Link Prediction for Event Logs in the Process Industry

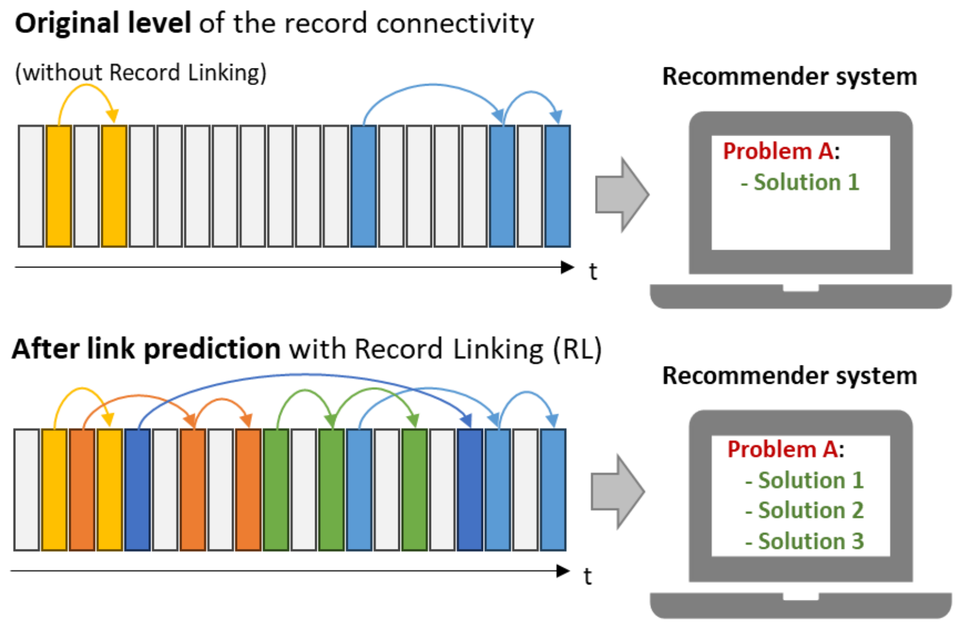
Anastasia Zhukova[1], Thomas Walton[2], Christian E. Lobmüller[2], and Bela Gipp[1]
[1]*University of Göttingen, Germany* [2]*eschbach GmbH, Germany*

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737
eschbach
LREC 20 Palma 26

Overview

- **Relying on a knowledge graph**, especially in low-resource settings, helps LLMs **compensate for not knowing a specific domain** well.
- **Improving the quality**, consistency, and connectivity of the underlying **linked data** is required to ensure the **effective use of the collected domain knowledge** in the domain-specific **RAG systems**.

How can we **adapt common NLP tasks** to a **link prediction** task aimed at **improving data quality and connectivity** in **German event logs** of daily operations?

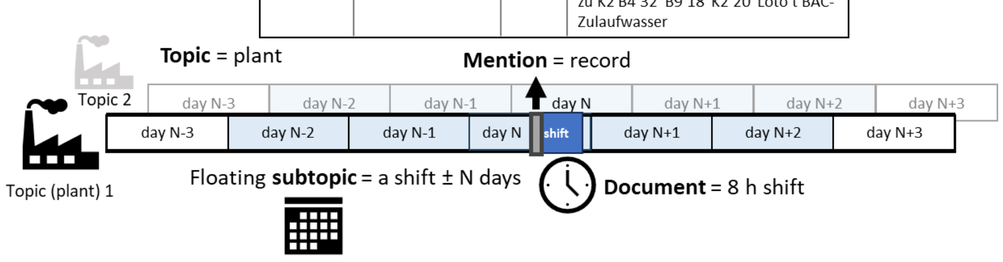


Background

Example of logs of daily operations

ID	Timestamp	Shift	Func. location	Description	Related to
001	29-01-26 10:47	Shift 1	A013-DR-330	Temperature spike detected in reactor chamber.	-
002	29-01-26 11:15	Shift 1	B716-RX-204	Routine inspection completed no anomalies detected.	-
003	29-01-26 15:02	Shift 2	C118-MX-118	Lubrication cycle executed successfully.	-
004	29-01-26 18:10	Shift 2	A013-DR-330	Cooling valve recalibrated and flow rate increased.	001
005	29-01-26 21:25	Shift 2	B716-FL-501	Filter replaced as part of scheduled maintenance.	-
006	29-01-26 22:55	Shift 3	C514-CN-210	Conveyor belt misalignment causing material spillage.	-
007	30-01-26 01:05	Shift 3	A013-DR-330	Additional insulation added to stabilize temperature fluctuations.	004
008	30-01-26 05:20	Shift 3	C514-CN-210	Belt realigned and tension adjusted.	006
009	30-01-26 06:18	Shift 1	A013-TK-777	Tank pressure levels within normal operating range.	-
010	30-01-26 10:02	Shift 1	C514-MX-118	Mixer calibration verified and logged.	-

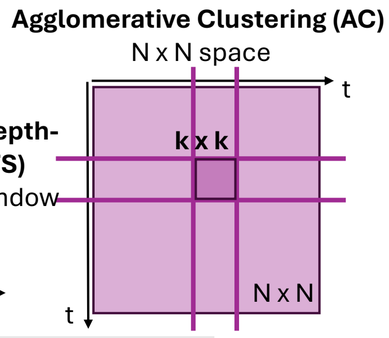
Time stamp	Functional locations	Product	Text logs
2021/08/01 10:04	FL 1-1-2 Tank	ABC	Gesendet an HAH Transfer von B6 nach B1 98779 H2 Wasser nach B6 98781 H2 Organik bleibt bei SFP Wasser D.O. 2-1 .59 2-3 11.06 Kohlenstofftransfer zu K2 B4 32' B9 18' K2 20' Loto't BAC-Zulaufwasser



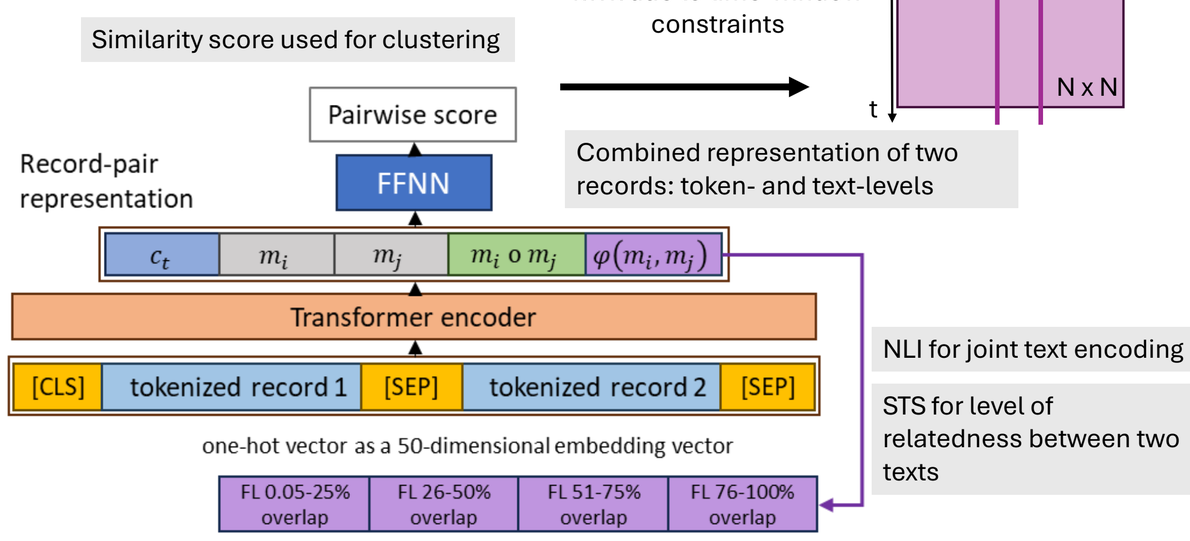
Task adaptation to cross-document coreference resolution (CDCR)

Methodology

Combined architecture of two state-of-the-art CDCR models



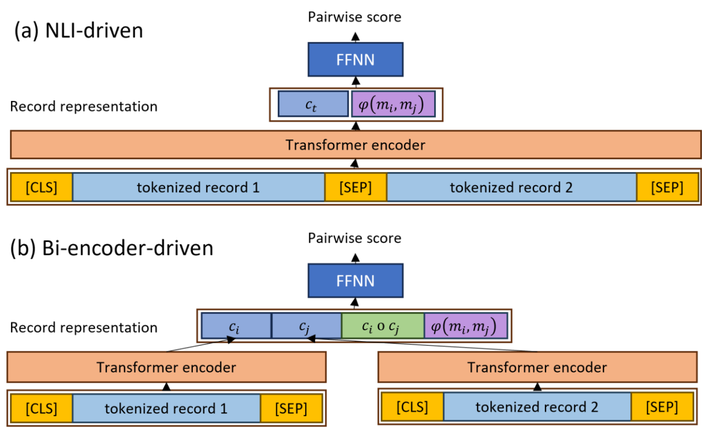
time-dependent depth-first search (tDFS)
k x k due to time-window constraints



Feature vector based on the record metadata (functional location)

Evaluation

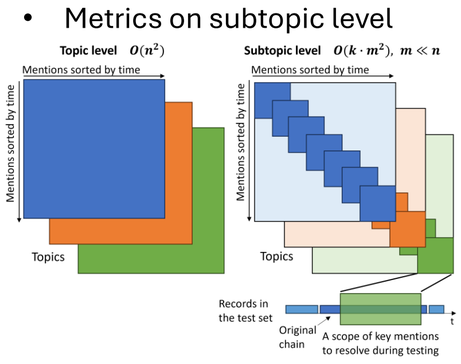
- Baselines



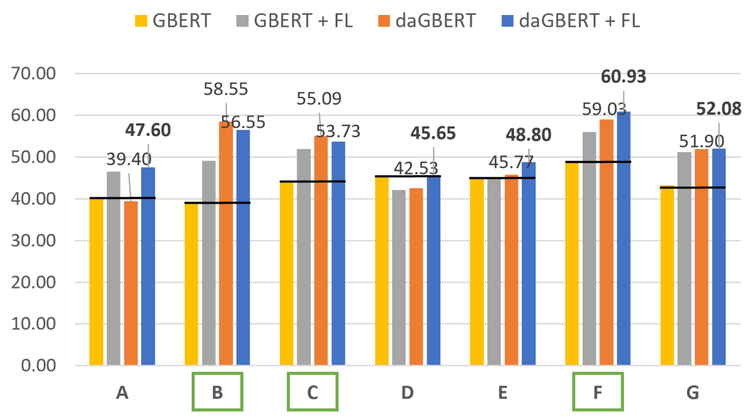
- Dataset

Plant	Train	Dev	Test
A	9579	1155	1174
B	-	-	930
C	1013	1016	1041
D	8341	1103	1103
E	8121	1110	1079
F	956	1090	905
G	8643	1253	1188
Total	36653	6727	7420

- Metrics on subtopic level



Architecture	Lang.model	Func.Loc.	F1-score	Clustering	CoNLL F1
NLI-driven	GBERT	-	78.83	AC	37.85
				tDFS	31.84
		+	76.58	AC	37.85
				tDFS	38.93
	daGBERT	-	72.64	AC	37.85
				tDFS	39.66
		+	68.52	AC	37.85
				tDFS	40.76
STS-driven	GBERT	-	67.34	AC	38.02
				tDFS	40.22
		+	66.64	AC	37.9
				tDFS	40.17
	mGTE	-	66.46	AC	37.85
				tDFS	37.81
		+	65.37	AC	37.85
				tDFS	34.3
	daGBERT	-	72.52	AC	38.21
				tDFS	40.98
		+	68.65	AC	38.49
				tDFS	39.79
RL (CDCR-driven)	GBERT	-	78.83	AC	38.23
				tDFS	43.79
		+	78.10	AC	39.2
				tDFS	48.83
	daGBERT	-	**81.05**	AC	41.44
				tDFS	*50.32*
		+	*80.51*	AC	41.33
				tDFS	**52.19**



Plants B, C, and F have no to little training data

daGBERT + FL > daGBERT > > GBERT + FL > GBERT

But for some plants funcl.location feature vector doesn't have too much influence
→ less desecrate vector is required

Findings

- **Record linking** performance emerges from the **interaction** of **architecture** (CDCR + NLI + STS), **representations**, and **clustering** — not individual components
- **LM domain adaptation** & **feature integration** are crucial: Domain-adapted model (daGBERT) + FL feature vector
- **tDFS provides temporal constraints** to avoid clustering false positives (e.g., similar incidents that occurred at different times)

Contact

Anastasia Zhukova

LinkedIn
Plant assistant project
https://gipplab.org/zhukova/
anastasia.zhukova@uni-goettingen.de