

mention (+ context)

At the 2000 Summer Olympics in Sydney, Sitnikov competed only in two swimming events. ... Three days later, in the **100 m freestyle**, Sitnikov placed fifty-third on the morning prelims. ...

Gold

Swimming at the 2000 Summer Olympics –
Men's 100 metre freestyle

Predicted

Swimming at the 2008 Summer Olympics –
Men's 100 metre freestyle

Analysis

mention (+ context)

... war er bei der Oscarverleihung 1935 erstmals für einen Oscar für den besten animierten Kurzfilm nominiert. Eine weitere Nominierung in dieser Kategorie erhielt er **1938** für "The Little Match Girl" (1937).

*... he was nominated for the Oscar for the best animated short film for the first time in 1935. He received another nomination in this category in **1938** for The Little Match Girl (1937).*

Gold

The 10th Academy Awards were originally scheduled ... but due to ... were held on March 10, 1938, ..

Predicted

The 9th Academy Awards were held on March 4, 1937, ...

Analysis

mention (+ context)

...攝津號與其姐妹艦河號於1914年10月至11月間參與了青島戰役的最後階段...

...Settsu and her sister Hewa participated in the final phase of the Tsingtao campaign between October and November 1914...

Siege of Tsingtao: The siege of Tsingtao (or Tsingtau) was the attack on the German port of Tsingtao (now Qingdao) ...

Predicted

Gold

青島戰役(,)是第一次世界大戰初期日本進攻國膠州灣殖民地及其首府青島的一場戰役，也是唯一的一場戰役。

Battle of the Yellow Sea

The Battle of Qingdao (,) was a battle in the Jiaozhou Bay colony and its capital Qingdao in the early World War I, and it was also the only battle.

Analysis

mention (+ context)

Ivanova won the silver medal at the 1978 World Junior Championships. She made her senior World debut at the **1979 World Championships**, finishing 18th. Ivanova was 16th at the 1980 Winter Olympics.

Gold

1979 World Figure Skating Championships

Predicted

FIBT World Championships 1979

Summary

- Task: multilingual linking of events to KB
- XLEL-WD: dataset compiled from Wikidata, Wikinews, and Wikipedia
- Linker: multilingual variants of BLINK for events



Questions?

Code: [adithya7/xlel-wd](#) 

Mentions: [adithya7/xlel_wd](#) ( datasets)

Dictionary: [adithya7/xlel_wd_dictionary](#) ( datasets)