TIGER: Temporally Improved Graph Entity Linker Supplementary Material

1. DATASET CONSTRUCTION

We combined the TempEL[1] dataset, a benchmark for temporal entity linking, with Wikidata5M[2] to explore the benefits of structured knowledge graphs. Our resultant dataset has five segments: two text-based (**entity description** and **mention context**) and three graph-based (**structure graph**, **feature graph**, and **feature matrix**). Both the structure and feature graphs depict yearly entity relationships. Continual entities refer to existing entities in previous years, and new entities refer to newly appeared, previously non-existent entities. The construction process is shown in Figure S1, with detailed steps provided in supplementary material. We make the dataset available in the TIGER Github repository: https://anonymous.4open.science/r/TIGER-Temporally-Improved-Graph-Entity-Linker-0CC3/README.md. The data are made available at 10.5281/zenodo.10977756. We now walk through each step.

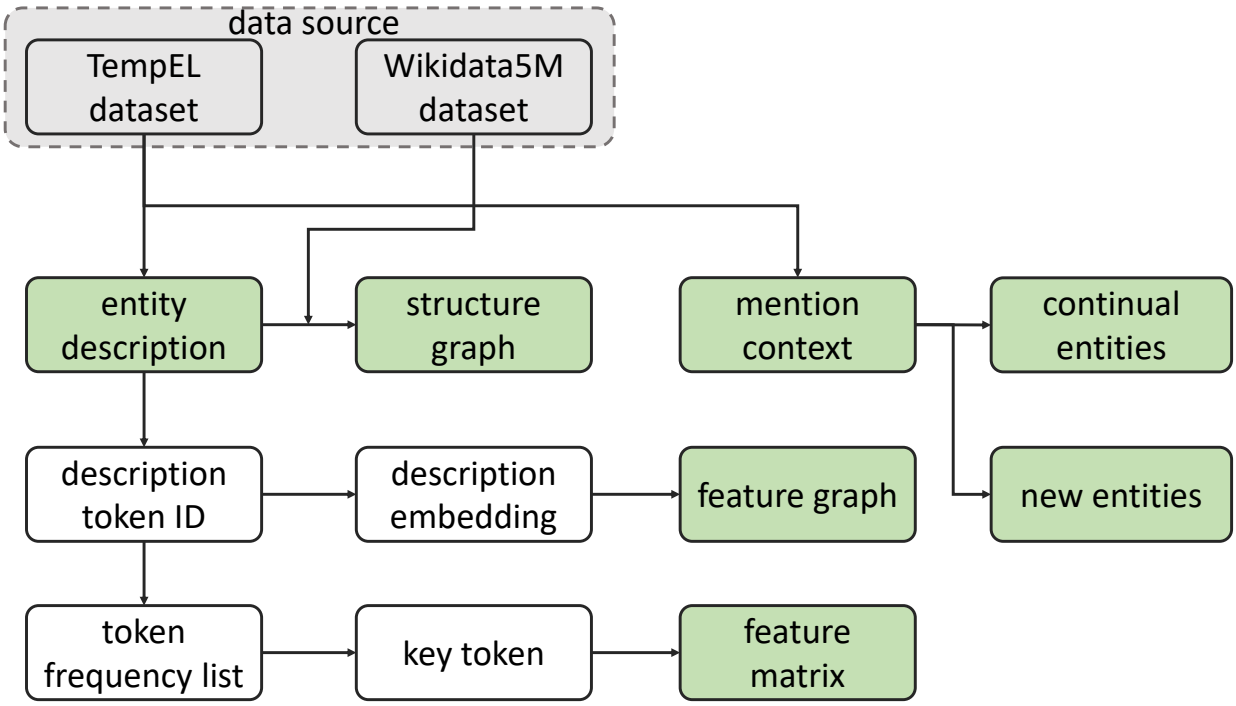


Fig. S1. Dataset construction process.

First, we categorized each year of data from the TempEL dataset into entity descriptions and mention context parts based on the year. The entity description comprises the title, text, document ID, and, importantly, the unique ID of the entity (its QID). The mention context consists of context left, context right, mention, label, QID, and category.

Second, we create a structure graph based on the relationship in the Wikidata5M dataset and the entity IDs in the TempEL dataset. There are numerous relationships among entities in the Wikidata5M dataset. To filter down the number of relationships, we matched these relationships' entity IDs (also QIDs) with the QID in the entity descriptions from the TempEL dataset. We will keep the relationship if both QIDs are in a

[1]https://cloud.ilabt.imec.be/index.php/s/RinXy8NgqdW58RW
[2]https://deepgraphlearning.github.io/project/wikidata5m

relationship in the Wikidata5M data and are present in the existing entity description. The structure graph is an $n \times n$ adjacency matrix, where n represents the total number of entities in the dataset. Each row indicates whether an entity has a connection with another entity. The adjacency matrix is made up of 0s and 1s. If entity i and entity j are connected, the value in the i^{th} row and j^{th} column of the matrix is 1; otherwise, it is 0.

Note that, at this point, we may have introduced temporal leakage. This is because the cutoff for the Wikidata5M dataset is July 2019. If we construct a structure graph based on Wikidata5M, temporal leakage might occur when dealing with data from years prior to 2019. For instance, if entity 1 and entity 2 had no relationship in 2013 but established one in 2019, our construction of the 2013 structure graph would inevitably link entity 1 and entity 2. However, the efficacy of our model can still be demonstrated. For example, when we only use data from 2019 to 2022, our model demonstrates a 16.24% performance boost over the state-of-the-art in a temporal setting when the time gap is one year and an improvement to 20.93% as the gap expands to three years.

Third, we built the feature graph using the embeddings from entity descriptions. We employed the pre-trained bert-base-uncased model to embed the entity description's textual information associated with the "text" key. By accessing the embedded information for each entity in the dataset, we established a kNN graph based on these entities, which we refer to as the feature graph. This graph highlights the connections between entities based on their entity descriptions. The feature graph is also an $n \times n$ adjacency matrix, where n represents the total number of entities in the dataset. Each row indicates whether an entity has a connection with other entities. If entity i and entity j are connected, the value in the i^{th} row and j^{th} column of the matrix is 1; otherwise, it is 0.

The difference between connections in the structure graph and those in the feature graph is that the connections in the structure graph are constructed based on actual existing links. For instance, on the Wikipedia page for "***Weightlifting (Q83462)***", the "***Summer Olympic Games (Q212434)***" is mentioned in the first two paragraphs. Hence, a connection between these two entities exists in the structure graph. However, connections in the feature graph are generated based on the embeddings of entity descriptions, which might not exist in the KG but can still benefit the model. For example, while "***Arizona_Wildcats (Q4620330)***" and "***Summer Olympic Games (Q212434)***" are not directly related, the embeddings of their descriptions have created a link between them. This makes "***Arizona_Wildcats (Q4620330)***" less likely to be confused with "***Arizona (Q816)***" or "***Wildcats (Q26665)***" a 1986 film by Michael Ritchie.

Fourth, we constructed a feature matrix representing each entity based on the tokens from entity descriptions in the dataset. After getting the token IDs for each entity using the pre-trained bert-base-uncased model, we filtered all token IDs based on their frequency of occurrence. We retained those token IDs that appeared between 46 and 200 times. We discarded highly frequent token IDs since these tokens, such as "is," "an," "the," and other common words, do not offer meaningful differentiation among entities. Also, the less frequent token IDs were removed due to the possibility of them being meaningless noise or random codes, and including an excess of these rare tokens would make the matrix too sparse, slowing down computation. The final feature matrix is an $n \times m$ dimensional matrix composed of 0s and 1s. Here, n represents the total number of entities in the dataset, while m is the number of retained token IDs. If the data in the i^{th} row and j^{th} column of the matrix is 1, it indicates that entity i contains the j^{th} token.

Finally, we generate distinct mention context subsets from all available mention context samples. Using the "category" in each sample as the standard, we further divided the training set into two sub-training sets: "Continual entities (existing entities in previous years)" and "New entities (newly appeared, previously non-existent entities)."

2. TRAINING DETAILS

Parameter Settings. We reuse the same hyperparameter settings from BLINK model and the same bert_uncased_L-8_H-512_A-8 pre-trained model to train the bi-encoder. The recall@N is used as the evaluation metric, where N equals 1, 2, 4, 8, 16, 32, and 64, respectively. If the correct answer appears within the top N predictions of the model, it is considered a correct prediction. The bi-encoder is trained on the ZESHEL dataset across five epochs, utilizing 128 mentions and 128 entity tokens at a learning rate 1e-05. Conversely, the bi-encoder undergoes training for one epoch on our dataset, maintaining similar mention and entity token quantities and learning rates. The training process employs an annual training approach and tests on all test sets. For instance, the results for 2019 are obtained by training on the 2019 training set and validating on the 2019 validation set. Once trained, the model is then saved. Subsequently, testing is performed on the test sets spanning the four years from 2019 to 2022.

Training Environment. Software versions: Python 3.9.16; PyTorch 1.7.1+cu110; Faiss-GPU 1.6.5; Numpy 1.24.2; SciPy 1.10.1; scikit-learn 1.2.1. All the experiments were run on a single A100 GPU with 40GB RAM.

3. ADDITIONAL RESULTS - NEW ENTITIES TRAINING SET

Tables S1 through Table S7 present a comparison between the baseline models (BLINK and SpEL) and our model TIGER. The dataset, "Extended TempEL: New entities," has a training set (1,764), a validation set ($\approx 42k$, same as original TempEL dataset), and a test set ($\approx 48k$, same as original TempEL dataset). Here, the training set contains only 1,764 samples because, in the original TempEL dataset, each year's dataset contains only 1,764 'new entities' samples. Rows represent training datasets, while columns represent testing datasets. For example, the number at the intersection of the first row and the tenth column in Table S1 is 0.1433, showing the model's performance when trained on 2013 data and tested on 2022 data. Performance evaluations for all models were based on recall metrics, specifically @1, @2, @4, @8, @16, @32, and @64.

Table S1. Results between our model and the baseline models @1 ("Extended TempEL: New entities" as the training set).

	Train \ Test	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
BLINK	2013	0.1440	0.1435	0.1431	0.1459	0.1473	0.1450	0.1431	0.1477	0.1445	0.1433
	2014	0.1221	0.1229	0.1229	0.1235	0.1222	0.1220	0.1202	0.1260	0.1228	0.1246
	2015	0.1380	0.1359	0.1371	0.1384	0.1397	0.1373	0.1371	0.1412	0.1376	0.1366
	2016	0.1310	0.1320	0.1308	0.1321	0.1330	0.1304	0.1314	0.1349	0.1316	0.1326
	2017	0.1403	0.1404	0.1404	0.1415	0.1428	0.1424	0.1402	0.1438	0.1410	0.1422
	2018	0.1284	0.1269	0.1268	0.1268	0.1301	0.1276	0.1284	0.1306	0.1268	0.1270
	2019	0.1400	0.1397	0.1418	0.1415	0.1427	0.1411	0.1409	0.1449	0.1406	0.1424
	2020	0.1228	0.1231	0.1206	0.1228	0.1245	0.1224	0.1206	0.1261	0.1216	0.1238
	2021	0.1270	0.1267	0.1270	0.1265	0.1271	0.1243	0.1239	0.1303	0.1294	0.1305
	2022	0.1266	0.1264	0.1264	0.1275	0.1297	0.1271	0.1265	0.1313	0.1288	0.1300
SpEL	2013	0.1740	0.1735	0.1931	0.1759	0.1773	0.1950	0.1831	0.1777	0.1945	0.1933
	2014	0.1721	0.1529	0.1729	0.1735	0.1622	0.1520	0.1502	0.1760	0.1728	0.1646
	2015	0.1680	0.1659	0.1771	0.1784	0.1797	0.1873	0.1671	0.1812	0.1676	0.1866
	2016	0.1610	0.1820	0.1708	0.1821	0.1630	0.1604	0.1714	0.1749	0.1616	0.1626
	2017	0.1803	0.1804	0.1804	0.1815	0.1828	0.1724	0.1902	0.1838	0.1810	0.1722
	2018	0.1684	0.1569	0.1668	0.1568	0.1601	0.1776	0.1584	0.1806	0.1668	0.1670
	2019	0.1800	0.1697	0.1718	0.1915	0.1727	0.1811	0.1809	0.1849	0.1706	0.1924
	2020	0.1628	0.1731	0.1606	0.1728	0.1545	0.1724	0.1706	0.1661	0.1516	0.1638
	2021	0.1770	0.1567	0.1770	0.1765	0.1671	0.1643	0.1739	0.1603	0.1794	0.1705
	2022	0.1566	0.1564	0.1464	0.1475	0.1497	0.1571	0.1565	0.1613	0.1488	0.1600
TIGER	2013	0.2245	0.2242	0.2257	0.2299	0.2283	0.2290	0.2245	0.2285	0.2294	0.2304
	2014	0.2011	0.2022	0.2011	0.2029	0.2062	0.2019	0.2010	0.2048	0.2031	0.2003
	2015	0.2307	0.2320	0.2313	0.2319	0.2355	0.2349	0.2343	0.2359	0.2355	0.2346
	2016	0.2154	0.2185	0.2161	0.2206	0.2206	0.2208	0.2180	0.2222	0.2185	0.2197
	2017	0.2292	0.2306	0.2297	0.2313	0.2335	0.2328	0.2313	0.2330	0.2320	0.2318
	2018	0.1725	0.1707	0.1710	0.1706	0.1753	0.1740	0.1738	0.1768	0.1750	0.1762
	2019	0.2156	0.2195	0.2172	0.2166	0.2208	0.2198	0.2174	0.2158	0.2156	0.2169
	2020	0.1570	0.1550	0.1573	0.1579	0.1606	0.1604	0.1591	0.1627	0.1604	0.1605
	2021	0.2012	0.2025	0.2013	0.2057	0.2049	0.2084	0.2049	0.2072	0.2079	0.2076
	2022	0.1523	0.1529	0.1548	0.1579	0.1598	0.1590	0.1566	0.1589	0.1566	0.1574

Table S2. Results between our model and the baseline models @2 ("Extended TempEL: New entities" as the training set).

	Train \ Test	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
BLINK	2013	0.2131	0.2141	0.2136	0.2139	0.2156	0.2157	0.2141	0.2182	0.2133	0.2117
	2014	0.1845	0.1850	0.1839	0.1840	0.1851	0.1841	0.1809	0.1858	0.1826	0.1858
	2015	0.2035	0.2035	0.2061	0.2059	0.2072	0.2064	0.2053	0.2083	0.2048	0.2034
	2016	0.1987	0.1983	0.1972	0.1985	0.1999	0.1983	0.1955	0.2007	0.1965	0.1990
	2017	0.2099	0.2100	0.2103	0.2117	0.2129	0.2133	0.2092	0.2133	0.2109	0.2119
	2018	0.1902	0.1910	0.1928	0.1916	0.1937	0.1917	0.1915	0.1942	0.1896	0.1912
	2019	0.2073	0.2079	0.2109	0.2112	0.2123	0.2128	0.2087	0.2137	0.2092	0.2108
	2020	0.1862	0.1855	0.1848	0.1853	0.1891	0.1871	0.1838	0.1896	0.1844	0.1871
	2021	0.1910	0.1926	0.1900	0.1908	0.1915	0.1892	0.1868	0.1938	0.1945	0.1927
	2022	0.1912	0.1917	0.1909	0.1915	0.1943	0.1911	0.1889	0.1951	0.1925	0.1937
SpEL	2013	0.2531	0.2541	0.2536	0.2739	0.2656	0.2657	0.2541	0.2582	0.2633	0.2617
	2014	0.2445	0.2250	0.2339	0.2340	0.2351	0.2341	0.2209	0.2358	0.2226	0.2358
	2015	0.2435	0.2635	0.2461	0.2459	0.2472	0.2564	0.2553	0.2483	0.2548	0.2634
	2016	0.2387	0.2483	0.2472	0.2485	0.2399	0.2483	0.2455	0.2507	0.2365	0.2490
	2017	0.2499	0.2700	0.2503	0.2717	0.2729	0.2633	0.2592	0.2533	0.2509	0.2519
	2018	0.2502	0.2510	0.2328	0.2516	0.2437	0.2417	0.2315	0.2442	0.2496	0.2512
	2019	0.2473	0.2479	0.2609	0.2612	0.2623	0.2528	0.2587	0.2537	0.2692	0.2608
	2020	0.2262	0.2255	0.2248	0.2353	0.2491	0.2271	0.2338	0.2296	0.2344	0.2471
	2021	0.2510	0.2526	0.2500	0.2408	0.2315	0.2392	0.2268	0.2338	0.2545	0.2527
	2022	0.2112	0.2217	0.2109	0.2215	0.2143	0.2211	0.2089	0.2251	0.2225	0.2137
TIGER	2013	0.3213	0.3242	0.3235	0.3286	0.3276	0.3280	0.3208	0.3250	0.3280	0.3301
	2014	0.2921	0.2935	0.2910	0.2928	0.2956	0.2938	0.2919	0.2954	0.2958	0.2924
	2015	0.3283	0.3286	0.3306	0.3317	0.3361	0.3357	0.3302	0.3327	0.3345	0.3347
	2016	0.3113	0.3125	0.3148	0.3160	0.3181	0.3167	0.3158	0.3172	0.3161	0.3159
	2017	0.3270	0.3290	0.3290	0.3303	0.3326	0.3296	0.3275	0.3300	0.3293	0.3316
	2018	0.2555	0.2545	0.2556	0.2569	0.2594	0.2590	0.2584	0.2632	0.2601	0.2627
	2019	0.3103	0.3134	0.3139	0.3137	0.3160	0.3146	0.3121	0.3111	0.3124	0.3138
	2020	0.2346	0.2338	0.2362	0.2365	0.2409	0.2414	0.2394	0.2424	0.2434	0.2420
	2021	0.2924	0.2964	0.2936	0.2967	0.3005	0.3009	0.2979	0.3013	0.3026	0.3002
	2022	0.2314	0.2336	0.2355	0.2366	0.2396	0.2390	0.2382	0.2407	0.2367	0.2383

Table S3. Results between our model and the baseline models @4 ("Extended TempEL: New entities" as the training set).

	Train \ Test	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
BLINK	2013	0.2953	0.2951	0.2971	0.2986	0.3001	0.2994	0.2973	0.3010	0.2961	0.2952
	2014	0.2583	0.2599	0.2584	0.2591	0.2620	0.2610	0.2587	0.2612	0.2581	0.2609
	2015	0.2847	0.2836	0.2856	0.2861	0.2882	0.2866	0.2848	0.2889	0.2840	0.2847
	2016	0.2774	0.2778	0.2794	0.2782	0.2808	0.2795	0.2749	0.2788	0.2768	0.2784
	2017	0.2908	0.2935	0.2943	0.2943	0.2976	0.2948	0.2924	0.2955	0.2919	0.2946
	2018	0.2689	0.2676	0.2674	0.2682	0.2697	0.2700	0.2652	0.2703	0.2654	0.2685
	2019	0.2912	0.2933	0.2930	0.2927	0.2946	0.2941	0.2926	0.2951	0.2914	0.2944
	2020	0.2642	0.2633	0.2641	0.2654	0.2683	0.2661	0.2648	0.2680	0.2627	0.2649
	2021	0.2686	0.2701	0.2699	0.2690	0.2715	0.2698	0.2663	0.2722	0.2739	0.2719
	2022	0.2707	0.2706	0.2705	0.2718	0.2728	0.2712	0.2684	0.2741	0.2719	0.2731
SpEL	2013	0.3553	0.3551	0.3471	0.3586	0.3701	0.3694	0.3673	0.3710	0.3561	0.3652
	2014	0.3183	0.3099	0.3084	0.3291	0.3120	0.3310	0.3287	0.3312	0.3081	0.3209
	2015	0.3547	0.3336	0.3356	0.3561	0.3382	0.3466	0.3448	0.3489	0.3340	0.3547
	2016	0.3274	0.3278	0.3294	0.3282	0.3508	0.3495	0.3349	0.3288	0.3268	0.3284
	2017	0.3608	0.3635	0.3443	0.3443	0.3576	0.3648	0.3424	0.3455	0.3419	0.3546
	2018	0.3289	0.3276	0.3174	0.3182	0.3197	0.3200	0.3152	0.3303	0.3254	0.3385
	2019	0.3412	0.3633	0.3630	0.3627	0.3646	0.3441	0.3626	0.3651	0.3414	0.3544
	2020	0.3342	0.3333	0.3341	0.3254	0.3183	0.3261	0.3248	0.3280	0.3127	0.3249
	2021	0.3386	0.3201	0.3299	0.3390	0.3415	0.3198	0.3263	0.3322	0.3239	0.3419
	2022	0.3107	0.2906	0.2905	0.2918	0.2928	0.3112	0.3084	0.3041	0.3019	0.3031
TIGER	2013	0.4288	0.4313	0.4336	0.4378	0.4353	0.4334	0.4295	0.4337	0.4338	0.4376
	2014	0.3923	0.3961	0.3930	0.3965	0.3975	0.3961	0.3940	0.3979	0.3968	0.3940
	2015	0.4343	0.4380	0.4394	0.4391	0.4407	0.4417	0.4384	0.4394	0.4417	0.4427
	2016	0.4165	0.4181	0.4187	0.4207	0.4211	0.4227	0.4205	0.4223	0.4198	0.4202
	2017	0.4310	0.4348	0.4340	0.4346	0.4365	0.4343	0.4327	0.4369	0.4339	0.4386
	2018	0.3523	0.3531	0.3525	0.3543	0.3566	0.3555	0.3531	0.3595	0.3574	0.3599
	2019	0.4156	0.4190	0.4170	0.4191	0.4203	0.4190	0.4163	0.4176	0.4195	0.4210
	2020	0.3302	0.3312	0.3334	0.3374	0.3378	0.3395	0.3361	0.3391	0.3425	0.3425
	2021	0.3949	0.3993	0.3984	0.4011	0.4045	0.4006	0.4017	0.4043	0.4072	0.4062
	2022	0.3268	0.3294	0.3289	0.3325	0.3355	0.3351	0.3330	0.3358	0.3348	0.3349

Table S4. Results between our model and the baseline models @8 ("Extended TempEL: New entities" as the training set).

	Train \ Test	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
BLINK	2013	0.3888	0.3892	0.3913	0.3942	0.3937	0.3918	0.3910	0.3918	0.3881	0.3883
	2014	0.3485	0.3489	0.3502	0.3499	0.3507	0.3494	0.3474	0.3492	0.3471	0.3493
	2015	0.3754	0.3761	0.3792	0.3788	0.3819	0.3779	0.3745	0.3767	0.3747	0.3750
	2016	0.3711	0.3693	0.3717	0.3715	0.3734	0.3716	0.3686	0.3720	0.3694	0.3704
	2017	0.3862	0.3886	0.3900	0.3889	0.3917	0.3881	0.3862	0.3888	0.3872	0.3871
	2018	0.3604	0.3596	0.3598	0.3587	0.3600	0.3588	0.3553	0.3600	0.3545	0.3573
	2019	0.3860	0.3895	0.3892	0.3911	0.3906	0.3904	0.3875	0.3901	0.3885	0.3915
	2020	0.3538	0.3558	0.3591	0.3589	0.3606	0.3573	0.3559	0.3603	0.3558	0.3585
	2021	0.3604	0.3604	0.3620	0.3612	0.3644	0.3636	0.3584	0.3654	0.3688	0.3640
	2022	0.3619	0.3631	0.3634	0.3642	0.3653	0.3628	0.3620	0.3655	0.3649	0.3646
SpEL	2013	0.4488	0.4692	0.4613	0.4642	0.4637	0.4618	0.4610	0.4618	0.4581	0.4483
	2014	0.4185	0.4289	0.4102	0.4199	0.4207	0.4094	0.4074	0.4292	0.4071	0.4293
	2015	0.4554	0.4561	0.4592	0.4488	0.4619	0.4379	0.4545	0.4567	0.4447	0.4450
	2016	0.4511	0.4493	0.4417	0.4415	0.4534	0.4316	0.4386	0.4420	0.4294	0.4404
	2017	0.4562	0.4586	0.4700	0.4489	0.4617	0.4581	0.4662	0.4588	0.4472	0.4671
	2018	0.4304	0.4296	0.4298	0.4387	0.4400	0.4388	0.4353	0.4200	0.4145	0.4373
	2019	0.4460	0.4695	0.4592	0.4611	0.4706	0.4604	0.4475	0.4701	0.4585	0.4715
	2020	0.4338	0.4158	0.4391	0.4389	0.4406	0.4373	0.4259	0.4203	0.4158	0.4185
	2021	0.4304	0.4404	0.4320	0.4312	0.4244	0.4236	0.4184	0.4354	0.4288	0.4340
	2022	0.3919	0.3931	0.4034	0.4142	0.4053	0.4028	0.4020	0.3955	0.3949	0.3946
TIGER	2013	0.5369	0.5393	0.5416	0.5448	0.5431	0.5412	0.5384	0.5411	0.5428	0.5453
	2014	0.4959	0.4968	0.4980	0.5002	0.5016	0.5004	0.5000	0.5015	0.5015	0.5002
	2015	0.5429	0.5471	0.5489	0.5493	0.5519	0.5488	0.5472	0.5465	0.5482	0.5506
	2016	0.5237	0.5255	0.5255	0.5282	0.5291	0.5276	0.5257	0.5283	0.5300	0.5310
	2017	0.5377	0.5414	0.5380	0.5405	0.5452	0.5409	0.5414	0.5459	0.5435	0.5451
	2018	0.4578	0.4593	0.4599	0.4610	0.4631	0.4607	0.4596	0.4635	0.4639	0.4662
	2019	0.5293	0.5284	0.5292	0.5328	0.5324	0.5308	0.5292	0.5310	0.5303	0.5330
	2020	0.4387	0.4388	0.4413	0.4471	0.4448	0.4452	0.4444	0.4453	0.4482	0.4498
	2021	0.5034	0.5063	0.5060	0.5095	0.5126	0.5089	0.5094	0.5125	0.5165	0.5145
	2022	0.4336	0.4363	0.4376	0.4403	0.4420	0.4405	0.4399	0.4407	0.4405	0.4427

Table S5. Results between our model and the baseline models @16 ("Extended TempEL: New entities" as the training set).

	Train \ Test	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
BLINK	2013	0.4861	0.4869	0.4897	0.4907	0.4933	0.4906	0.4878	0.4884	0.4846	0.4848
	2014	0.4452	0.4476	0.4477	0.4479	0.4494	0.4486	0.4469	0.4492	0.4449	0.4474
	2015	0.4728	0.4760	0.4768	0.4782	0.4785	0.4767	0.4715	0.4719	0.4699	0.4711
	2016	0.4708	0.4709	0.4737	0.4729	0.4747	0.4737	0.4702	0.4712	0.4690	0.4702
	2017	0.4834	0.4867	0.4886	0.4883	0.4893	0.4894	0.4864	0.4871	0.4849	0.4856
	2018	0.4559	0.4585	0.4589	0.4599	0.4584	0.4580	0.4548	0.4556	0.4513	0.4536
	2019	0.4884	0.4922	0.4928	0.4931	0.4947	0.4933	0.4921	0.4918	0.4913	0.4908
	2020	0.4561	0.4580	0.4608	0.4618	0.4625	0.4611	0.4582	0.4591	0.4557	0.4561
	2021	0.4611	0.4608	0.4633	0.4626	0.4671	0.4655	0.4613	0.4634	0.4696	0.4652
	2022	0.4600	0.4644	0.4641	0.4643	0.4671	0.4639	0.4635	0.4642	0.4630	0.4657
SpEL	2013	0.5661	0.5669	0.5597	0.5707	0.5733	0.5506	0.5578	0.5484	0.5446	0.5448
	2014	0.5252	0.5176	0.5077	0.5279	0.5194	0.5086	0.5069	0.5292	0.5249	0.5274
	2015	0.5428	0.5460	0.5568	0.5582	0.5585	0.5467	0.5415	0.5519	0.5299	0.5511
	2016	0.5308	0.5309	0.5437	0.5529	0.5547	0.5537	0.5502	0.5412	0.5490	0.5302
	2017	0.5434	0.5467	0.5486	0.5483	0.5593	0.5494	0.5664	0.5671	0.5549	0.5656
	2018	0.5359	0.5285	0.5289	0.5199	0.5384	0.5280	0.5248	0.5256	0.5113	0.5336
	2019	0.5684	0.5522	0.5728	0.5631	0.5547	0.5733	0.5621	0.5618	0.5713	0.5508
	2020	0.5261	0.5280	0.5308	0.5318	0.5425	0.5311	0.5282	0.5391	0.5257	0.5361
	2021	0.5411	0.5208	0.5333	0.5326	0.5371	0.5455	0.5213	0.5334	0.5396	0.5352
	2022	0.5100	0.5044	0.5141	0.5143	0.5071	0.5039	0.5135	0.5142	0.5030	0.5157
TIGER	2013	0.6426	0.6422	0.6444	0.6444	0.6460	0.6460	0.6449	0.6441	0.6467	0.6489
	2014	0.5986	0.6005	0.5999	0.5993	0.6022	0.5995	0.6027	0.6042	0.6030	0.6033
	2015	0.6461	0.6503	0.6518	0.6510	0.6535	0.6497	0.6490	0.6487	0.6496	0.6530
	2016	0.6302	0.6291	0.6324	0.6338	0.6342	0.6324	0.6325	0.6336	0.6352	0.6352
	2017	0.6377	0.6403	0.6419	0.6412	0.6457	0.6412	0.6420	0.6446	0.6419	0.6437
	2018	0.5651	0.5670	0.5682	0.5678	0.5687	0.5679	0.5652	0.5699	0.5690	0.5711
	2019	0.6366	0.6373	0.6366	0.6376	0.6383	0.6366	0.6355	0.6354	0.6367	0.6376
	2020	0.5506	0.5514	0.5516	0.5559	0.5575	0.5571	0.5575	0.5567	0.5587	0.5617
	2021	0.6068	0.6109	0.6113	0.6111	0.6162	0.6115	0.6126	0.6141	0.6193	0.6189
	2022	0.5459	0.5490	0.5509	0.5504	0.5548	0.5510	0.5514	0.5521	0.5539	0.5551

Table S6. Results between our model and the baseline models @32 ("Extended TempEL: New entities" as the training set).

	Train \ Test	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
BLINK	2013	0.5882	0.5907	0.5918	0.5907	0.5920	0.5899	0.5871	0.5870	0.5832	0.5855
	2014	0.5467	0.5505	0.5511	0.5507	0.5515	0.5509	0.5482	0.5491	0.5456	0.5471
	2015	0.5722	0.5766	0.5781	0.5773	0.5780	0.5772	0.5744	0.5738	0.5706	0.5709
	2016	0.5735	0.5771	0.5771	0.5760	0.5789	0.5761	0.5743	0.5738	0.5708	0.5715
	2017	0.5854	0.5894	0.5892	0.5873	0.5914	0.5900	0.5874	0.5873	0.5843	0.5860
	2018	0.5561	0.5588	0.5601	0.5587	0.5602	0.5598	0.5562	0.5563	0.5517	0.5536
	2019	0.5925	0.5960	0.5976	0.5943	0.5981	0.5961	0.5967	0.5940	0.5934	0.5930
	2020	0.5596	0.5627	0.5650	0.5640	0.5658	0.5647	0.5631	0.5642	0.5588	0.5606
	2021	0.5644	0.5657	0.5675	0.5684	0.5716	0.5707	0.5696	0.5699	0.5765	0.5738
	2022	0.5634	0.5679	0.5668	0.5677	0.5685	0.5679	0.5681	0.5677	0.5664	0.5660
SpEL	2013	0.6582	0.6507	0.6518	0.6607	0.6520	0.6499	0.6471	0.6570	0.6432	0.6555
	2014	0.6067	0.6205	0.6111	0.6107	0.6215	0.6309	0.6182	0.6091	0.6056	0.6271
	2015	0.6322	0.6566	0.6481	0.6373	0.6480	0.6572	0.6444	0.6438	0.6406	0.6409
	2016	0.6435	0.6571	0.6571	0.6360	0.6489	0.6361	0.6443	0.6538	0.6408	0.6515
	2017	0.6554	0.6594	0.6692	0.6673	0.6514	0.6700	0.6674	0.6473	0.6443	0.6460
	2018	0.6161	0.6188	0.6401	0.6287	0.6402	0.6298	0.6162	0.6263	0.6117	0.6136
	2019	0.6525	0.6760	0.6776	0.6543	0.6681	0.6561	0.6567	0.6540	0.6534	0.6730
	2020	0.6196	0.6227	0.6450	0.6340	0.6358	0.6247	0.6331	0.6342	0.6388	0.6206
	2021	0.6444	0.6457	0.6275	0.6284	0.6416	0.6507	0.6496	0.6399	0.6565	0.6438
	2022	0.6234	0.6179	0.6168	0.6177	0.6285	0.6279	0.6181	0.6277	0.6264	0.6160
TIGER	2013	0.7365	0.7348	0.7381	0.7347	0.7379	0.7365	0.7360	0.7365	0.7396	0.7401
	2014	0.6923	0.6938	0.6937	0.6924	0.6960	0.6954	0.6986	0.6991	0.6982	0.7016
	2015	0.7367	0.7403	0.7406	0.7409	0.7429	0.7390	0.7377	0.7395	0.7398	0.7416
	2016	0.7246	0.7258	0.7285	0.7288	0.7291	0.7267	0.7288	0.7304	0.7305	0.7300
	2017	0.7287	0.7313	0.7327	0.7330	0.7371	0.7324	0.7332	0.7338	0.7324	0.7370
	2018	0.6640	0.6657	0.6679	0.6672	0.6684	0.6688	0.6671	0.6688	0.6671	0.6702
	2019	0.7290	0.7292	0.7315	0.7304	0.7330	0.7309	0.7313	0.7283	0.7310	0.7320
	2020	0.6629	0.6651	0.6661	0.6649	0.6698	0.6695	0.6700	0.6686	0.6713	0.6721
	2021	0.7031	0.7066	0.7051	0.7061	0.7100	0.7082	0.7081	0.7107	0.7125	0.7125
	2022	0.6540	0.6556	0.6598	0.6593	0.6616	0.6613	0.6617	0.6608	0.6594	0.6631

Table S7. Results between our model and the baseline models @64 ("Extended TempEL: New entities" as the training set).

	Train \ Test	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
BLINK	2013	0.6831	0.6858	0.6879	0.6850	0.6861	0.6867	0.6840	0.6834	0.6809	0.6832
	2014	0.6499	0.6526	0.6516	0.6508	0.6523	0.6524	0.6495	0.6504	0.6471	0.6492
	2015	0.6718	0.6747	0.6761	0.6746	0.6738	0.6729	0.6721	0.6705	0.6696	0.6705
	2016	0.6745	0.6774	0.6775	0.6763	0.6784	0.6752	0.6747	0.6759	0.6743	0.6751
	2017	0.6845	0.6867	0.6873	0.6850	0.6876	0.6876	0.6843	0.6849	0.6842	0.6859
	2018	0.6583	0.6603	0.6617	0.6574	0.6600	0.6595	0.6582	0.6571	0.6554	0.6565
	2019	0.6934	0.6954	0.6954	0.6949	0.6967	0.6953	0.6945	0.6921	0.6934	0.6939
	2020	0.6642	0.6672	0.6676	0.6662	0.6676	0.6663	0.6658	0.6654	0.6632	0.6654
	2021	0.6681	0.6702	0.6711	0.6708	0.6738	0.6730	0.6726	0.6713	0.6762	0.6736
	2022	0.6639	0.6677	0.6690	0.6677	0.6705	0.6709	0.6691	0.6683	0.6707	0.6699
SpEL	2013	0.7531	0.7558	0.7479	0.7350	0.7361	0.7367	0.7440	0.7334	0.7309	0.7432
	2014	0.7199	0.7226	0.7216	0.7208	0.7223	0.7024	0.7195	0.7104	0.7071	0.7092
	2015	0.7218	0.7347	0.7261	0.7346	0.7238	0.7429	0.7221	0.7305	0.7296	0.7405
	2016	0.7445	0.7474	0.7375	0.7463	0.7484	0.7252	0.7447	0.7259	0.7443	0.7451
	2017	0.7345	0.7567	0.7573	0.7450	0.7476	0.7376	0.7443	0.7549	0.7542	0.7459
	2018	0.7083	0.7303	0.7117	0.7174	0.7100	0.7095	0.7182	0.7071	0.7054	0.7265
	2019	0.7634	0.7554	0.7654	0.7549	0.7467	0.7553	0.7645	0.7621	0.7634	0.7539
	2020	0.7242	0.7172	0.7376	0.7262	0.7176	0.7163	0.7258	0.7254	0.7132	0.7154
	2021	0.7381	0.7402	0.7211	0.7208	0.7238	0.7430	0.7326	0.7313	0.7362	0.7236
	2022	0.6939	0.6977	0.7090	0.7177	0.7205	0.7209	0.6991	0.7183	0.7007	0.6999
TIGER	2013	0.8149	0.8136	0.8142	0.8136	0.8150	0.8132	0.8127	0.8136	0.8152	0.8169
	2014	0.7751	0.7770	0.7781	0.7785	0.7787	0.7784	0.7789	0.7804	0.7797	0.7839
	2015	0.8113	0.8149	0.8145	0.8146	0.8155	0.8129	0.8118	0.8137	0.8139	0.8156
	2016	0.8048	0.8073	0.8106	0.8099	0.8121	0.8096	0.8099	0.8105	0.8101	0.8100
	2017	0.8053	0.8068	0.8102	0.8096	0.8129	0.8086	0.8079	0.8098	0.8096	0.8128
	2018	0.7523	0.7557	0.7575	0.7575	0.7605	0.7577	0.7571	0.7588	0.7575	0.7607
	2019	0.8096	0.8106	0.8125	0.8111	0.8125	0.8110	0.8110	0.8090	0.8112	0.8130
	2020	0.7597	0.7649	0.7642	0.7642	0.7663	0.7677	0.7679	0.7659	0.7671	0.7684
	2021	0.7861	0.7882	0.7898	0.7878	0.7927	0.7899	0.7904	0.7926	0.7948	0.7955
	2022	0.7513	0.7544	0.7569	0.7571	0.7579	0.7578	0.7581	0.7589	0.7582	0.7616

4. ADDITIONAL RESULTS - CONTINUAL ENTITIES TRAINING SET

Tables S8 through Table S14 present a comparison between the baseline models (BLINK and SpEL) and our model TIGER. The dataset, "Extended TempEL: Continual entities," has a training set (1,764), a validation set ($\approx 42k$, same as original TempEL dataset), and a test set ($\approx 48k$, same as original TempEL dataset). Here, the training set contains only 1,764 samples because, in the original TempEL dataset, each year's dataset contains only 1,764 'new entities' samples. For comparison, we also have 1,764 'continual entity' samples. Rows represent training datasets, while columns represent testing datasets. For example, the number at the intersection of the first row and the tenth column in Table S8 is 0.1741, showing the model's performance when trained on 2013 data and tested on 2022 data. Performance evaluations for all models were based on recall metrics, specifically @1, @2, @4, @8, @16, @32, and @64.

Table S8. Results between our model and the baseline models @1 ("Extended TempEL: Continual entities" as the training set).

	Train \ Test	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
BLINK	2013	0.1710	0.1719	0.1712	0.1727	0.1719	0.1736	0.1701	0.1761	0.1743	0.1741
	2014	0.1585	0.1593	0.1598	0.1613	0.1608	0.1587	0.1584	0.1625	0.1621	0.1620
	2015	0.1854	0.1870	0.1878	0.1879	0.1897	0.1887	0.1855	0.1913	0.1898	0.1909
	2016	0.1581	0.1589	0.1579	0.1595	0.1609	0.1602	0.1575	0.1614	0.1597	0.1589
	2017	0.1926	0.1953	0.1950	0.1947	0.1961	0.1944	0.1901	0.1984	0.1952	0.1944
	2018	0.1797	0.1782	0.1779	0.1786	0.1774	0.1799	0.1760	0.1833	0.1817	0.1801
	2019	0.1757	0.1767	0.1766	0.1763	0.1775	0.1775	0.1729	0.1785	0.1783	0.1774
	2020	0.1841	0.1867	0.1873	0.1869	0.1890	0.1874	0.1843	0.1901	0.1875	0.1866
	2021	0.1759	0.1752	0.1742	0.1760	0.1769	0.1754	0.1735	0.1797	0.1771	0.1758
	2022	0.1623	0.1641	0.1645	0.1666	0.1673	0.1656	0.1634	0.1686	0.1665	0.1665
SpEL	2013	0.2275	0.2133	0.2269	0.2133	0.2296	0.2237	0.2130	0.2284	0.2153	0.2264
	2014	0.2039	0.2160	0.2065	0.2179	0.2095	0.2091	0.2041	0.2191	0.2078	0.2213
	2015	0.2409	0.2430	0.2337	0.2331	0.2356	0.2301	0.2298	0.2373	0.2493	0.2430
	2016	0.1983	0.2124	0.2165	0.2188	0.2151	0.2044	0.2011	0.2197	0.2173	0.2166
	2017	0.2446	0.2447	0.2418	0.2347	0.2516	0.2346	0.2317	0.2421	0.2414	0.2391
	2018	0.2205	0.2342	0.2353	0.2291	0.2281	0.2271	0.2358	0.2243	0.2335	0.2342
	2019	0.2268	0.2356	0.2246	0.2285	0.2305	0.2278	0.2234	0.2275	0.2253	0.2209
	2020	0.2362	0.2426	0.2341	0.2348	0.2473	0.2391	0.2309	0.2496	0.2457	0.2314
	2021	0.2302	0.2251	0.2185	0.2218	0.2170	0.2312	0.2190	0.2315	0.2217	0.2279
	2022	0.2199	0.2187	0.2088	0.2223	0.2126	0.2240	0.2201	0.2191	0.2234	0.2193
TIGER	2013	0.2516	0.2522	0.2499	0.2516	0.2506	0.2512	0.2499	0.2500	0.2515	0.2519
	2014	0.2689	0.2696	0.2699	0.2707	0.2727	0.2713	0.2690	0.2733	0.2710	0.2684
	2015	0.2779	0.2805	0.2790	0.2788	0.2835	0.2797	0.2762	0.2777	0.2740	0.2764
	2016	0.2968	0.2963	0.2964	0.2991	0.3015	0.2969	0.2937	0.2985	0.2976	0.2962
	2017	0.3127	0.3132	0.3081	0.3099	0.3098	0.3068	0.3046	0.3078	0.3059	0.3031
	2018	0.2919	0.2930	0.2932	0.2940	0.2973	0.2920	0.2917	0.2953	0.2943	0.2939
	2019	0.2992	0.3043	0.3066	0.3051	0.3044	0.3011	0.3017	0.3035	0.3053	0.3038
	2020	0.2880	0.2928	0.2909	0.2914	0.2942	0.2908	0.2890	0.2925	0.2895	0.2894
	2021	0.2790	0.2804	0.2808	0.2836	0.2857	0.2830	0.2812	0.2842	0.2843	0.2822
	2022	0.2831	0.2875	0.2865	0.2868	0.2867	0.2814	0.2783	0.2803	0.2795	0.2801

Table S9. Results between our model and the baseline models @2 ("Extended TempEL: Continual entities" as the training set).

	Train \ Test	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
BLINK	2013	0.2534	0.2539	0.2531	0.2556	0.2549	0.2540	0.2514	0.2575	0.2566	0.2551
	2014	0.2363	0.2358	0.2381	0.2384	0.2400	0.2381	0.2349	0.2389	0.2400	0.2397
	2015	0.2689	0.2736	0.2726	0.2742	0.2760	0.2739	0.2710	0.2775	0.2771	0.2750
	2016	0.2332	0.2373	0.2357	0.2364	0.2386	0.2375	0.2325	0.2387	0.2375	0.2361
	2017	0.2804	0.2809	0.2831	0.2816	0.2850	0.2824	0.2805	0.2871	0.2858	0.2846
	2018	0.2605	0.2610	0.2613	0.2613	0.2625	0.2640	0.2581	0.2643	0.2651	0.2649
	2019	0.2605	0.2592	0.2612	0.2583	0.2624	0.2593	0.2565	0.2626	0.2633	0.2630
	2020	0.2701	0.2716	0.2727	0.2726	0.2743	0.2733	0.2673	0.2739	0.2722	0.2730
	2021	0.2581	0.2586	0.2596	0.2581	0.2612	0.2595	0.2550	0.2608	0.2624	0.2598
	2022	0.2423	0.2430	0.2441	0.2443	0.2474	0.2453	0.2422	0.2480	0.2481	0.2456
SpEL	2013	0.3104	0.3087	0.3132	0.3079	0.3209	0.3072	0.3185	0.3095	0.3196	0.3124
	2014	0.3011	0.2988	0.3049	0.3010	0.3093	0.2942	0.2940	0.3030	0.3000	0.3092
	2015	0.3351	0.3357	0.3244	0.3274	0.3271	0.3306	0.3245	0.3379	0.3283	0.3390
	2016	0.2859	0.3067	0.2954	0.2968	0.2962	0.2891	0.2907	0.2988	0.3056	0.2975
	2017	0.3393	0.3447	0.3508	0.3501	0.3545	0.3354	0.3348	0.3483	0.3462	0.3481
	2018	0.3151	0.3150	0.3154	0.3250	0.3212	0.3218	0.3094	0.3315	0.3268	0.3310
	2019	0.3107	0.3273	0.3219	0.3168	0.3134	0.3108	0.3142	0.3305	0.3247	0.3221
	2020	0.3225	0.3389	0.3231	0.3324	0.3361	0.3334	0.3231	0.3403	0.3332	0.3297
	2021	0.3137	0.3171	0.3181	0.3168	0.3145	0.3261	0.3076	0.3225	0.3245	0.3204
	2022	0.3063	0.2943	0.3025	0.3079	0.3145	0.3075	0.2989	0.3157	0.3162	0.3006
TIGER	2013	0.3609	0.3603	0.3597	0.3610	0.3598	0.3582	0.3590	0.3607	0.3618	0.3617
	2014	0.3786	0.3790	0.3771	0.3808	0.3821	0.3785	0.3778	0.3832	0.3802	0.3795
	2015	0.3865	0.3900	0.3886	0.3890	0.3916	0.3898	0.3860	0.3900	0.3889	0.3881
	2016	0.4126	0.4140	0.4127	0.4149	0.4164	0.4125	0.4088	0.4141	0.4136	0.4136
	2017	0.4286	0.4302	0.4284	0.4272	0.4307	0.4245	0.4248	0.4309	0.4282	0.4268
	2018	0.4072	0.4066	0.4087	0.4095	0.4112	0.4054	0.4048	0.4103	0.4073	0.4101
	2019	0.4189	0.4221	0.4230	0.4213	0.4250	0.4225	0.4210	0.4244	0.4244	0.4245
	2020	0.4018	0.4063	0.4044	0.4045	0.4072	0.4025	0.4017	0.4055	0.4048	0.4037
	2021	0.3918	0.3941	0.3948	0.3964	0.3994	0.3954	0.3933	0.3994	0.3976	0.3964
	2022	0.3934	0.3980	0.3976	0.3973	0.3963	0.3919	0.3880	0.3939	0.3921	0.3921

Table S10. Results between our model and the baseline models @4 ("Extended TempEL: Continual entities" as the training set).

	Train \ Test	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
BLINK	2013	0.3479	0.3500	0.3494	0.3500	0.3549	0.3510	0.3469	0.3536	0.3554	0.3522
	2014	0.3260	0.3289	0.3301	0.3305	0.3341	0.3302	0.3264	0.3337	0.3355	0.3327
	2015	0.3651	0.3680	0.3693	0.3721	0.3722	0.3691	0.3673	0.3712	0.3749	0.3716
	2016	0.3262	0.3266	0.3260	0.3277	0.3308	0.3287	0.3220	0.3287	0.3316	0.3280
	2017	0.3800	0.3813	0.3832	0.3836	0.3861	0.3835	0.3799	0.3854	0.3861	0.3830
	2018	0.3576	0.3570	0.3569	0.3585	0.3603	0.3600	0.3540	0.3596	0.3637	0.3593
	2019	0.3570	0.3567	0.3596	0.3567	0.3619	0.3582	0.3546	0.3604	0.3625	0.3624
	2020	0.3668	0.3697	0.3713	0.3729	0.3738	0.3716	0.3657	0.3722	0.3730	0.3724
	2021	0.3540	0.3567	0.3538	0.3549	0.3609	0.3565	0.3519	0.3575	0.3599	0.3575
	2022	0.3352	0.3362	0.3382	0.3374	0.3410	0.3391	0.3368	0.3418	0.3423	0.3410
SpEL	2013	0.4081	0.4103	0.4155	0.4107	0.4328	0.4125	0.4111	0.4206	0.4169	0.4320
	2014	0.3903	0.3985	0.3950	0.4009	0.4118	0.4048	0.4038	0.3974	0.4065	0.3960
	2015	0.4395	0.4365	0.4351	0.4401	0.4473	0.4354	0.4365	0.4450	0.4357	0.4335
	2016	0.3889	0.3933	0.4010	0.3998	0.4024	0.4024	0.4002	0.3932	0.3995	0.3880
	2017	0.4458	0.4593	0.4531	0.4577	0.4601	0.4485	0.4505	0.4622	0.4519	0.4462
	2018	0.4341	0.4215	0.4361	0.4352	0.4353	0.4210	0.4330	0.4267	0.4292	0.4361
	2019	0.4274	0.4309	0.4353	0.4343	0.4387	0.4211	0.4215	0.4360	0.4268	0.4294
	2020	0.4431	0.4424	0.4447	0.4348	0.4481	0.4413	0.4370	0.4437	0.4471	0.4332
	2021	0.4297	0.4172	0.4196	0.4165	0.4373	0.4262	0.4137	0.4367	0.4350	0.4237
	2022	0.4065	0.4143	0.3991	0.4023	0.4109	0.4126	0.4133	0.4084	0.4107	0.4155
TIGER	2013	0.4768	0.4750	0.4748	0.4755	0.4752	0.4746	0.4742	0.4763	0.4764	0.4777
	2014	0.4901	0.4927	0.4926	0.4927	0.4909	0.4891	0.4872	0.4921	0.4903	0.4913
	2015	0.4997	0.5042	0.5037	0.5023	0.5047	0.5017	0.4996	0.5041	0.5036	0.5025
	2016	0.5271	0.5318	0.5305	0.5309	0.5329	0.5277	0.5256	0.5312	0.5285	0.5297
	2017	0.5441	0.5473	0.5474	0.5463	0.5479	0.5441	0.5424	0.5487	0.5457	0.5442
	2018	0.5215	0.5232	0.5236	0.5256	0.5271	0.5231	0.5200	0.5259	0.5230	0.5256
	2019	0.5374	0.5428	0.5439	0.5397	0.5417	0.5377	0.5364	0.5423	0.5424	0.5426
	2020	0.5180	0.5218	0.5192	0.5176	0.5189	0.5138	0.5155	0.5201	0.5175	0.5174
	2021	0.5073	0.5093	0.5086	0.5093	0.5112	0.5081	0.5067	0.5141	0.5118	0.5133
	2022	0.5093	0.5126	0.5137	0.5117	0.5125	0.5097	0.5038	0.5083	0.5093	0.5100

Table S11. Results between our model and the baseline models @8 ("Extended TempEL: Continual entities" as the training set).

	Train \ Test	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
BLINK	2013	0.4531	0.4544	0.4538	0.4565	0.4586	0.4550	0.4531	0.4574	0.4609	0.4564
	2014	0.4314	0.4327	0.4322	0.4347	0.4385	0.4338	0.4314	0.4360	0.4409	0.4393
	2015	0.4700	0.4721	0.4723	0.4749	0.4774	0.4749	0.4714	0.4745	0.4776	0.4761
	2016	0.4258	0.4285	0.4296	0.4306	0.4354	0.4316	0.4260	0.4296	0.4363	0.4317
	2017	0.4863	0.4887	0.4880	0.4906	0.4918	0.4902	0.4886	0.4898	0.4931	0.4905
	2018	0.4616	0.4638	0.4600	0.4639	0.4656	0.4624	0.4604	0.4620	0.4657	0.4642
	2019	0.4614	0.4663	0.4668	0.4658	0.4689	0.4670	0.4627	0.4650	0.4698	0.4695
	2020	0.4719	0.4753	0.4762	0.4772	0.4790	0.4766	0.4730	0.4776	0.4785	0.4795
	2021	0.4551	0.4591	0.4594	0.4602	0.4640	0.4611	0.4586	0.4609	0.4651	0.4640
	2022	0.4396	0.4421	0.4409	0.4433	0.4454	0.4428	0.4391	0.4442	0.4473	0.4455
SpEL	2013	0.5321	0.5425	0.5336	0.5353	0.5355	0.5303	0.5268	0.5385	0.5376	0.5383
	2014	0.5158	0.5082	0.5131	0.5224	0.5201	0.5120	0.5027	0.5133	0.5236	0.5185
	2015	0.5572	0.5510	0.5449	0.5557	0.5642	0.5459	0.5607	0.5482	0.5627	0.5588
	2016	0.5118	0.5015	0.5160	0.5091	0.5144	0.5134	0.5124	0.4999	0.5228	0.5168
	2017	0.5704	0.5785	0.5739	0.5770	0.5805	0.5718	0.5753	0.5598	0.5661	0.5720
	2018	0.5337	0.5441	0.5473	0.5364	0.5554	0.5361	0.5482	0.5411	0.5387	0.5428
	2019	0.5331	0.5535	0.5431	0.5415	0.5478	0.5497	0.5485	0.5466	0.5470	0.5481
	2020	0.5589	0.5637	0.5597	0.5614	0.5648	0.5559	0.5433	0.5519	0.5523	0.5617
	2021	0.5252	0.5304	0.5374	0.5456	0.5390	0.5418	0.5325	0.5314	0.5522	0.5344
	2022	0.5105	0.5245	0.5150	0.5320	0.5258	0.5162	0.5221	0.5164	0.5350	0.5308
TIGER	2013	0.5888	0.5910	0.5876	0.5890	0.5910	0.5899	0.5880	0.5888	0.5886	0.5917
	2014	0.5973	0.6006	0.5999	0.6011	0.6004	0.5977	0.5970	0.6016	0.6003	0.6005
	2015	0.6069	0.6105	0.6114	0.6103	0.6113	0.6091	0.6071	0.6100	0.6094	0.6120
	2016	0.6353	0.6379	0.6378	0.6381	0.6408	0.6357	0.6345	0.6398	0.6381	0.6400
	2017	0.6513	0.6519	0.6522	0.6504	0.6514	0.6495	0.6494	0.6511	0.6486	0.6498
	2018	0.6289	0.6297	0.6290	0.6315	0.6340	0.6287	0.6276	0.6309	0.6285	0.6327
	2019	0.6453	0.6480	0.6508	0.6452	0.6487	0.6455	0.6455	0.6485	0.6476	0.6517
	2020	0.6253	0.6252	0.6252	0.6252	0.6257	0.6243	0.6232	0.6279	0.6255	0.6271
	2021	0.6138	0.6164	0.6149	0.6152	0.6189	0.6162	0.6159	0.6228	0.6194	0.6229
	2022	0.6195	0.6233	0.6243	0.6214	0.6207	0.6203	0.6166	0.6196	0.6173	0.6206

Table S12. Results between our model and the baseline models @16 ("Extended TempEL: Continual entities" as the training set).

	Train \ Test	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
BLINK	2013	0.5589	0.5606	0.5615	0.5625	0.5653	0.5625	0.5614	0.5622	0.5683	0.5640
	2014	0.5393	0.5424	0.5436	0.5450	0.5473	0.5416	0.5414	0.5443	0.5490	0.5457
	2015	0.5774	0.5790	0.5787	0.5801	0.5813	0.5820	0.5821	0.5814	0.5838	0.5829
	2016	0.5339	0.5377	0.5375	0.5372	0.5430	0.5396	0.5381	0.5367	0.5412	0.5389
	2017	0.5901	0.5945	0.5952	0.5970	0.5985	0.5970	0.5966	0.5966	0.5986	0.5970
	2018	0.5677	0.5707	0.5691	0.5710	0.5725	0.5695	0.5680	0.5700	0.5736	0.5705
	2019	0.5715	0.5742	0.5744	0.5752	0.5772	0.5737	0.5726	0.5740	0.5774	0.5782
	2020	0.5779	0.5806	0.5835	0.5842	0.5856	0.5817	0.5819	0.5839	0.5840	0.5849
	2021	0.5631	0.5665	0.5662	0.5682	0.5728	0.5684	0.5667	0.5680	0.5718	0.5695
	2022	0.5465	0.5499	0.5498	0.5514	0.5524	0.5479	0.5469	0.5498	0.5529	0.5541
SpEL	2013	0.6296	0.6228	0.6277	0.6404	0.6316	0.6352	0.6397	0.6279	0.6365	0.6422
	2014	0.6073	0.6098	0.6121	0.6203	0.6121	0.6203	0.6033	0.6203	0.6277	0.6063
	2015	0.6480	0.6492	0.6482	0.6405	0.6443	0.6454	0.6429	0.6609	0.6579	0.6518
	2016	0.5947	0.6009	0.5981	0.6037	0.6130	0.6174	0.5982	0.6100	0.6082	0.6008
	2017	0.6690	0.6680	0.6662	0.6626	0.6731	0.6720	0.6645	0.6729	0.6620	0.6600
	2018	0.6462	0.6472	0.6292	0.6368	0.6415	0.6393	0.6377	0.6495	0.6531	0.6395
	2019	0.6370	0.6438	0.6362	0.6397	0.6407	0.6443	0.6516	0.6406	0.6494	0.6516
	2020	0.6539	0.6515	0.6518	0.6566	0.6529	0.6423	0.6565	0.6622	0.6599	0.6627
	2021	0.6354	0.6291	0.6351	0.6413	0.6341	0.6355	0.6310	0.6330	0.6371	0.6336
	2022	0.6191	0.6215	0.6271	0.6283	0.6156	0.6164	0.6094	0.6146	0.6266	0.6277
TIGER	2013	0.6926	0.6948	0.6924	0.6930	0.6952	0.6925	0.6917	0.6939	0.6926	0.6975
	2014	0.6973	0.6975	0.6990	0.6989	0.6996	0.6970	0.6960	0.6988	0.6984	0.7019
	2015	0.7074	0.7094	0.7094	0.7082	0.7124	0.7069	0.7064	0.7080	0.7074	0.7093
	2016	0.7315	0.7321	0.7339	0.7316	0.7360	0.7320	0.7313	0.7339	0.7334	0.7352
	2017	0.7387	0.7378	0.7401	0.7378	0.7405	0.7381	0.7379	0.7399	0.7370	0.7382
	2018	0.7220	0.7249	0.7236	0.7243	0.7262	0.7229	0.7235	0.7229	0.7217	0.7251
	2019	0.7400	0.7405	0.7434	0.7397	0.7425	0.7406	0.7397	0.7423	0.7415	0.7444
	2020	0.7194	0.7187	0.7204	0.7186	0.7212	0.7200	0.7218	0.7232	0.7227	0.7249
	2021	0.7135	0.7133	0.7127	0.7121	0.7171	0.7150	0.7153	0.7182	0.7156	0.7196
	2022	0.7167	0.7193	0.7178	0.7164	0.7183	0.7187	0.7163	0.7181	0.7154	0.7189

Table S13. Results between our model and the baseline models @32 ("Extended TempEL: Continual entities" as the training set).

	Train \ Test	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
BLINK	2013	0.6631	0.6651	0.6671	0.6682	0.6688	0.6688	0.6648	0.6659	0.6687	0.6684
	2014	0.6459	0.6524	0.6515	0.6515	0.6548	0.6508	0.6492	0.6531	0.6553	0.6542
	2015	0.6797	0.6825	0.6840	0.6839	0.6849	0.6837	0.6837	0.6830	0.6851	0.6849
	2016	0.6425	0.6452	0.6438	0.6447	0.6471	0.6443	0.6442	0.6435	0.6463	0.6465
	2017	0.6912	0.6955	0.6975	0.6987	0.6994	0.6982	0.6986	0.6986	0.6993	0.6982
	2018	0.6693	0.6737	0.6720	0.6731	0.6752	0.6724	0.6710	0.6708	0.6749	0.6735
	2019	0.6743	0.6789	0.6807	0.6805	0.6815	0.6782	0.6795	0.6797	0.6820	0.6826
	2020	0.6813	0.6829	0.6849	0.6855	0.6878	0.6858	0.6849	0.6860	0.6862	0.6884
	2021	0.6658	0.6703	0.6698	0.6703	0.6734	0.6716	0.6697	0.6718	0.6745	0.6734
	2022	0.6536	0.6560	0.6570	0.6568	0.6600	0.6565	0.6565	0.6590	0.6604	0.6607
SpEL	2013	0.7148	0.7248	0.7236	0.7270	0.7261	0.7289	0.7243	0.7341	0.7297	0.7208
	2014	0.6996	0.7158	0.7155	0.7027	0.7214	0.7140	0.6999	0.7036	0.7099	0.7120
	2015	0.7368	0.7446	0.7458	0.7411	0.7386	0.7526	0.7407	0.7389	0.7507	0.7480
	2016	0.6936	0.7022	0.7054	0.7133	0.7033	0.7053	0.7124	0.7107	0.6975	0.7066
	2017	0.7476	0.7583	0.7606	0.7633	0.7533	0.7541	0.7610	0.7547	0.7679	0.7502
	2018	0.7233	0.7296	0.7247	0.7251	0.7279	0.7381	0.7298	0.7248	0.7370	0.7392
	2019	0.7251	0.7358	0.7317	0.7326	0.7473	0.7425	0.7371	0.7399	0.7370	0.7407
	2020	0.7404	0.7517	0.7484	0.7371	0.7494	0.7548	0.7408	0.7476	0.7531	0.7506
	2021	0.7315	0.7322	0.7204	0.7231	0.7241	0.7253	0.7241	0.7407	0.7284	0.7246
	2022	0.7202	0.7215	0.7148	0.7132	0.7189	0.7104	0.7106	0.7092	0.7253	0.7137
TIGER	2013	0.7819	0.7833	0.7826	0.7815	0.7858	0.7820	0.7812	0.7823	0.7814	0.7860
	2014	0.7854	0.7854	0.7879	0.7860	0.7882	0.7839	0.7857	0.7869	0.7853	0.7895
	2015	0.7909	0.7922	0.7937	0.7920	0.7949	0.7911	0.7908	0.7919	0.7916	0.7918
	2016	0.8122	0.8129	0.8157	0.8138	0.8166	0.8126	0.8144	0.8146	0.8141	0.8159
	2017	0.8138	0.8142	0.8165	0.8138	0.8163	0.8140	0.8136	0.8154	0.8111	0.8139
	2018	0.8024	0.8025	0.8042	0.8026	0.8071	0.8042	0.8039	0.8030	0.8035	0.8075
	2019	0.8171	0.8176	0.8193	0.8175	0.8220	0.8190	0.8198	0.8202	0.8184	0.8207
	2020	0.8001	0.8002	0.8010	0.7992	0.8039	0.8007	0.8035	0.8033	0.8023	0.8051
	2021	0.7967	0.7974	0.8000	0.7979	0.8040	0.7994	0.7993	0.8019	0.7997	0.8031
	2022	0.7974	0.7990	0.7985	0.7980	0.8013	0.7990	0.7989	0.8010	0.7984	0.8037

Table S14. Results between our model and the baseline models @64 ("Extended TempEL: Continual entities" as the training set).

	Train \ Test	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
BLINK	2013	0.7594	0.7617	0.7625	0.7613	0.7628	0.7621	0.7614	0.7628	0.7651	0.7639
	2014	0.7489	0.7538	0.7532	0.7537	0.7546	0.7529	0.7517	0.7541	0.7551	0.7539
	2015	0.7731	0.7758	0.7768	0.7776	0.7766	0.7754	0.7749	0.7746	0.7780	0.7769
	2016	0.7421	0.7449	0.7419	0.7456	0.7441	0.7434	0.7425	0.7413	0.7441	0.7440
	2017	0.7840	0.7880	0.7890	0.7886	0.7880	0.7873	0.7864	0.7867	0.7899	0.7873
	2018	0.7650	0.7669	0.7660	0.7660	0.7668	0.7649	0.7641	0.7657	0.7688	0.7666
	2019	0.7695	0.7745	0.7759	0.7758	0.7764	0.7752	0.7749	0.7750	0.7770	0.7755
	2020	0.7742	0.7764	0.7765	0.7763	0.7788	0.7760	0.7767	0.7779	0.7788	0.7793
	2021	0.7621	0.7655	0.7659	0.7659	0.7669	0.7660	0.7640	0.7651	0.7675	0.7667
	2022	0.7506	0.7565	0.7560	0.7550	0.7588	0.7550	0.7556	0.7562	0.7583	0.7575
SpEL	2013	0.8156	0.8058	0.8220	0.8155	0.8169	0.8190	0.8204	0.8169	0.8144	0.8096
	2014	0.7936	0.8021	0.8042	0.8071	0.8033	0.8041	0.8011	0.7942	0.8085	0.8102
	2015	0.8169	0.8340	0.8209	0.8244	0.8253	0.8170	0.8341	0.8257	0.8338	0.8243
	2016	0.7976	0.7905	0.7877	0.7877	0.8029	0.7973	0.7997	0.7969	0.7891	0.7875
	2017	0.8438	0.8436	0.8469	0.8424	0.8422	0.8405	0.8463	0.8311	0.8405	0.8427
	2018	0.8223	0.8249	0.8197	0.8173	0.8131	0.8220	0.8046	0.8091	0.8136	0.8186
	2019	0.8173	0.8266	0.8273	0.8228	0.8275	0.8227	0.8322	0.8344	0.8310	0.8241
	2020	0.8172	0.8313	0.8325	0.8222	0.8337	0.8258	0.8237	0.8197	0.8381	0.8195
	2021	0.8099	0.8073	0.8132	0.8251	0.8130	0.8109	0.8213	0.8128	0.8117	0.8082
	2022	0.8000	0.8153	0.7979	0.8052	0.8024	0.8027	0.8145	0.8108	0.8160	0.8148
TIGER	2013	0.8558	0.8564	0.8556	0.8552	0.8571	0.8556	0.8554	0.8560	0.8557	0.8590
	2014	0.8568	0.8564	0.8574	0.8556	0.8582	0.8536	0.8555	0.8556	0.8545	0.8579
	2015	0.8597	0.8604	0.8614	0.8594	0.8643	0.8607	0.8600	0.8596	0.8585	0.8620
	2016	0.8761	0.8764	0.8783	0.8763	0.8807	0.8774	0.8780	0.8792	0.8782	0.8796
	2017	0.8750	0.8756	0.8766	0.8746	0.8767	0.8737	0.8741	0.8739	0.8723	0.8747
	2018	0.8680	0.8687	0.8702	0.8706	0.8739	0.8696	0.8694	0.8692	0.8678	0.8727
	2019	0.8784	0.8786	0.8796	0.8781	0.8802	0.8786	0.8808	0.8797	0.8788	0.8806
	2020	0.8657	0.8666	0.8671	0.8652	0.8686	0.8664	0.8679	0.8685	0.8668	0.8693
	2021	0.8656	0.8668	0.8686	0.8664	0.8686	0.8669	0.8677	0.8682	0.8662	0.8688
	2022	0.8652	0.8685	0.8672	0.8654	0.8683	0.8663	0.8652	0.8665	0.8654	0.8679

5. ADDITIONAL RESULTS - MITIGATING TEMPORAL DEGRADATION

Table S15 to Table S18 illuminate the effectiveness of our model in mitigating temporal degradation using results derived from the expanded TempEL dataset. In Table S15 and Table S16, we use "Extended TempEL: New entities" as the training set. In Table S17 and Table S18, we use "Extended TempEL: Continual entities" as the training set. Each column in the table represents the years' gap between the training and testing datasets, as denoted by the digits from 0 to 9. For instance, 0 implies that training and testing datasets come from the same year, while 9 indicates that the model was trained in 2013 and tested in 2022. The rows are divided based on various metrics, includes @1, @2, @4, @8, @16, @32 and @64. "Boost" displays a comparison between our model TIGER and SpEL model, calculated as Boost $= \frac{\text{Our Model's Result} - \text{Baseline Model's Result}}{\text{Baseline Model's Result}}$.

Table S15 and Table S17 show models' performance on the "Only Forward" setting. "Only Forward" means the results include only one scenario: the model is trained in the past and tested in the future. Table S16 and Table S18 show models' performance on the "Forward and Backward" setting. "Forward and Backward" means the results include two scenarios: training in the past and testing in the future, and training in the future and testing in the past. For instance, in Table S15, when @1 and the gap is 9, our model's result under "Only Forward" is 0.2304, representing only the situation of training in 2013 and testing in 2022. However, in Table S16 under "Forward and Backward," our model's result is 0.1914, which is the average of two situations: training in 2013 and testing in 2022, and training in 2022 and testing in 2013. It is worth noting that when the gap is 0, it means the training and testing datasets come from the same year. Thus, the "Only Forward" and "Forward and Backward" values are equal.

Table S15. Temporal degradation mitigation performance on the "Only Forward" setting using "Extended TempEL: New entities" as training dataset.

		0	1	2	3	4	5	6	7	8	9
@1	BLINK	0.1333	0.1340	0.1340	0.1357	0.1349	0.1360	0.1348	0.1357	0.1346	0.1433
	SpEL	0.1733	0.1695	0.1765	0.1771	0.1699	0.1720	0.1723	0.1790	0.1796	0.1933
	TIGER	0.2032	0.2076	0.2086	0.2163	0.2158	0.2232	0.2211	0.2221	0.2149	0.2304
	Boost (%)	17.23	22.46	18.22	22.11	27.04	29.76	28.32	24.04	19.66	19.19
@2	BLINK	0.1994	0.1999	0.2004	0.2021	0.2013	0.2027	0.2009	0.2014	0.1996	0.2117
	SpEL	0.2444	0.2455	0.2504	0.2535	0.2513	0.2447	0.2484	0.2481	0.2496	0.2617
	TIGER	0.2948	0.3009	0.3018	0.3114	0.3101	0.3201	0.3167	0.3185	0.3102	0.3301
	Boost (%)	20.65	22.55	20.54	22.82	23.41	30.82	27.46	28.39	24.30	26.14
@4	BLINK	0.2794	0.2789	0.2804	0.2825	0.2809	0.2837	0.2802	0.2813	0.2785	0.2952
	SpEL	0.3324	0.3411	0.3379	0.3396	0.3425	0.3457	0.3402	0.3446	0.3385	0.3652
	TIGER	0.3975	0.4042	0.4060	0.4161	0.4143	0.4250	0.4223	0.4244	0.4139	0.4376
	Boost (%)	19.56	18.50	20.16	22.52	20.96	22.96	24.13	23.16	22.27	19.82
@8	BLINK	0.3720	0.3717	0.3735	0.3752	0.3724	0.3745	0.3713	0.3713	0.3687	0.3883
	SpEL	0.4370	0.4439	0.4422	0.4437	0.4424	0.4445	0.4438	0.4380	0.4437	0.4483
	TIGER	0.5050	0.5122	0.5133	0.5234	0.5215	0.5326	0.5298	0.5311	0.5215	0.5453
	Boost (%)	15.57	15.39	16.07	17.95	17.88	19.82	19.37	21.26	17.53	21.64
@16	BLINK	0.4717	0.4716	0.4724	0.4737	0.4705	0.4728	0.4693	0.4681	0.4660	0.4848
	SpEL	0.5437	0.5427	0.5499	0.5452	0.5422	0.5448	0.5368	0.5415	0.5360	0.5448
	TIGER	0.6109	0.6163	0.6175	0.6257	0.6235	0.6353	0.6335	0.6334	0.6250	0.6489
	Boost (%)	12.35	13.56	12.29	14.77	15.00	16.60	18.01	16.97	16.60	19.11
@32	BLINK	0.5747	0.5745	0.5743	0.5751	0.5715	0.5737	0.5696	0.5678	0.5652	0.5855
	SpEL	0.6407	0.6412	0.6393	0.6451	0.6398	0.6397	0.6371	0.6345	0.6352	0.6555
	TIGER	0.7081	0.7122	0.7132	0.7188	0.7173	0.7284	0.7262	0.7254	0.7206	0.7401
	Boost (%)	10.51	11.08	11.55	11.42	12.11	13.86	13.99	14.33	13.45	12.91
@64	BLINK	0.6741	0.6739	0.6735	0.6742	0.6712	0.6734	0.6698	0.6670	0.6651	0.6832
	SpEL	0.7331	0.7350	0.7310	0.7370	0.7279	0.7354	0.7323	0.7270	0.7201	0.7432
	TIGER	0.7920	0.7951	0.7955	0.7993	0.7977	0.8057	0.8043	0.8030	0.7996	0.8169
	Boost (%)	8.03	8.17	8.83	8.46	9.59	9.57	9.83	10.45	11.04	9.92

Table S16. Temporal degradation mitigation performance on the "Forward and Backward" setting using "Extended TempEL: New entities" as training dataset.

		0	1	2	3	4	5	6	7	8	9
@1	BLINK	0.1333	0.1326	0.1331	0.1332	0.1329	0.1325	0.1321	0.1305	0.1306	0.1350
	SpEL	0.1733	0.1687	0.1737	0.1725	0.1688	0.1685	0.1709	0.1672	0.1731	0.1750
	TIGER	0.2032	0.2037	0.2040	0.2052	0.2028	0.2031	0.2018	0.1968	0.1960	0.1914
	Boost (%)	17.23	20.74	17.42	18.98	20.17	20.53	18.10	17.70	13.18	9.37
@2	BLINK	0.1994	0.1985	0.1992	0.1996	0.1990	0.1981	0.1973	0.1957	0.1955	0.2015
	SpEL	0.2444	0.2457	0.2461	0.2482	0.2465	0.2401	0.2423	0.2390	0.2430	0.2365
	TIGER	0.2948	0.2958	0.2961	0.2978	0.2944	0.2942	0.2926	0.2870	0.2866	0.2808
	Boost (%)	20.65	20.39	20.29	19.99	19.43	22.50	20.79	20.09	17.97	18.74
@4	BLINK	0.2794	0.2777	0.2793	0.2796	0.2787	0.2787	0.2771	0.2748	0.2741	0.2830
	SpEL	0.3324	0.3344	0.3356	0.3354	0.3404	0.3387	0.3321	0.3298	0.3266	0.3380
	TIGER	0.3975	0.3985	0.3991	0.4005	0.3970	0.3967	0.3959	0.3886	0.3880	0.3822
	Boost (%)	19.56	19.18	18.94	19.43	16.63	17.13	19.19	17.84	18.83	13.09
@8	BLINK	0.3720	0.3704	0.3721	0.3731	0.3713	0.3708	0.3692	0.3653	0.3652	0.3751
	SpEL	0.4370	0.4398	0.4421	0.4409	0.4388	0.4398	0.4354	0.4319	0.4277	0.4201
	TIGER	0.5050	0.5061	0.5063	0.5082	0.5046	0.5042	0.5042	0.4960	0.4957	0.4895
	Boost (%)	15.57	15.08	14.54	15.26	15.00	14.64	15.80	14.83	15.89	16.51
@16	BLINK	0.4717	0.4702	0.4720	0.4727	0.4709	0.4703	0.4689	0.4642	0.4644	0.4724
	SpEL	0.5437	0.5402	0.5414	0.5419	0.5392	0.5383	0.5364	0.5309	0.5294	0.5274
	TIGER	0.6109	0.6111	0.6116	0.6126	0.6088	0.6096	0.6105	0.6021	0.6015	0.5974
	Boost (%)	12.35	13.12	12.95	13.04	12.90	13.26	13.81	13.41	13.62	13.27
@32	BLINK	0.5747	0.5730	0.5745	0.5748	0.5729	0.5723	0.5711	0.5659	0.5657	0.5745
	SpEL	0.6407	0.6419	0.6420	0.6441	0.6412	0.6393	0.6336	0.6309	0.6332	0.6395
	TIGER	0.7081	0.7080	0.7087	0.7090	0.7055	0.7069	0.7079	0.7009	0.7000	0.6971
	Boost (%)	10.51	10.29	10.39	10.07	10.03	10.58	11.73	11.09	10.55	9.01
@64	BLINK	0.6741	0.6731	0.6741	0.6748	0.6732	0.6730	0.6723	0.6674	0.6665	0.6736
	SpEL	0.7331	0.7320	0.7316	0.7348	0.7307	0.7320	0.7311	0.7257	0.7190	0.7186
	TIGER	0.7920	0.7920	0.7925	0.7921	0.7895	0.7902	0.7923	0.7856	0.7849	0.7841
	Boost (%)	8.03	8.20	8.32	7.81	8.05	7.95	8.38	8.25	9.17	9.12

Table S17. Temporal degradation mitigation performance on the "Only Forward" setting using "Extended TempEL: Continual entities" as training dataset.

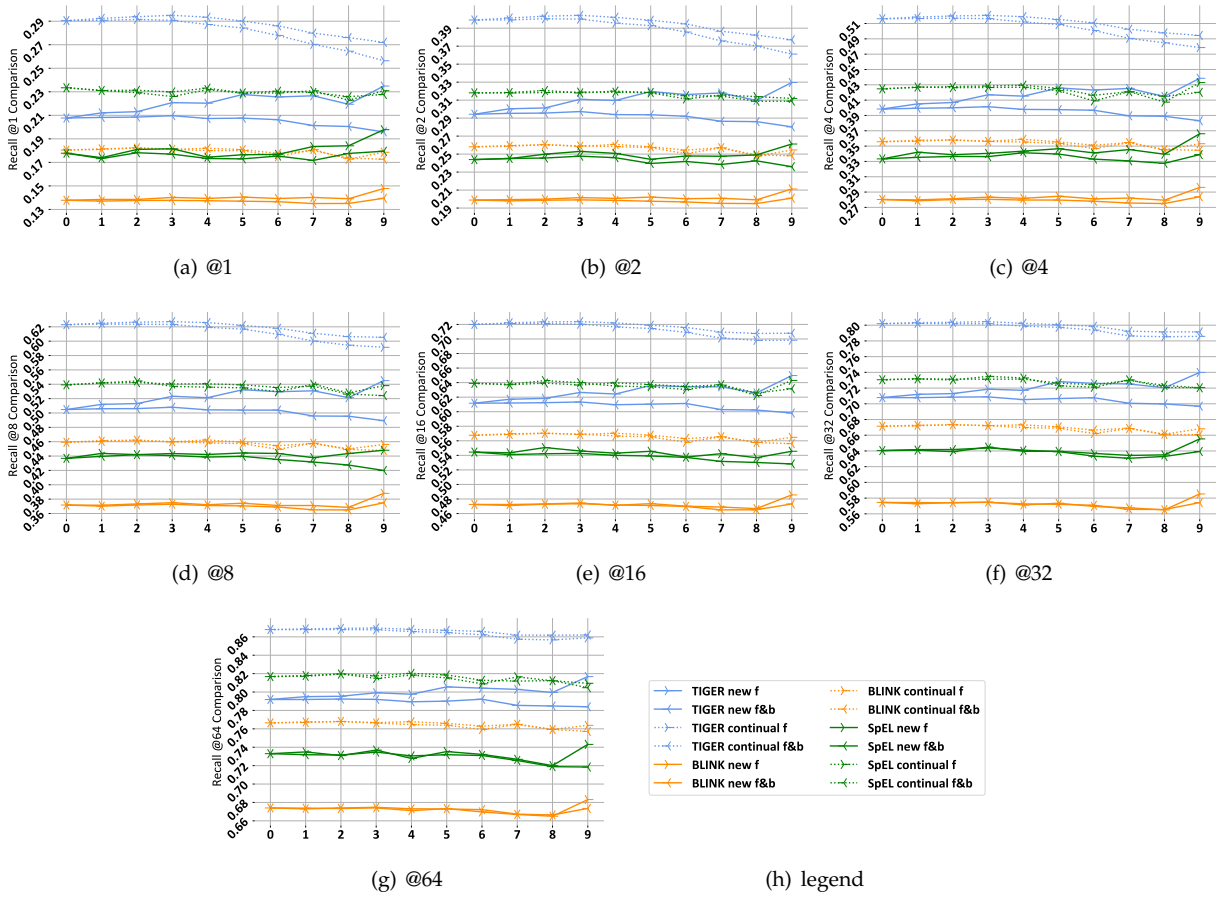
		0	1	2	3	4	5	6	7	8	9
@1	BLINK	0.1760	0.1770	0.1776	0.1767	0.1755	0.1755	0.1703	0.1764	0.1682	0.1741
	SpEL	0.2289	0.2266	0.2247	0.2215	0.2273	0.2243	0.2245	0.2264	0.2183	0.2264
	TIGER	0.2860	0.2862	0.2870	0.2862	0.2827	0.2797	0.2734	0.2658	0.2600	0.2519
	Boost (%)	24.95	26.31	27.71	29.22	24.39	24.71	21.76	17.40	19.08	11.26
@2	BLINK	0.2586	0.2600	0.2608	0.2596	0.2589	0.2577	0.2509	0.2575	0.2482	0.2551
	SpEL	0.3186	0.3185	0.3189	0.3194	0.3193	0.3186	0.3118	0.3162	0.3144	0.3124
	TIGER	0.3996	0.3997	0.4010	0.4006	0.3961	0.3933	0.3862	0.3763	0.3707	0.3617
	Boost (%)	25.40	25.53	25.74	25.44	24.07	23.46	23.84	19.03	17.89	15.78
@4	BLINK	0.3548	0.3568	0.3569	0.3552	0.3544	0.3526	0.3459	0.3536	0.3441	0.3522
	SpEL	0.4238	0.4262	0.4254	0.4256	0.4259	0.4214	0.4081	0.4202	0.4065	0.4320
	TIGER	0.5153	0.5156	0.5160	0.5154	0.5111	0.5077	0.4999	0.4897	0.4839	0.4777
	Boost (%)	21.59	20.96	21.29	21.12	20.00	20.48	22.51	16.54	19.04	10.58
@8	BLINK	0.4594	0.4617	0.4622	0.4601	0.4585	0.4575	0.4496	0.4581	0.4501	0.4564
	SpEL	0.5394	0.5421	0.5448	0.5372	0.5362	0.5352	0.5299	0.5403	0.5281	0.5383
	TIGER	0.6232	0.6240	0.6238	0.6235	0.6195	0.6170	0.6098	0.6004	0.5946	0.5917
	Boost (%)	15.54	15.11	14.50	16.06	15.54	15.28	15.07	11.12	12.59	9.92
@16	BLINK	0.5668	0.5689	0.5695	0.5683	0.5658	0.5647	0.5571	0.5647	0.5570	0.5640
	SpEL	0.6382	0.6369	0.6420	0.6391	0.6344	0.6335	0.6297	0.6358	0.6214	0.6422
	TIGER	0.7192	0.7205	0.7204	0.7195	0.7158	0.7136	0.7083	0.7005	0.6973	0.6975
	Boost (%)	12.69	13.12	12.21	12.59	12.83	12.64	12.48	10.18	12.21	8.61
@32	BLINK	0.6717	0.6729	0.6735	0.6724	0.6699	0.6691	0.6624	0.6687	0.6615	0.6684
	SpEL	0.7308	0.7318	0.7305	0.7351	0.7331	0.7231	0.7213	0.7307	0.7209	0.7208
	TIGER	0.8022	0.8026	0.8020	0.8021	0.7990	0.7975	0.7939	0.7865	0.7855	0.7860
	Boost (%)	9.77	9.67	9.80	9.11	8.98	10.29	10.07	7.64	8.96	9.05
@64	BLINK	0.7666	0.7676	0.7681	0.7664	0.7647	0.7640	0.7594	0.7649	0.7595	0.7639
	SpEL	0.8169	0.8181	0.8197	0.8149	0.8185	0.8155	0.8090	0.8166	0.8123	0.8096
	TIGER	0.8680	0.8680	0.8680	0.8678	0.8658	0.8647	0.8623	0.8575	0.8568	0.8590
	Boost (%)	6.25	6.10	5.90	6.49	5.78	6.03	6.59	5.01	5.48	6.10

Table S18. Temporal degradation mitigation performance on the "Forward and Backward" setting using "Extended TempEL: Continual entities" as training dataset.

		0	1	2	3	4	5	6	7	8	9
@1	BLINK	0.1760	0.1765	0.1779	0.1766	0.1775	0.1764	0.1731	0.1755	0.1691	0.1682
	SpEL	0.2289	0.2266	0.2268	0.2254	0.2286	0.2246	0.2260	0.2249	0.2214	0.2232
	TIGER	0.2860	0.2879	0.2893	0.2905	0.2889	0.2856	0.2816	0.2754	0.2716	0.2675
	Boost (%)	24.95	27.05	27.53	28.88	26.40	27.15	24.60	22.46	22.69	19.87
@2	BLINK	0.2586	0.2592	0.2610	0.2591	0.2613	0.2586	0.2549	0.2576	0.2494	0.2487
	SpEL	0.3186	0.3190	0.3212	0.3185	0.3205	0.3190	0.3154	0.3151	0.3092	0.3094
	TIGER	0.3996	0.4018	0.4038	0.4048	0.4026	0.3993	0.3953	0.3871	0.3828	0.3776
	Boost (%)	25.40	25.96	25.73	27.09	25.60	25.18	25.33	22.84	23.80	22.05
@4	BLINK	0.3548	0.3555	0.3572	0.3554	0.3580	0.3545	0.3502	0.3537	0.3446	0.3437
	SpEL	0.4238	0.4257	0.4267	0.4275	0.4286	0.4244	0.4155	0.4200	0.4142	0.4193
	TIGER	0.5153	0.5175	0.5189	0.5198	0.5180	0.5144	0.5099	0.5017	0.4969	0.4935
	Boost (%)	21.59	21.57	21.63	21.58	20.87	21.20	22.72	19.45	19.96	17.71
@8	BLINK	0.4594	0.4607	0.4623	0.4600	0.4626	0.4597	0.4547	0.4577	0.4494	0.4480
	SpEL	0.5394	0.5419	0.5425	0.5406	0.5409	0.5394	0.5357	0.5375	0.5265	0.5244
	TIGER	0.6232	0.6253	0.6267	0.6274	0.6261	0.6223	0.6182	0.6112	0.6066	0.6056
	Boost (%)	15.54	15.39	15.52	16.07	15.74	15.36	15.40	13.70	15.22	15.48
@16	BLINK	0.5668	0.5679	0.5696	0.5680	0.5696	0.5670	0.5623	0.5647	0.5568	0.5553
	SpEL	0.6382	0.6362	0.6385	0.6359	0.6388	0.6366	0.6338	0.6363	0.6249	0.6307
	TIGER	0.7192	0.7214	0.7227	0.7230	0.7213	0.7181	0.7151	0.7087	0.7068	0.7071
	Boost (%)	12.69	13.39	13.18	13.69	12.92	12.80	12.82	11.38	13.11	12.12
@32	BLINK	0.6717	0.6721	0.6737	0.6724	0.6734	0.6709	0.6667	0.6691	0.6612	0.6610
	SpEL	0.7308	0.7320	0.7315	0.7315	0.7316	0.7265	0.7245	0.7299	0.7237	0.7205
	TIGER	0.8022	0.8034	0.8039	0.8047	0.8026	0.8008	0.7989	0.7926	0.7917	0.7917
	Boost (%)	9.77	9.74	9.90	10.01	9.71	10.23	10.27	8.59	9.39	9.88
@64	BLINK	0.7666	0.7670	0.7681	0.7669	0.7678	0.7661	0.7630	0.7651	0.7594	0.7573
	SpEL	0.8169	0.8172	0.8197	0.8177	0.8204	0.8187	0.8129	0.8120	0.8125	0.8048
	TIGER	0.8680	0.8687	0.8693	0.8697	0.8682	0.8672	0.8660	0.8620	0.8619	0.8621
	Boost (%)	6.25	6.30	6.05	6.36	5.82	5.93	6.54	6.16	6.09	7.12

6. DIMINISHING PERFORMANCE GAINS

Figure S2 displays recall@N results from the Expanded TempEL dataset. We assessed our proposed model against the baseline bi-encoder using recall metrics. The x-axis indicates the year gap between training and testing sets, while the y-axis represents the recall rate. Overall, our model consistently outperforms the baseline models.



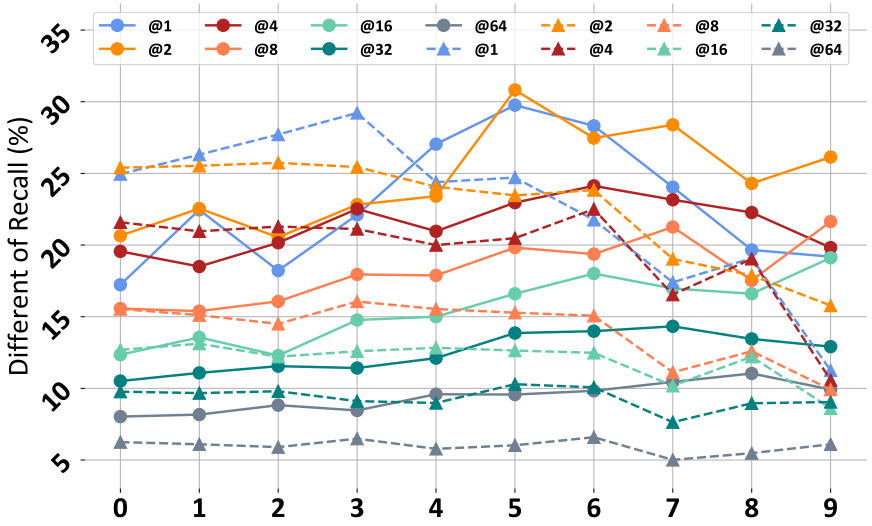
(a) @1 (b) @2 (c) @4

(d) @8 (e) @16 (f) @32

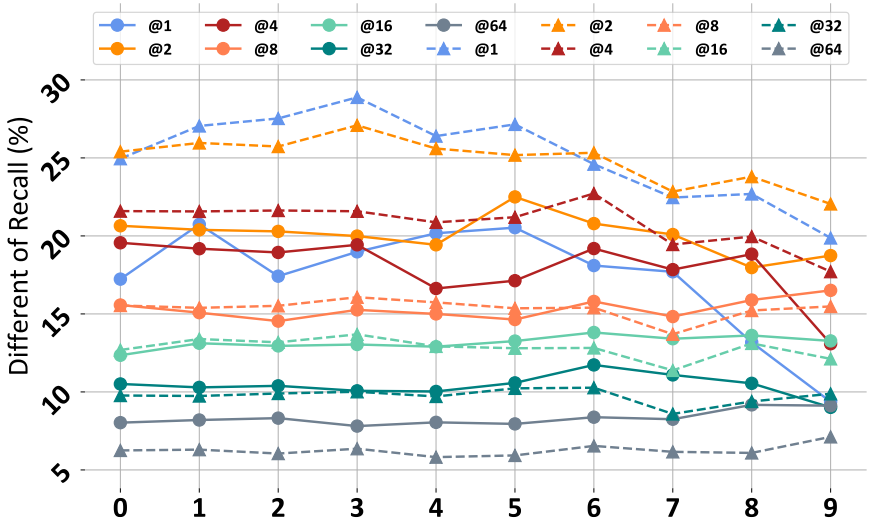
(g) @64 (h) legend

Fig. S2. The improvement in the metrics @1, @2, @4, @8, @16, @32 and @64.

It is also worth noting that the improvement effect of our model diminishes gradually as the metric threshold becomes more lenient, as shown in Figure S3. A plausible explanation for this observation is when using @64 threshold, the model only needs to correctly predict one out of the top 64 answers, allowing for a higher tolerance of errors. Consequently, the relative performance improvement of our model becomes less evident at these higher thresholds. This indicates a trade-off between prediction accuracy and the employed scale, suggesting the need for careful balance in practical applications.



(a) Only Forward



(b) Forward and Backward

Fig. S3. Different of Recall changes as the metric changes.