



# Event Series Completion

Jakob Krude, Ayan Banerjee

September 2023

Knowledge Graph Lab I SS23

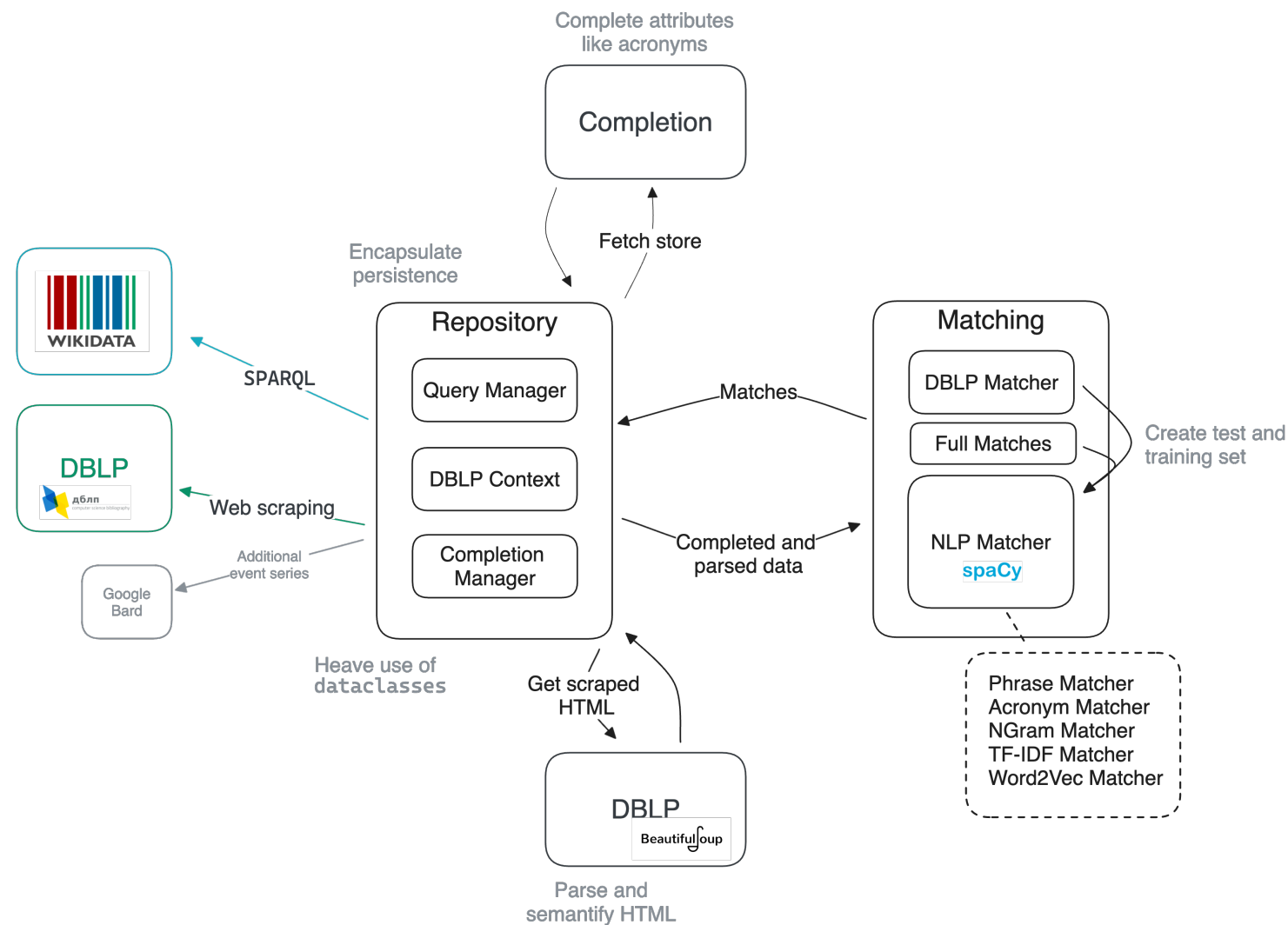
# Outline

---

1. Task description
2. Project Architecture
3. Matching through DBLP
  1. Data Insights
4. Matching Events directly
  1. Matching Results
5. Outlook



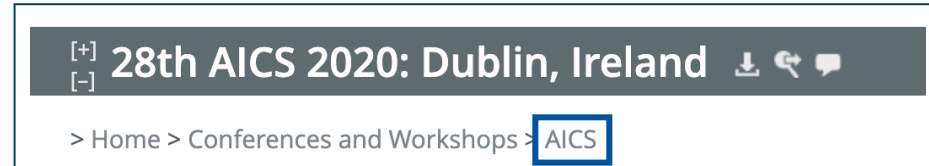
# Project Architecture



# 1. Approach: Matching Through DBLP

# 1. Approach

## Why do we use DBLP



# 1. Approach

## Its complicated

[+] 8. iStar@RE 2015: Ottawa, Canada [download] [share] [comment]

> Home > Conferences and Workshops > iStar  
> Home > Conferences and Workshops > RE

[+] International i\* Workshop (iStar) [download] [share] [comment]

> Home > Conferences and Workshops

[+] IEEE International Requirements Engineering Conference (RE) [home] [download] [share] [comment]  
IEEE International Conference on Requirements Engineering (ICRE)

> Home > Conferences and Workshops

☰ ▼ Dagstuhl



# 1. Approach

## There is more

### [+] 1. SocInf@IJCAI 2015: Buenos Aires, Argentina [-]

> Home > Conferences and Workshops > IJCAI

☰ ▼ Dagstuhl

### [+] International Joint Conference on Artificial Intelligence (IJCAI) [-]

> Home > Conferences and Workshops

☰ ▼ Dagstuhl

#### [+] Venue Information [-]

#### [+] Venue statistics

- *access*: some or all publications **openly available** 8
  - *has part*: Workshop on Artificial Intelligence in Affective Computing (AffComp)
  - *has part*: International Workshop AI and Feedback (AInF)
  - *has part*: Workshop on Computer Games (CGW)
  - *has part*: International Workshop on Defeasible and Ampliative Reasoning (DARE)
  - *has part*: International Workshop "What can FCA do for Artificial Intelligence?" (FCA4AI)
  - *has part*: Workshop on Fuzzy Logic in AI (FLinAI)
  - *has part*: Workshop on Information Integration on the Web (IIWeb)
  - *has part*: Workshop on Intelligent Techniques for Web Personalization & Recommender Systems (ITWP)
  - *has part*: Workshop on Discovering Meaning On the Go in Large Heterogeneous Data (LHD)
  - *has part*: Workshop on Machine Learning for Interactive Systems (MLIS)
  - *has part*: Workshop on Ontology Learning (OL)
  - *has part*: International Workshop on Social Influence Analysis (SocInf)
  - *has part* (2016, 2018, 2022): International Workshop on Statistical Relational AI (StarAI)
  - *has part*: Workshop on Ubiquitous Data Mining (UDM)
- show less*



# 1. Approach

## Wikidata

28th Irish Conference on Artificial Intelligence and Cognitive Science

(Q113582107)

DBLP event ID



conf/aics/aics2020

▼ 0 references



## DBLP

[+] 28th AICS 2020: Dublin, Ireland [-] ⬇️ ⬅️ 💬

> Home > Conferences and Workshops > AICS



Irish Conference on Artificial Intelligence and Cognitive Science (Q105700366)



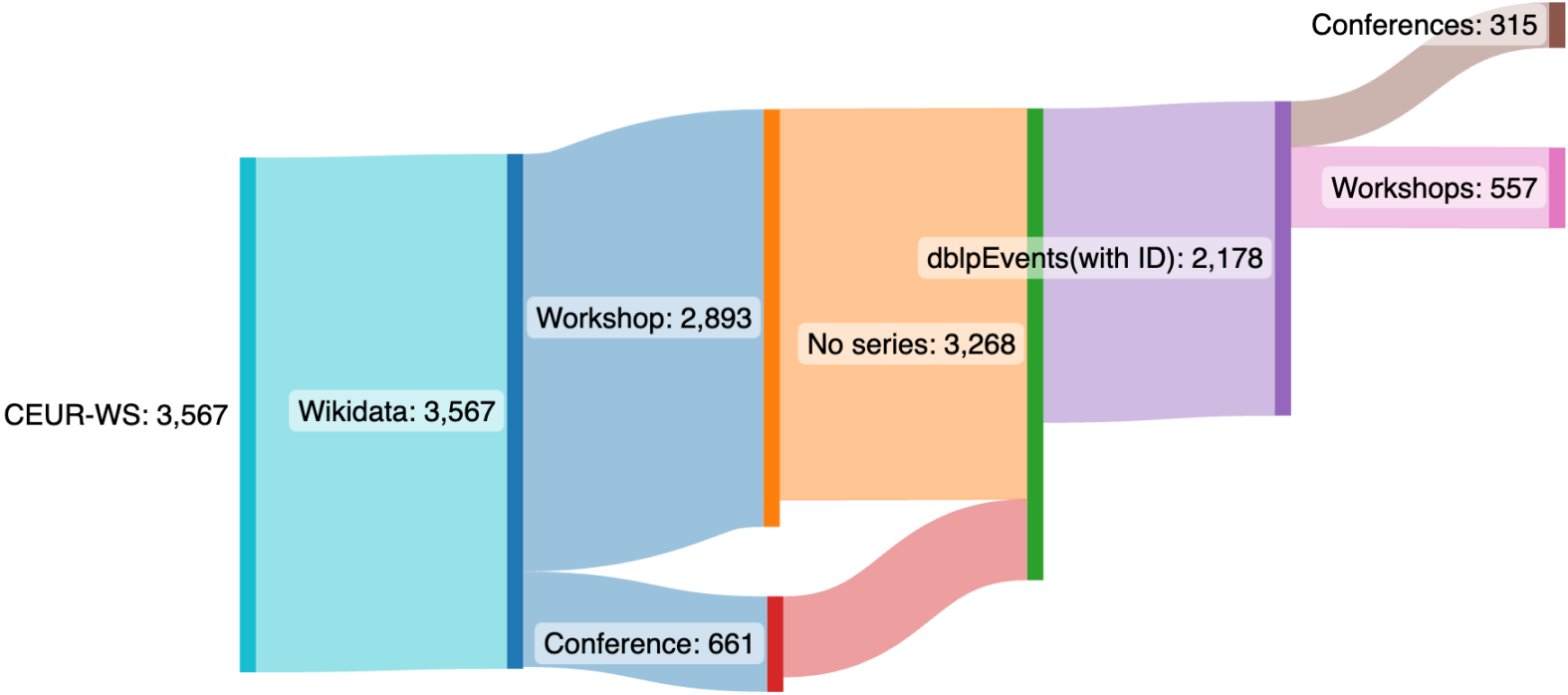
[+] Irish Conference on Artificial Intelligence and Cognitive Science (AICS) [-] ⬇️ ⬅️ 💬

> Home > Conferences and Workshops

☰ ▼ Dagstuhl

AICS 2021: Dublin, Ireland

■ ⬇️ ⬅️ 💬 Arjun Pakrashi, Ellen Rushe, Mehran Hossein Zadeh Bazargani, Brian Mac Namee: **The 29th Irish Conference on Artificial Intelligence and Cognitive Science 2021, Dublin, Republic of Ireland, December 9-10, 2021.** CEUR Workshop Proceedings 3105, CEUR-WS.org 2022 [contents]



## **2. Approach: Matching Events Directly**

## 2. Approach

---

28th Irish Conference on Artificial Intelligence and Cognitive Science

(Q113582107)

---



Irish Conference on Artificial Intelligence and Cognitive Science (Q105700366)

---

Directly match event to series

- Title
- Label
- Acronym

Using multiple Matching Algorithms

- Full Match
- Phrase Match
- Acronym Match
- N-Grams
- Tf-Idf
- Word2Vec

## Full Match

- The idea here is to fully match the content of *title/label* from event-to-event series .
  - Title from the Wikidata event to event series

### Workshop on Linked Data on the Web (Q113649749)

---

academic workshop

### Workshop on Linked Data on the Web (Q105491258)

---

academic workshop series

Linked Data on the Web | LDOW

- CEUR-WS title from ceurspt to event series title.

```
"cvb.number": 2481,  
"cvb.urn": "urn:nbn:de:0074-2481-7",  
"cvb.archive": "http://sunsite.informatik.rwth-aachen.de/ftp/pub/publications/CEUR-WS/Vol-2481.zip",  
"cvb.desc": "?",  
"cvb.hl": "CLiC-it 2019 Italian Conference on Computational Linguistics",  
"cvb.volume_number": "Vol-2481",  
"cvb.ceurpubdate": "2019-10-24",  
"cvb.voltitle": "Italian Conference on Computational Linguistics",  
"cvb.colocated": null,  
"cvb.homepage": "http://clic2019.di.uniba.it/index_en.html",  
"cvb.h3": "Proceedings of the Sixth Italian Conference on Computational Linguistics",
```

### Italian Conference on Computational Linguistics (Q105700367)

---

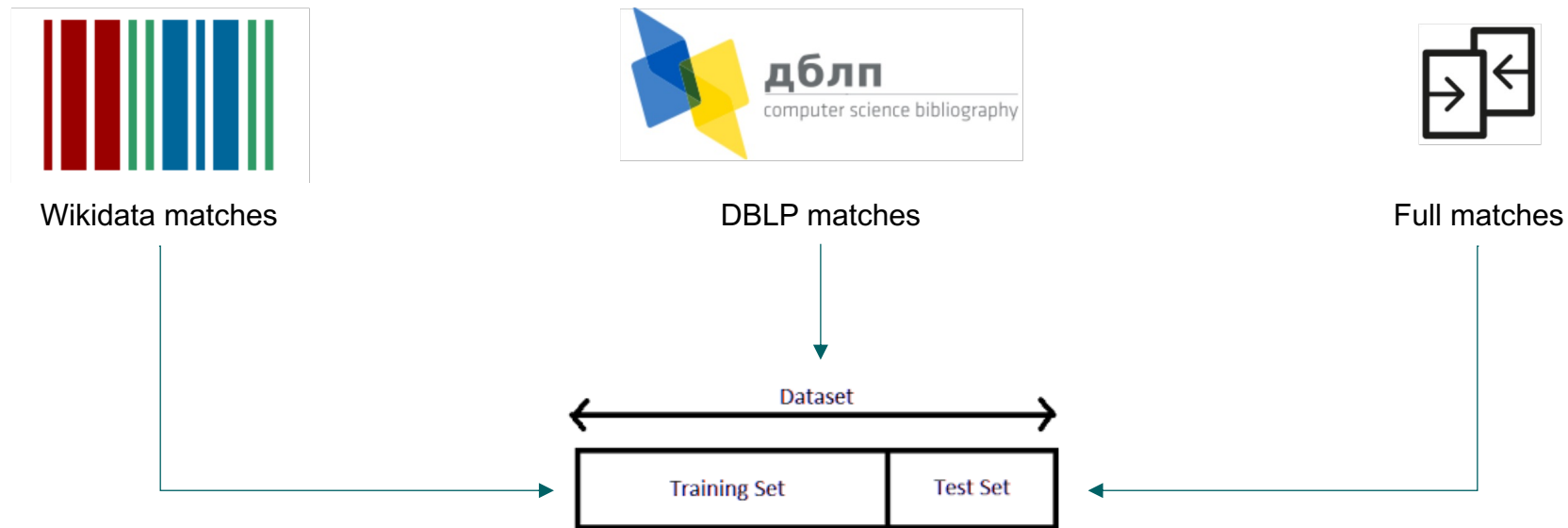
conference series

CLiC-it

# Matching Algorithms

## Creation of training and test set

- To evaluate how our matching algorithms work, we use the existing matches to create our training/test set.






# Matching Algorithms


## Phrase Match

- NLP technique to match phrases using the [spaCy](#) PhraseMatcher.

Event

5th International Conference on Recent Trends and Applications in Computer Science and Information Technology (Q118698035)			
academic conference 			
<a href="#">In more languages</a> <small>Configure</small>			
Language	Label	Description	Also known as
English	5th International Conference on Recent Trends and Applications in Computer Science and Information Technology	academic conference	
German	No label defined	No description defined	
Italian	No label defined	No description defined	
French	No label defined	No description defined	
<a href="#">All entered languages</a>			

Event series

International Conference on Recent Trends and Applications in Computer Science and Information Technology (Q105695983)			
conference series 			
RTA-CSIT			
<a href="#">In more languages</a> <small>Configure</small>			
Language	Label	Description	Also known as
English	International Conference on Recent Trends and Applications in Computer Science and Information Technology	conference series	RTA-CSIT
German	No label defined	No description defined	
Italian	No label defined	No description defined	
French	No label defined	No description defined	
<a href="#">All entered languages</a>			

- Clearly, we can see that the length of the event series label is shorter than the event label.
- Our assumption is event series label are **contained** in event titles

# Matching Algorithms

## Acronym Match

- Abstract matcher on the “acronym” attribute of Wikidata event and event series using **spaCy** PhraseMatcher.

### Event

15th Central European Workshop Services and their Composition (ZEUS 2023) (Q118164550)

academic workshop

In more languages

Configure

Language	Label	Description	Also known as
English	15th Central European Workshop Services and their Composition (ZEUS 2023)	academic workshop	
German	No label defined	No description defined	
Italian	No label defined	No description defined	
French	No label defined	No description defined	

Statements

instance of

academic workshop

1 reference

+ add value

short name

ZEUS 2023 (English)

1 reference

+ add value

Wikipedia

Wikibooks

Wikinews

Wikiquote

Wikisource

Wikiversity

Wikivoyage

Wiktionary

Multilingual sites

### Event series

Central-European Workshop on Services and their Composition (Q105695450)

conference series

ZEUS

In more languages

Configure

Language	Label	Description	Also known as
English	Central-European Workshop on Services and their Composition	conference series	ZEUS
German	No label defined	No description defined	
Italian	No label defined	No description defined	
French	No label defined	No description defined	

Statements

instance of

scientific conference series

1 reference

+ add value

short name

ZEUS (English)

1 reference

+ add value

16

Event Series Completion  
Knowledge Graph Lab | Ayan Banerjee, Jakob krude

RWTHAACHEN  
UNIVERSITY

# Matching Algorithms

## Ngram Match

- NLP technique to match contiguous sequence of n items from text or speech.
  - In our case we split event/event series titles into independent words that acts as *grams*.
    - For Example

1st International Workshop on Knowledge Graph Summarization (Q115053287)



['1st', 'International', 'Workshop', 'Knowledge', 'Graph', 'Summarization']

1st	International	Workshop	Knowledge	Graph	Summarization
-----	---------------	----------	-----------	-------	---------------

International Workshop On Knowledge Graph Construction (Q117087609)



['International', 'Workshop', 'Knowledge', 'Graph', 'Summarization']

International	Workshop	Knowledge	Graph	Summarization
---------------	----------	-----------	-------	---------------

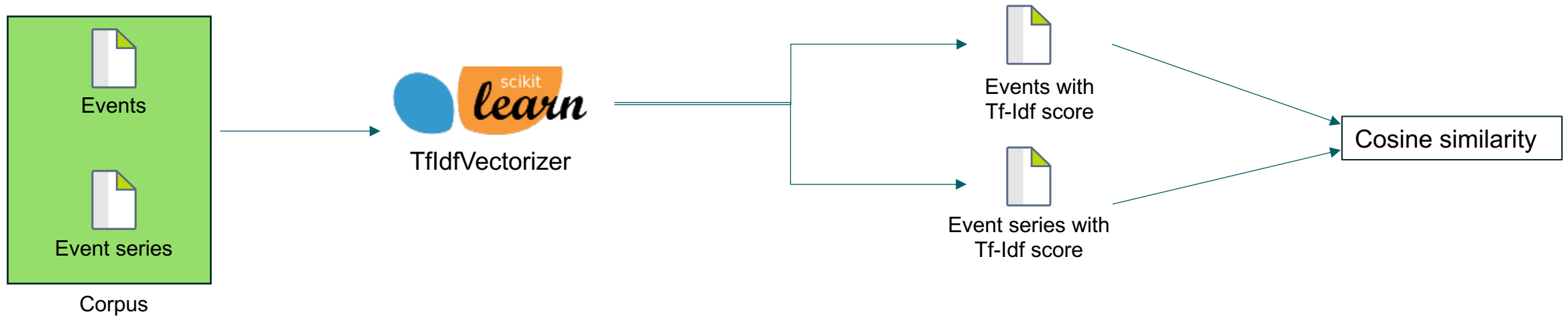
- Similarity is found between them using Jaccard Index.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

## Term frequency-Inverse Document Frequency (TF-IDF) Match

- NLP technique to calculate the relevance of an event/event series title/label.
  - $tf(t, d)$  = count of  $t$  in  $d$  / number of words in  $d$  where
    - $t$  – event/event series title
    - $d$  – documents comprising of the event/event series
  - $idf(t) = N/df(t)$  where
    - $df(t)$  occurrence of  $t$  in documents

$$tf-idf(t, d) = tf(t, d) * idf(t)$$



## Word2Vec Match

- Family of model architectures and optimizations that can be used to learn *word embeddings*.
  - Word embeddings are *vectors* with dense representation in which similar words have a similar encoding.
  - We apply the Continuous Bag of Words(CBOW) method to predict the words.

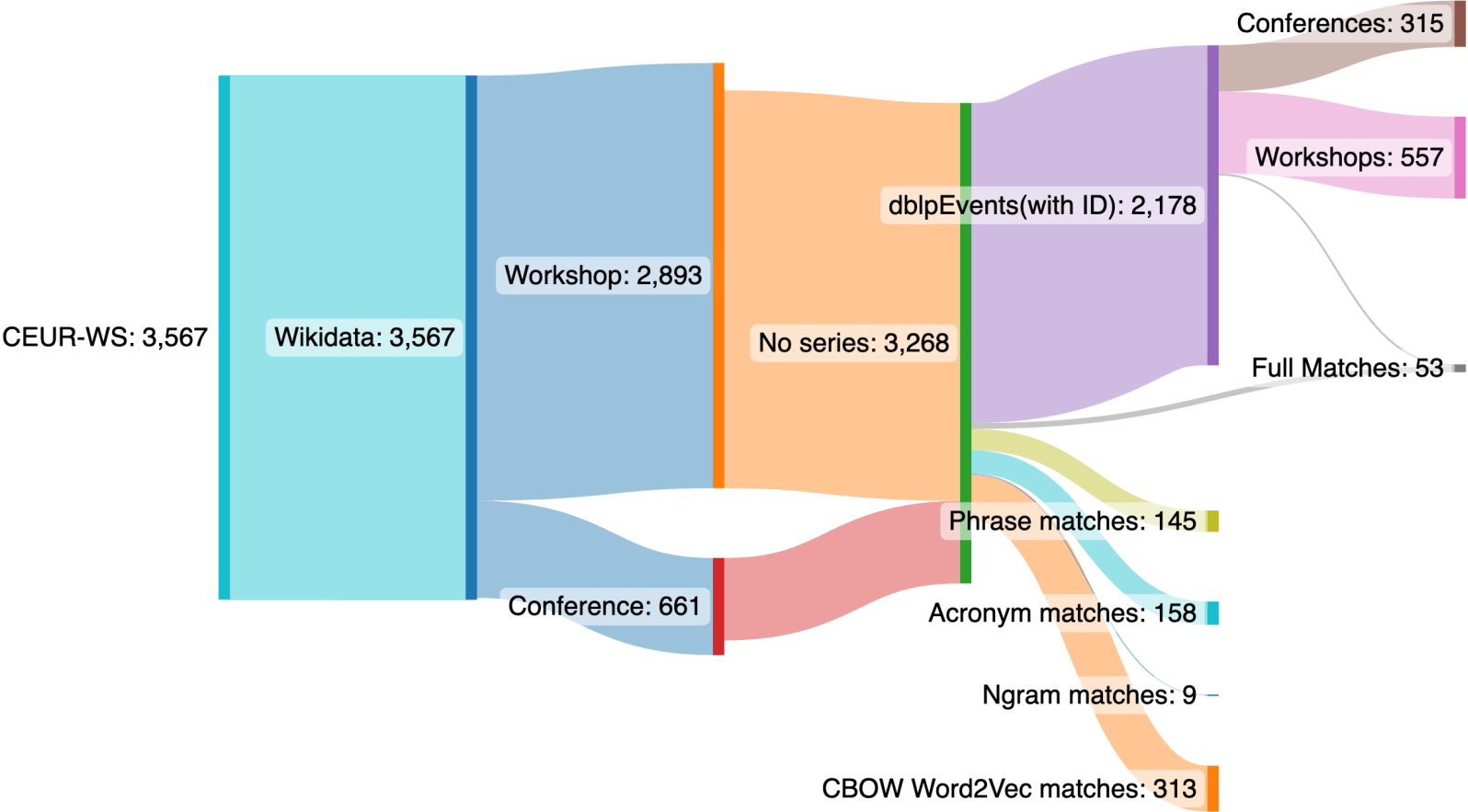


## NLP Matching results

---

Algorithms	Found Matches	Precision	Recall	F1-score
Phrase Matching	145	64.16%	39.12%	48.60%
Acronym Matching	138	90.80%	84.90%	87.80%
N-Grams	9	75.35%	23.83%	36.21%
TF-IDF	142	79.51%	56.90%	66.36%
Word2Vec	313	76.98%	43.36%	55.47%





## Future Work

- Fine tune hyperparameter further
- Integrate data sources into database(e.g – Neo4j)
- Push the found series back to Wikidata
- Data exploration
  - Distribution of locations
  - Trend in workshop topics
- Use additional data sources such as WikiCFP, Scholia
- Comprehensive usage of Large Language Models(LLMs)
- Optimize performance

# Thank You!