IELM: An Open Information Extraction Benchmark for Pre-Trained Language Models

Chenguang Wang[1], **Xiao Liu**[2], **Dawn Song**[3]
[1]Washington University in St. Louis, [2]Tsinghua University, [3]UC Berkeley
chenguangwang@wustl.edu, liuxiao21@mails.tsinghua.edu.cn,
dawnsong@berkeley.edu

Abstract

We introduce a new open information extraction (OIE) benchmark for pre-trained language models (LM). Recent studies have demonstrated that pre-trained LMs, such as BERT and GPT, may store linguistic and relational knowledge. In particular, LMs are able to answer "fill-in-the-blank" questions when given a pre-defined relation category. Instead of focusing on pre-defined relations, we create an OIE benchmark aiming to fully examine the open relational information present in the pre-trained LMs. We accomplish this by turning pre-trained LMs into zero-shot OIE systems. Surprisingly, pre-trained LMs are able to obtain competitive performance on both standard OIE datasets (CaRB and Re-OIE2016) and two new large-scale factual OIE datasets (TAC KBP-OIE and Wikidata-OIE) that we establish via distant supervision. For instance, the zero-shot pre-trained LMs outperform the F1 score of the state-of-the-art supervised OIE methods on our factual OIE datasets without needing to use any training sets.[1]

1 Introduction

Pre-trained language models (LM), such as BERT (Devlin et al., 2018) and GPT-3 (Brown et al., 2020), have revolutionized NLP over the last several years and advanced the state-of-the-art results in a wide set of downstream NLP tasks. Recent studies show that a considerable amount of linguistic (Hewitt and Manning, 2019; Clark et al., 2019) and relational knowledge (Petroni et al., 2019; Talmor et al., 2019; Jiang et al., 2020; Petroni et al., 2020) has been captured by the pre-trained LMs via pre-training on large-scale textual corpora. These approaches often design "fill-in-the-blank" questions based on pre-defined relations. For example, a question "Bob Dylan was born in _" is manually created for the LMs to answer the "birthplace" relation of "Bob Dylan".

Most existing approaches that evaluate what pre-trained LMs have learned are based on benchmarks with pre-defined relation categories. Yet, the benchmarks present two limitations. First, most benchmarks only cover a limited number of pre-defined relations. Therefore, it is unclear whether the pre-trained LMs have stored general open relation information. For example, the Google-RE in LAMA benchmark (Petroni et al., 2019) includes only three relations (i.e., "birthplace", "birthdate", and "deathplace"), while there are hundreds of relations available in the real world scenario. Second, a majority of benchmarks evaluate LMs in a close manner. This means that the gold relation is given to the models. For example, "was born in" is given as the model's input. Besides, the existing benchmarks often provide a single gold relation for each input sentence. However, an input sentence may indicate multiple relations, e.g., containing both "birthplace" and "birthdate" information about an argument or entity. We are curious: instead of the limited relational information, can we systematically benchmark the general information stored in the pre-trained LMs?

In this work, we set up a new open information extraction (OIE) benchmark, called IELM, towards testing the general and open relational information stored in pre-trained LMs. We refer to OIE as it is a task that is designed to extract open relations from massive corpora without requiring a pre-defined relation category. As shown in Figure 1, we successfully convert pre-trained LMs to zero-shot OIE systems. We apply them to two standard OIE datasets, including CaRB (Bhardwaj et al., 2019) and Re-OIE2016 (Stanovsky and Dagan, 2016; Zhan and Zhao, 2020), as well as two new large-scale factual OIE datasets in our IELM benchmark. We show that the zero-shot pre-trained LMs outperform the fully supervised state-of-the-arts on fac-

[1]Our code and datasets are available at https://github.com/cgraywang/IELM.

Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 8417–8437
December 7-11, 2022 ©2022 Association for Computational Linguistics

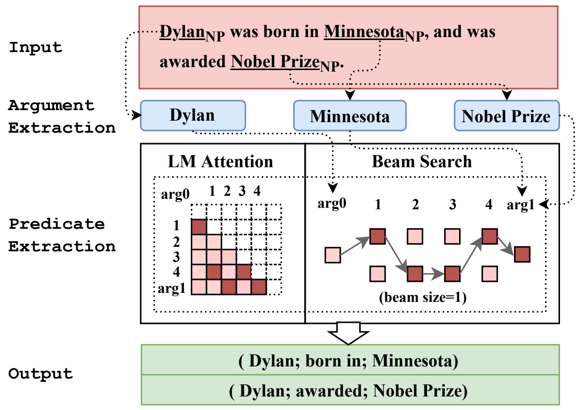


Figure 1: Summary of our approach. The zero-shot open information extraction system takes a noun phrase (NP) chunked sentence as input, and outputs a set of triples. The approach first conducts argument extraction by encoding NPs as argument pairs, then performs predicate extraction via decoding using the parameters (i.e., attention scores) of the pre-trained language models. The output extractions are then evaluated on our IELM benchmark.

tual OIE datasets. Standard OIE datasets rely on human annotations and often consist of thousands of gold triples and sentences. Unlike those datasets, we create two large-scale OIE datasets, namely TAC KBP-OIE and Wikidata-OIE, via distant supervision from knowledge graphs. For example, Wikidata-OIE is constructed via aligning English Wikipedia to Wikidata triples, resulting in millions of triples and documents. The design of zero-shot LMs for OIE is important: by encoding the noun chunks as arguments in the input, we only make use of the parameters of pre-trained LMs to decode the predicates (or relations) between the arguments. To the best of our knowledge, this is the first attempt to systematically evaluate pre-trained LMs in a zero-shot OIE setting. To summarize, our key contributions are the following.

1. We benchmark the general relational information in pre-trained LMs on our IELM benchmark. Besides two standard OIE datasets (CaRB and Re-OIE2016), we also create two large-scale factual OIE datasets for our benchmark. The two new OIE datasets are called TAC KBP-OIE and Wikidata-OIE, which are constructed via distant supervision from two knowledge graphs (TAC KBP and Wikidata). Our benchmark is a general OIE benchmark, helping the development of future OIE systems.

2. We enable the zero-shot capabilities of pre-trained LMs for OIE by encoding the arguments in the input and decoding predicates using the parameters of pre-trained LMs. The pre-trained LMs are particularly good at recovering factual arguments and predicates.

3. We test the OIE performance of 6 pre-trained LMs (BERT and GPT-2 (Radford et al., 2019) families) and 14 OIE systems on IELM benchmark. The zero-shot LMs achieve state-of-the-art OIE performance on TAC KBP-OIE and Wikidata-OIE, even outperforming fully supervised OIE systems.

2 Language Models as Zero-Shot Information Extractors

For open information extraction (OIE), we take an input as a NP-chunked sentence and output a set of triples. Below is an example.

> **Input** Dylan$_{\mathrm{NP}}$ was born in Minnesota$_{\mathrm{NP}}$, and was awarded Nobel Prize$_{\mathrm{NP}}$.
> **Output** (Dylan; born in; Minnesota), (Dylan; awarded; Nobel Prize).

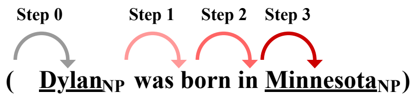
NP denotes the noun phrase.

2.1 Argument Extraction

Follow traditional linguistic OIE systems such as Stanford OpenIE (Angeli et al., 2015) and OpenIE5 (Saha et al., 2017, 2018), we treat each NP pair as an argument pair (e.g., "Dylan" and "Minnesota"). We then utilize the parameters of LMs to extract the predicates (e.g., "born in") between the pair in the input as below.

2.2 Predicate Extraction

The predicate extraction problem is formulated as extracting a set of sequences in the input that are associated with an argument pair. We particularly use the attention scores in a pre-trained LM to measure the relevance between a sequence and the argument pair. We frame the process as a search problem. Given an argument pair, we aim to search for the sequences with the largest attention scores connecting the pair. To compute a score for every possible sequence is computationally expensive especially when the sequence length is large, the exhaustive search is therefore intractable. We adapt beam search as an approximate strategy to efficiently explore the search space. Beam search maintains the k-best candidates. This means the time cost of beam search does not depend on the sequence length, but on the size of the beam and



Step	Action	Partial candidate	Total score
0	START	(Dylan;	0
1	YIELD	(Dylan; born	0.2
2	YIELD	(Dylan; born in;	0.5
3	STOP	(Dylan; born in; Minnesota)	0.7

(a) Predicate extraction example.

Key: / Query:	Dylan	was	born	in	Minnesota
Dylan	X	X	X	X	X
was	0.1	X	X	X	X
born	0.2	0.1	X	X	X
in	0.05	0.05	0.3	X	X
Minnesota	0.1	0.06	0.1	0.2	X

(b) Attention matrix.

Figure 2: Illustration of predicate extraction with a pre-trained language model (LM). The upper part of (a) represents the general search steps of producing the triple *(Dylan; born in; Minnesota)* from the input "Dylan$_{\text{NP}}$ was born in Minnesota$_{\text{NP}}$" encoded with argument noun phrases (NP). The lower portion shows the corresponding step-by-step process. (b) shows the attention scores generated through the forward pass of the LM over the corresponding input.

the average length of the candidates. In general, the beam search starts with the first argument (e.g., "Dylan"). At each step, beam search simply selects top-k next tokens with the largest attention scores, and just keeps k partial candidates with the highest scores, where k is the beam size. When a candidate produces the second argument (e.g., "Minnesota"), the candidate is complete.

We show a running example as follows. Let's first consider the search from left to right with beam size equal to 1. An example search process is shown in Figure 2. Given an argument pair "Dylan" and "Minnesota", at each step, the search performs one of the following actions:

- `START` the search from first argument. The first argument is added as an initial candidate into the beam. In Figure 2(a), at step 0, "Dylan" is added into the beam. The total attention score is initialized to 0.

- `YIELD` a new partial candidate in the beam if the current candidate has not reached the second argument. This action conducts the following: The next largest attended token is appended to the end of the current candidate to yield the new candidate. The total score is increased by the associated attention score. At step 1 of Figure 2(a), "born" is appended to the current candidate to yield the partial candidate, since "born" has the largest attention score (0.2 as highlighted in Figure 2(b)) with "Dylan" in the attention matrix. The total score becomes 0.2. Note that we only consider the single head attention in this example for simplicity. "x" in Figure 2(b) marks the tokens (prior to the current token) that are not considered in the search to prevent searching backward. Step 2 takes the same action, and the score becomes 0.5.

- `STOP` the search step if the candidate has reached the second argument, then add the candidate as a valid triple into the beam. As beam size equals to 1, *(Dylan; born in; Minnesota)* is returned for the given pair. The final score of the triple is 0.7.

We also notice triples are often in reverse order in the sentence, thus enabling bidirectionality by running the algorithm in both directions (left to right and right to left). We merge the subwords as words, and only consider word-level attention. The beam search is implemented by the breadth-first search, which is efficient as the time complexity is $O(k \cdot d)$. d is the maximum depth of the search tree.

3 The IELM Benchmark

3.1 Datasets

3.1.1 Standard OIE

We adopt two standard OIE datasets below.

CaRB CaRB (Bhardwaj et al., 2019) is a crowd-sourced OIE dataset, where the input sentences are from the OIE2016 (Stanovsky and Dagan, 2016).

Re-OIE2016 Re-OIE2016 (Zhan and Zhao, 2020) is also generated based on the input sentences in the OIE2016, and is further enhanced by human annotations.

3.1.2 Factual OIE

In addition, we introduce two large-scale factual OIE datasets based on knowledge graphs (KG).

Method	AIDA	
	dev	test
Spitkovsky and Chang 2012	26.0	28.2
Kolitsas et al. 2018*	-	82.4
Ours	63.8	64.5

Table 1: Evaluation of unsupervised entity linking of Wikidata-OIE on AIDA benchmark. An asterisk (*) indicates a supervised method.

TAC KBP-OIE TAC Knowledge Base Population (KBP) Slot Filling is a task to search a document collection to fill in a target entity for predefined relations (slots) with a given entity in a reference KG. We adapt the dataset as an OIE dataset. In particular, we use a document sub-collection of the TAC KBP 2013 task (Surdeanu, 2013) as the input, and use the official human annotations regarding the documents as gold extractions.

Wikidata-OIE Besides TAC KBP-OIE, we create a larger factual OIE dataset based on the English Wikipedia. Different from TAC KBP, there are no gold triple annotations for Wikipedia. Since a large amount of Wikidata triples are from English Wikipedia, we create the dataset using distant supervision (Zhang et al., 2017) by aligning Wikidata triples to Wikipedia text. We employ an unsupervised entity linker based on a pre-built mention-to-entity dictionary (Spitkovsky and Chang, 2012) to extract potential gold arguments for scalability considerations. The entity linker links an arbitrary entity mention in a sentence to a Wikipedia anchor, which is further linked to a Wikidata entity. For each sentence in Wikipedia articles containing two linked arguments, if there is a Wikidata triple describing a relation holding the two arguments, we denote the Wikidata triple as a gold extraction.

Unlike TAC KBP-OIE which is built based on human annotations, Wikidata-OIE is derived from automatic annotations. Therefore, we evaluate our unsupervised entity linker on the standard AIDA benchmark (Hoffart et al., 2011) consisting of Wikipedia entities. Table 1 shows that it significantly improves the unsupervised performance (Spitkovsky and Chang, 2012) and reaches competitiveness with a supervised method (Kolitsas et al., 2018). Given the scale of Wikidata-OIE, we sacrifice acceptable effectiveness for efficiency.

The statistics of the datasets are shown in Table 2. For CaRB and Re-OIE2016, we report the statistics of the corresponding test sets. We include a dataset comparison in Appendix A.3.

Dataset	# of triples	# of arguments	# of predicates	# of documents
Re-OIE2016	1,508	3,328	1,506	595
CaRB	2,715	6,226	2,715	641
TAC KBP-OIE	27,655	39,661	41	3,877,207
Wikidata-OIE	27,368,562	6,047,494	1,156	6,047,494

Table 2: Dataset statistics of the IELM benchmark.

3.2 Pre-Trained Language Models for OIE

Unidirectional Language Models Given an input sequence $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$, unidirectional LMs assign a joint probability to the sequence by factorizing it as $p(\mathbf{x}) = \prod_t p(x_t|x_{t-1}, \ldots, x_1)$, where $p(x_t|x_{t-1}, \ldots, x_1) = \sigma(\mathbf{W}\mathbf{h}_t + \mathbf{b})$. $\mathbf{h}_t$ is the output vector of a neural network at position t.

We consider GPT-2 (Radford et al., 2019), where $\mathbf{h}_t$ is produced by Transformer decoders (Vaswani et al., 2017). GPT-2 is pre-trained on WebText containing 40GB of text. We explore all four pre-trained GPT-2s with different model sizes: GPT-2 (117M), GPT-2$_{\text{MEDIUM}}$ (345M), GPT-2$_{\text{LARGE}}$ (774M), and GPT-2$_{\text{XL}}$ (1558M).

Bidirectional Language Models Different from unidirectional LMs that predict the next word given the previous words, bidirectional LMs take both left and right context of the target word into consideration, formally, $p(x_t) = p(x_t|x_1, \ldots, x_{t-1}, x_{t+1}, \ldots, x_N)$.

We use BERT (Devlin et al., 2018) that enables bidirectional context modeling via a masked LM objective and utilizing the Transformer architecture. BERT is pre-trained on BooksCorpus and English Wikipedia. We use both its pre-trained settings: BERT$_{\text{BASE}}$ (109M) and BERT$_{\text{LARGE}}$ (335M).

3.3 Comparison Methods

We compare our method with a wide set of OIE systems including both neural and traditional linguistic OIE systems. Most OIE systems are based on supervised learning, which are indicated with asterisks (*) in Table 3. We provide details of the comparison systems in Appendix A.5.

3.4 Evaluation Method

3.4.1 Standard OIE

On CaRB and Re-OIE2016, we follow the original evaluation proposed in (Bhardwaj et al., 2019) and (Stanovsky and Dagan, 2016; Zhan and Zhao, 2020), and report precision, recall, F1, area under the curve (AUC) for compared OIE systems. AUC is calculated from a plot of the precision and recall values for all potential confidence thresholds. The F1 is the maximum value among the precision-recall pairs. We follow the matching function proposed for each dataset, i.e., lexical match for Re-OIE2016, and tuple match for CaRB. The CaRB evaluation function is stricter as it penalizes long extractions.

3.4.2 Factual OIE

We report precision, recall, and F1 of the OIE systems on two large-scale factual OIE datasets: TAC KBP-OIE and Wikidata-OIE. We introduce *exact match* as the matching function for them as below.

Matching Function The matching functions for standard OIE datasets are generally flexible. For example, the lexical match of Re-OIE2016 judges an argument or predicate as correct if and only if it includes the syntactic head of the gold argument or predicate. Unlike these matching functions, our exact matching function requires both arguments and predicates are linked to the gold extractions.

For TAC KBP-OIE, we judge an argument to be correct if and only if it matches the name of the gold argument and the span position of the gold argument in the sentence. The main challenge is how to properly link a predicate, since there are often many ways to express it. We follow Stanford OpenIE (Angeli et al., 2015) to produce the predicate mapping between the OIE relations and TAC KBP predicates. A predicate is correct if the pair of OIE relation and gold predicate exists in the predicate mapping. The predicate mapping is constructed in two steps. First, a collection of predicate mappings was constructed by a single annotator in approximately a day. Second, predicate mappings were finalized through the following learning procedure. This process matches OIE relations to the TAC KBP predicates by searching for co-occurrent relations in a large distantly-labeled corpus, and decides pairs of OIE relations and TAC KBP predicates that have a high PMI[2]. The basic idea is that the more often the argument pairs of the triples and TAC KBP triples are linked, the more likely the corresponding relations or predicates are linked to each other. Example predicate mappings are shown in Appendix A.4.

For Wikidata-OIE, we link an argument based on the entity linker used in Wikidata-OIE construction (Sec. 3.1). An argument is correct if the linked argument matches the name of the gold argument and the span position of the gold argument in the sentence. The predicate mapping is bootstrapped from TAC KBP-OIE's mapping. In addition, we normalize each predicate phrase of the triples by lemmatization, and removing inflection, auxiliary verbs, adjectives, adverbs. One author manually filters out the bad predicate pairs. This process takes approximately a day. A predicate is correct if the OIE to gold predicate pair exists in the bootstrapped predicate mapping. An annotator randomly subsampled and checked 100 aligned triple-sentence pairs and concluded with a 93% accuracy of extracted triples.

Metrics We use the official scorer of TAC KBP Slot Filling 2013 to calculate precision, recall, and F1 for TAC KBP-OIE. Besides, like previous OIE systems, for LMs, we adopt two constraints from ReVerb (Fader et al., 2011), the antecessor of OpenIE5: (i) the frequency of predicates must be above a threshold aiming to avoid triples to be over-specified, and (ii) a predicate must be a contiguous sequence in the sentence avoiding predicates that have no meaningful interpretation. We set these parameters according to Sec. 4.4.

We only report precision, recall, and F1 based on the parameter study in Sec. 4.4 for pre-trained LMs on the IELM benchmark. We do not compute AUC as pre-trained LMs are treated as zero-shot OIE systems. We therefore do not tune the results with respect to different confidence. Our main focus is to benchmark the OIE performance of the LMs under a unified setting. Another reason is that it is computationally expensive to get the AUC on the two large-scale datasets: TAC KBP-OIE and Wikidata-OIE. We also do not report AUC for the compared OIE systems on TAC KBP-OIE and Wikidata-OIE. Instead, we use the confidence threshold that obtains the best F1 on Re-OIE2016 to compute the scores.

4 Results

In this section, we show that pre-trained LMs are effective zero-shot OIE systems, and exceed the previous state-of-the-art OIE systems on our large-scale factual OIE datasets in IELM benchmark. To keep our evaluation as simple as possible, the hyper-parameters and settings are shared across datasets. More experimental details are described in the appendix.

Method	#Params	CaRB				Re-OIE2016				TAC KBP-OIE			Wikidata-OIE		
		P	R	F1	AUC	P	R	F1	AUC	P	R	F1	P	R	F1
MinIE (Gashteovski et al., 2017)	-	-	-	41.9	-	48.2	71.1	57.5	47.2	-	-	-	-	-	-
ClausIE (Del Corro and Gemulla, 2013)	-	-	-	44.9	22.4	-	-	64.2	46.4	-	-	-	-	-	-
OLLIE (Schmitz et al., 2012)	-	-	-	41.1	22.4	-	-	49.5	31.3	-	-	-	-	-	-
PropS (Stanovsky et al., 2016)	-	-	-	31.9	12.6	-	-	64.2	43.3	-	-	-	-	-	-
OpenIE4 (Christensen et al., 2011)	-	-	-	48.8	27.2	-	-	68.3	50.9	52.9	14.2	22.4	30.6	14.0	19.2
OpenIE5 (Saha et al., 2017, 2018)	-	-	-	48.0	25.0	61.1	76.5	67.9	45.8	57.0	14.6	23.2	26.3	14.4	18.6
Stanford OpenIE* (Angeli et al., 2015)	-	-	-	23.0	13.4	-	-	16.7	11.5	61.6	17.4	27.1	23.3	13.1	16.8
SenseOIE* (Roy et al., 2019)	640K	-	-	28.2	-	-	-	-	-	-	-	-	-	-	-
SpanOIE* (Zhan and Zhao, 2020)	963K	60.9	41.6	49.4	30.0	79.7	74.5	77.0	65.8	-	-	-	-	-	-
NeuralOIE* (Cui et al., 2018)	5M	-	-	51.6	32.8	79.2	77.5	78.4	73.0	-	-	-	-	-	-
Multi^{2}OIE* (Ro et al., 2020)	110M	60.9	45.8	52.3	32.6	86.9	81.0	83.9	74.6	-	-	-	-	-	-
RnnOIE* (Stanovsky et al., 2018)	965K	55.6	40.2	46.7	26.8	84.2	73.9	78.7	68.3	50.0	14.6	22.6	29.9	15.9	20.7
IMOJIE* (Kolluru et al., 2020b)	110M	64.7	45.6	53.5	33.3	88.1	67.1	76.2	63.1	58.2	14.9	23.8	31.2	16.2	21.3
OpenIE6* (Kolluru et al., 2020a)	220M	-	-	54.0	35.7	75.3	78.2	76.7	73.8	60.0	14.9	23.9	29.8	15.3	20.3
BERT$_{\text{BASE}}$ (zero-shot) (ours)	109M	21.2	18.3	19.7	-	25.9	34.0	29.4	-	61.6	18.8	28.8	32.0	18.9	23.7
BERT$_{\text{LARGE}}$ (zero-shot) (ours)	335M	22.4	20.2	21.2	-	30.7	38.5	34.1	-	61.7	19.0	29.1	32.3	19.1	24.0
GPT-2 (zero-shot) (ours)	117M	23.1	19.8	21.3	-	25.1	38.3	30.3	-	61.6	18.2	28.1	32.4	18.1	23.2
GPT-2$_{\text{MEDIUM}}$ (zero-shot) (ours)	345M	23.7	20.0	21.7	-	26.8	39.9	32.0	-	62.1	18.7	28.7	33.1	18.2	23.5
GPT-2$_{\text{LARGE}}$ (zero-shot) (ours)	774M	24.2	20.5	22.2	-	27.4	41.6	33.0	-	62.4	19.0	29.1	34.2	18.3	23.8
GPT-2$_{\text{XL}}$ (zero-shot) (ours)	1558M	24.5	21.0	22.7	-	29.3	43.4	35.0	-	62.7	19.5	29.7	35.7	18.5	24.4

Table 3: Compare the quality of different OIE systems. An asterisk (*) indicates a supervised method.

4.1 OIE Results

Table 3 shows the results. While zero-shot OIE systems synthesized by pre-trained LMs obtain notably lower scores compared to previous OIE systems on standard OIE datasets, they outperform the previous OIE systems on factual OIE datasets. We also find that larger LMs obtain improved results on all datasets. For example, BERT$_{\text{LARGE}}$ outperforms BERT$_{\text{BASE}}$ due to its larger model size. GPT-2s share similar trends. This is because larger LMs store richer relational information. This finding is consistent with previous studies (Petroni et al., 2019, 2020).

4.1.1 Standard OIE

The main reasons for the degraded performance of pre-trained LMs on standard OIE datasets are three-fold. First, the comparison methods mainly involve supervised systems that are trained on OIE datasets, which are denoted with asterisks (*) in Table 3. Besides the supervised systems, the remaining comparison systems all require human involvement, such as providing linguistic patterns for the extraction. In contrast, the pre-trained LMs are used as zero-shot OIE systems without using any training sets. Second, the zero-shot OIE modules still have room to improve. For example, approximately 30.0% of the argument extraction errors are due to the spaCy noun chunker. 16.9% of the gold extractions contain predicates outside the argument pairs. The current predicate extraction only allows searching between the arguments, and thus cannot handle such cases. Third, standard OIE benchmarks such as CaRB and Re-OIE2016 mainly examine the general information extraction capability. The zero-shot approach is not able to recall the information of interest in LMs. We might need a specific module (e.g., ranking) to locate such information. Interestingly, pre-trained LMs achieve comparable performance with supervised Stanford OpenIE. The result indicates pre-trained LMs contain informative patterns that are useful for OIE.

4.1.2 Factual OIE

As shown in Table 3, the best zero-shot OIE system based on GPT-2$_{\text{XL}}$ obtains a +2.6% and a +3.1% absolute F1 improvement on TAC KBP-OIE and Wikidata-OIE respectively over the previous supervised state-of-the-art. Due to the computation cost of OIE systems (Sec. 4.3), we only select several best performed OIE systems on the standard OIE datasets for the large-scale OIE experiments including: linguistic OIE systems (OpenIE4, OpenIE5, Stanford OpenIE) and neural OIE systems (RnnOIE, IMOJIE, OpenIE6).

Compared to the results on standard OIE datasets, pre-trained LMs consistently achieve state-of-the-art performance on both datasets. Both datasets emphasize measuring factual arguments and predicates in the reference KGs. Previous studies (Petroni et al., 2019, 2020) show that LMs have stored a considerable amount of factual information via pre-training on large-scale text. We draw the same conclusion. To the best of our knowledge, our IELM benchmark is the first benchmark that includes factual OIE datasets. More importantly, both

Method	MinIE	ClausIE	OLLIE	PropS	OpenIE 4	OpenIE 5	Stanford OpenIE		SpanOIE	RnnOIE
Sentences/sec.	8.9	4.0	14.5	4.6	20.1	3.1	2.5		19.4	149.2
Method	NeuralOIE	IMOJIE	Multi^{2}OIE	OpenIE 6	BERT$_{\text{BASE}}$	BERT$_{\text{LARGE}}$	GPT-2	GPT-2$_{\text{MEDIUM}}$	GPT-2$_{\text{LARGE}}$	GPT-2$_{\text{XL}}$
Sentences/sec.	11.5	2.6	21.2	142.0	16.2	11.9	13.9	12.7	11.5	11.0

Table 4: Runtime on Re-OIE2016.

linguistic and neural OIE systems are derived from manually designed linguistic patterns or learned patterns. The result shows that the pre-trained attention weights capture a more flexible set of factual patterns. The result also suggests that our approach is capable of using such patterns. In order to scale our approach to large-scale datasets, the argument and predicate extraction are both efficient by design. In particular, the beam search for predicate extraction is efficient in exploring the relational sequences in the input sentence. Besides, the attention scores used in the beam search are produced via a single forward pass of the pre-trained LM over the input sentence without fine-tuning.

Moreover, we find that BERT LMs outperform GPT-2 LMs under similar model sizes. On both datasets, BERT$_{\text{BASE}}$ performs better than GPT-2 in F1, and BERT$_{\text{LARGE}}$ outperforms GPT-2$_{\text{MEDIUM}}$ in F1. This is mainly because the recall of BERT LMs is higher than that of corresponding GPT-2 LMs. The result indicates that the Cloze-style loss function (i.e., masked LM) of BERT is more effective and flexible in recovering information than the autoregressive LM objective. We also notice that the precision of GPT-2 LMs is higher than that of BERT LMs. The reason is that the autoregressive LM objective captures more accurate information than Cloze-style loss does by preventing extra noise (e.g., masked tokens) in pre-training.

Pre-trained LMs achieve competitive precision, e.g., the precision is greater than 60% on TAC KBP-OIE. However, only moderate recalls are obtained. Therefore, improving recall is clearly the future direction. We find that both argument and predicate extraction can be further improved. For example, the main cause of the moderate recall is the incorrect arguments caused by spaCy noun chunks as summarized in Sec. 4.2. Besides, we can incorporate predicates that are not between the argument pairs into the extractions, as we observe a number of gold triples are in inverted sentences. We also notice that the F1 gain over previous state-of-the-arts on TAC KBP-OIE is smaller compared to that on Wikidata-OIE. A larger text corpus, e.g., Wikipedia, provides more information. We could improve the recall by running on larger corpora such as WebText2 and Common Crawl (Raffel et al., 2019; Brown et al., 2020) to collect more triples.

4.2 Error Analysis

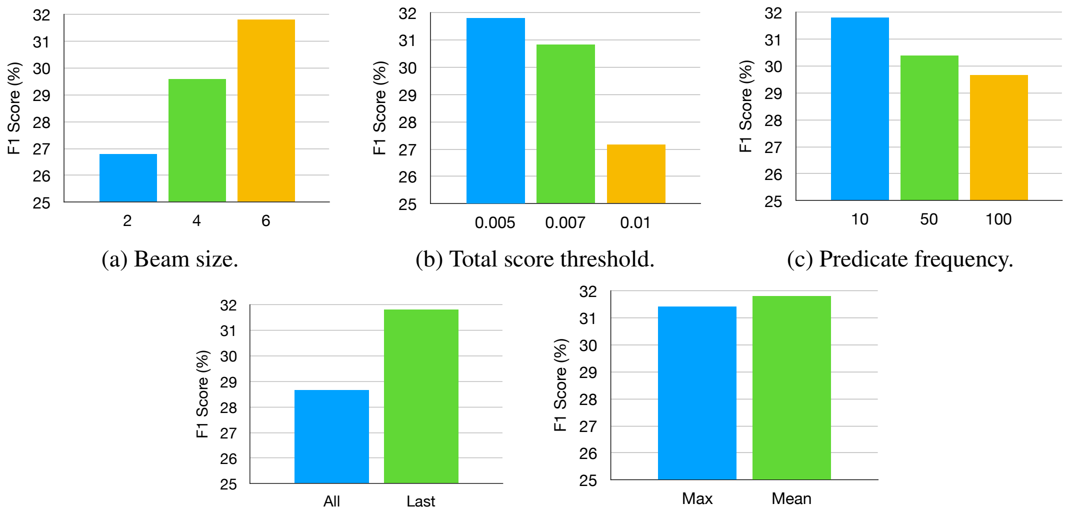
There is still significant room to improve the results. We argue that we are measuring a lower bound for what LMs know. To further understand the shortcomings of the current method, we conduct an error analysis of the errors in precision on all datasets. We choose BERT$_{\text{LARGE}}$ for the study. We sample 100 documents from the Wikidata-OIE dataset, and manually check the reasons for the errors. We find 33.1% of the errors are caused by *incorrect arguments*, while the predicate phrases are correct. The errors are due to the incorrect noun chunks detected by the spaCy. 18.3% of the errors are due to the *missing pairs in predicate mapping*. We also note that approximately 23.8% of the errors are actually *correct triples that are not covered by Wikidata*. For example, *(Bob_Dylan, residence, Nashville)* does not exist in Wikidata, but it is a correct triple. The rest of the errors made by BERT$_{\text{LARGE}}$ are *incorrect predicate phrases*, such as uninformative phrases. We find similar errors are made by BERT$_{\text{LARGE}}$ on other datasets. Based on the above analysis, enhancing argument detection and predicate mapping is helpful to further improve the results.

4.3 Runtime Analysis

The runtime of OIE systems is crucial in practice. We test the runtime of different OIE systems on Re-OIE2016. The results are in Table 4. We find ours is competitive in terms of efficiency given the size of the models.

4.4 Parameter Study

We study the effects of the key parameters using BERT$_{\text{BASE}}$ on TAC KBP-OIE as shown in Figure 3. We randomly sample 20% of the oracle query entities (provided by TAC KBP) as a hold-out



(a) Beam size. (b) Total score threshold. (c) Predicate frequency.

(d) Attention weights layers. (e) Attention weights reduction.

Figure 3: Parameter study with $\text{BERT}_{\text{BASE}}$ on TAC KBP-OIE hold-out subset.

dataset to tune the parameters, and use the best parameter setting achieved for all experiments. When studying the effect of a certain parameter, we keep the remaining parameters as default. We use F1 to measure the effects. Additional details are described in Appendix A.2.3.

5 Related Work

Pre-trained language models (LM), e.g., BERT (Devlin et al., 2018), GPT (Radford et al., 2018, 2019), and large LMs over 100B parameters (Brown et al., 2020; Chowdhery et al., 2022; Zeng et al., 2022) contain growing amount of linguistic and factual knowledge obtained via pre-training on large-scale corpora. To evaluate their abilities, researchers have created many knowledge benchmarks. LAMA leverages manually created prompts (Petroni et al., 2019, 2020). Recent studies have also developed soft prompts (Liu et al., 2021; Zhong et al., 2021)) for fact retrieval. KILT (Petroni et al., 2021) proposes a knowledge-intensive benchmark concerning several downstream tasks to evaluate LMs' ability in capturing knowledge. Wang et al. (2022) have utilized a set of knowledge-intensive structure prediction tasks to evaluate the knowledge in pre-trained LMs. Shen et al. (2022) have adapted KG completion as a benchmark to evaluate LMs. Besides relational knowledge, closed-book OpenQA (Roberts et al., 2020) benchmarks (in which LMs answer the open-domain questions without retrieving contexts) have also been adopted as a way to evaluate LMs' knowledge. While the existing benchmarks evaluate LMs in an implicit way, the main difference is that our benchmark explicitly and interpretably evaluates triples from the textual corpora extracted using model parameters (e.g. attentions). In the field of neural network interpretation (Linzen et al., 2016; Adi et al., 2016; Tenney et al., 2019), in particular the pre-trained deep LM analysis, substantial recent work focuses on both visualizing and analyzing the attention (Vig, 2019; Jain and Wallace, 2019; Clark et al., 2019; Michel et al., 2019; Ramsauer et al., 2020). Instead of analyzing or visualizing, our benchmark quantitatively evaluates the relational information with respect to open information extraction.

Open information extraction systems, e.g., OLLIE (Schmitz et al., 2012), Reverb (Fader et al., 2011), Stanford OpenIE (Angeli et al., 2015), OpenIE 5 (Saha et al., 2017, 2018), RnnOIE (Stanovsky et al., 2018), and OpenIE 6 (Kolluru et al., 2020a) aim to extract triples from web corpora for open schema KGs. Besides, NELL (Carlson et al., 2010), DeepDive (Niu et al., 2012), Knowledge Vault (Dong et al., 2014) extract information based on a fixed schema or ontology, where humans help improve the accuracy of the extractions. Probase (Wu et al., 2012) produces taxonomies instead of rich typed relations in general KGs. Our benchmark first evaluates LMs' unsupervised information extraction ability on common open information extraction datasets such as CaRB (Bhardwaj et al., 2019) and Re-OIE2016 (Zhan and Zhao, 2020), and then aligns the extracted triples to KG triples for large-scale knowledge extraction benchmark construction. Our algorithm is similar to the generation algorithm in DeepEx (Wang et al., 2021). The focus of this work is to benchmark the zero-shot OIE performance of pre-trained LMs on both standard and factual OIE datasets. To further improve the OIE performance, the ranking module in DeepEx can be useful. The structure pre-training proposed in (Wang et al., 2022) can also be helpful.

6 Conclusion

We benchmark the general relational information in pre-trained language models (LM) in an open information extraction (OIE) setup. We find that the pre-trained LMs contain a considerable amount of open relational information through large-scale evaluation on both standard OIE datasets and newly created large-scale factual OIE datasets in our IELM benchmark. We are able to turn pre-training LMs into zero-shot OIE systems to efficiently deliver the benchmark results. The reach of this result is broad and has potential downstream utility for deep neural network interpretation, information extraction, and knowledge graph construction. Although the results are promising, we argue that our results just indicate a lower bound about what the LMs have. We hope our results will foster further research in the LM OIE benchmark direction.

7 Limitations

For the limitations of our method, the argument extraction module of our algorithm relies on a third-party noun chunker. As reported, the noun chunker introduces the majority of the errors in our extraction results. A limitation in our benchmark is that we have not conducted a large-scale manual evaluation of our factual OIE datasets (TAC KBP-OIE and Wikidata-OIE). The main focus of our study is to provide a large-scale OIE benchmark. As a result, this makes our benchmark more challenging to be used than standard OIE datasets in terms of computation costs and infrastructure. Finally, we have only benchmarked BERT and GPT-2 on our datasets. Future work could include testing a wide range of language models on our benchmark.

8 Ethical Considerations

We hereby acknowledge that all of the co-authors of this work are aware of the provided *ACM Code of Ethics* and honor the code of conduct. This work is about benchmarking the zero-shot OIE capability of pre-trained language models including BERT and GPT. Our ethical considerations and the work's underlying future impacts are discussed in the following perspectives. Language models are known to present potential risks and limitations (Brown et al., 2020), and the corpus used in pre-training (such as Wikipedia) may introduce unwanted biases and toxicity. We do not anticipate the production of harmful outputs after using our method or datasets, especially for vulnerable populations.

9 Environmental Impact

We adopt the pre-trained language models BERT (Devlin et al., 2018) and GPT-2 series (Radford et al., 2019) in our IELM benchmark. The models' carbon footprints are estimated to be 22–28 kilograms (Gibney, 2022). Additionally, The focus of this study is to test the zero-shot OIE ability of pre-trained language models. We do not train language models on massive datasets. Instead, we only do inference on a few evaluation datasets. This cost is less than 0.1% energy than that of their pre-training. This demonstrates that developing proper zero-shot learning strategies for large language models can not only deepen our understanding of their latent mechanisms, but also further reduce the energy consumption and environmental impacts that language models with ever-growing size may cause.

Acknowledgements

We would like to thank the anonymous reviewers for their suggestions and comments. This material is in part based upon work supported by Berkeley DeepDrive and Berkeley Artificial Intelligence Research.

References

Yossi Adi, Einat Kermany, Yonatan Belinkov, Ofer Lavi, and Yoav Goldberg. 2016. Fine-grained analysis of sentence embeddings using auxiliary prediction tasks. *arXiv preprint arXiv:1608.04207*.

Gabor Angeli, Melvin Jose Johnson Premkumar, and Christopher D Manning. 2015. Leveraging linguistic structure for open domain information extraction. In *ACL*, pages 344–354.

Sangnie Bhardwaj, Samarth Aggarwal, and Mausam Mausam. 2019. Carb: A crowdsourced benchmark for open ie. In *EMNLP*, pages 6262–6267.

Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda

Askell, et al. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R Hruschka Jr, and Tom M Mitchell. 2010. Toward an architecture for never-ending language learning. In *AAAI*, 3.

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. 2022. Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311*.

Janara Christensen, Stephen Soderland, and Oren Etzioni. 2011. An analysis of open information extraction based on semantic role labeling. In *KCAP*, pages 113–120.

Kevin Clark, Urvashi Khandelwal, Omer Levy, and Christopher D Manning. 2019. What does bert look at? an analysis of bert's attention. *arXiv preprint arXiv:1906.04341*.

Lei Cui, Furu Wei, and Ming Zhou. 2018. Neural open information extraction. In *ACL*, pages 407–413.

Luciano Del Corro and Rainer Gemulla. 2013. Clausie: clause-based open information extraction. In *WWW*, pages 355–366.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *KDD*, pages 601–610.

Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying relations for open information extraction. In *EMNLP*, pages 1535–1545.

Kiril Gashteovski, Rainer Gemulla, and Luciano del Corro. 2017. Minie: minimizing facts in open information extraction. ACL.

Elizabeth Gibney. 2022. How to shrink ai's ballooning carbon footprint. *Nature*, 607(7920):648–648.

John Hewitt and Christopher D Manning. 2019. A structural probe for finding syntax in word representations. In *NAACL*, pages 4129–4138.

Johannes Hoffart, Mohamed Amir Yosef, Ilaria Bordino, Hagen Fürstenau, Manfred Pinkal, Marc Spaniol, Bilyana Taneva, Stefan Thater, and Gerhard Weikum. 2011. Robust disambiguation of named entities in text. In *EMNLP*, pages 782–792.

Sarthak Jain and Byron C Wallace. 2019. Attention is not explanation. *arXiv preprint arXiv:1902.10186*.

Zhengbao Jiang, Frank F Xu, Jun Araki, and Graham Neubig. 2020. How can we know what language models know? *TACL*, 8:423–438.

Nikolaos Kolitsas, Octavian-Eugen Ganea, and Thomas Hofmann. 2018. End-to-end neural entity linking. *arXiv preprint arXiv:1808.07699*.

Keshav Kolluru, Vaibhav Adlakha, Samarth Aggarwal, Soumen Chakrabarti, et al. 2020a. Openie6: Iterative grid labeling and coordination analysis for open information extraction. In *EMNLP*, pages 3748–3761.

Keshav Kolluru, Samarth Aggarwal, Vipul Rathore, Soumen Chakrabarti, et al. 2020b. Imojie: Iterative memory-based joint open information extraction. In *ACL*, pages 5871–5886.

Tal Linzen, Emmanuel Dupoux, and Yoav Goldberg. 2016. Assessing the ability of lstms to learn syntax-sensitive dependencies. *TACL*, pages 521–535.

Xiao Liu, Yanan Zheng, Zhengxiao Du, Ming Ding, Yujie Qian, Zhilin Yang, and Jie Tang. 2021. Gpt understands, too. *arXiv:2103.10385*.

Paul Michel, Omer Levy, and Graham Neubig. 2019. Are sixteen heads really better than one? In *NIPS*, pages 14014–14024.

Feng Niu, Ce Zhang, Christopher Ré, and Jude Shavlik. 2012. Elementary: Large-scale knowledge-base construction via machine learning and statistical inference. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 8(3):42–73.

Fabio Petroni, Patrick Lewis, Aleksandra Piktus, Tim Rocktäschel, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. 2020. How context affects language models' factual predictions. *arXiv preprint arXiv:2005.04611*.

Fabio Petroni, Aleksandra Piktus, Angela Fan, Patrick Lewis, Majid Yazdani, Nicola De Cao, James Thorne, Yacine Jernite, Vladimir Karpukhin, Jean Maillard, et al. 2021. Kilt: a benchmark for knowledge intensive language tasks. In *NAACL*, pages 2523–2544.

Fabio Petroni, Tim Rocktäschel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, Alexander H Miller, and Sebastian Riedel. 2019. Language models as knowledge bases? *arXiv preprint arXiv:1909.01066*.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI Blog*, (8):9.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. *CoRR*, abs/1910.10683.

Hubert Ramsauer, Bernhard Schäfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Lukas Gruber, Markus Holzleitner, Milena Pavlović, Geir Kjetil Sandve, Victor Greiff, et al. 2020. Hopfield networks is all you need. *arXiv preprint arXiv:2008.02217*.

Youngbin Ro, Yukyung Lee, and Pilsung Kang. 2020. Multi^2oie: Multilingual open information extraction based on multi-head attention with bert. In *Findings of EMNLP*, pages 1107–1117.

Adam Roberts, Colin Raffel, and Noam Shazeer. 2020. How much knowledge can you pack into the parameters of a language model? *arXiv preprint arXiv:2002.08910*.

Arpita Roy, Youngja Park, Taesung Lee, and Shimei Pan. 2019. Supervising unsupervised open information extraction models. In *EMNLP*, pages 728–737.

Swarnadeep Saha, Harinder Pal, et al. 2017. Bootstrapping for numerical open ie. In *ACL*, pages 317–323.

Swarnadeep Saha et al. 2018. Open information extraction from conjunctive sentences. In *COLING*, pages 2288–2299.

Michael Schmitz, Stephen Soderland, Robert Bart, Oren Etzioni, et al. 2012. Open language learning for information extraction. In *EMNLP*, pages 523–534.

Jianhao Shen, Chenguang Wang, Linyuan Gong, and Dawn Song. 2022. Joint language semantic and structure embedding for knowledge graph completion. In *COLING*.

Valentin I Spitkovsky and Angel X Chang. 2012. A cross-lingual dictionary for english wikipedia concepts.

Gabriel Stanovsky and Ido Dagan. 2016. Creating a large benchmark for open information extraction. In *EMNLP*, pages 2300–2305.

Gabriel Stanovsky, Jessica Ficler, Ido Dagan, and Yoav Goldberg. 2016. Getting more out of syntax with props. *arXiv preprint arXiv:1603.01648*.

Gabriel Stanovsky, Julian Michael, Luke Zettlemoyer, and Ido Dagan. 2018. Supervised open information extraction. In *NAACL*, pages 885–895.

Mihai Surdeanu. 2013. Overview of the tac2013 knowledge base population evaluation: English slot filling and temporal slot filling. *TAC*, page 2.

Alon Talmor, Yanai Elazar, Yoav Goldberg, and Jonathan Berant. 2019. olmpics–on what language model pre-training captures. *arXiv preprint arXiv:1912.13283*.

Ian Tenney, Patrick Xia, Berlin Chen, Alex Wang, Adam Poliak, R Thomas McCoy, Najoung Kim, Benjamin Van Durme, Samuel R Bowman, Dipanjan Das, et al. 2019. What do you learn from context? probing for sentence structure in contextualized word representations. *arXiv preprint arXiv:1905.06316*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *NIPS*, pages 5998–6008.

Jesse Vig. 2019. Visualizing attention in transformerbased language models. *arXiv preprint arXiv:1904.02679*.

Chenguang Wang, Xiao Liu, Zui Chen, Haoyun Hong, Jie Tang, and Dawn Song. 2021. Zero-shot information extraction as a unified text-to-triple translation. In *EMNLP*.

Chenguang Wang, Xiao Liu, Zui Chen, Haoyun Hong, Jie Tang, and Dawn Song. 2022. DeepStruct: Pretraining of language models for structure prediction. In *ACL*.

Wentao Wu, Hongsong Li, Haixun Wang, and Kenny Q Zhu. 2012. Probase: A probabilistic taxonomy for text understanding. In *SIGMOD*, pages 481–492.

Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang, Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu, Wendi Zheng, Xiao Xia, et al. 2022. Glm-130b: An open bilingual pre-trained model. *arXiv preprint arXiv:2210.02414*.

Junlang Zhan and Hai Zhao. 2020. Span model for open information extraction on accurate corpus. In *AAAI*, pages 9523–9530.

Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D Manning. 2017. Position-aware attention and supervised data improve slot filling. In *EMNLP*.

Zexuan Zhong, Dan Friedman, and Danqi Chen. 2021. Factual probing is [mask]: Learning vs. learning to recall. In *NAACL*, pages 5017–5033.

A The IELM Benchmark Details

Additional details of our open information extraction (OIE) benchmark IELM are described in this section.

A.1 Wikidata-OIE

In this section, we describe some technical details regarding the construction and evaluation of Wikidata-OIE.

A.1.1 Entity Linking

We use an unsupervised entity linker for both Wikidata-OIE dataset construction and OIE evaluation. The entity linker is originally developed in (Spitkovsky and Chang, 2012), which is based on a mention-to-entity dictionary. We build an enhanced dictionary as follows: we add new Wikipedia anchors to the dictionary which results in 26 million entries compared to the original 21 million entries. Then a Wikipedia anchor to the Wikidata item dictionary is used to further link the entities (or arguments) to Wikidata. If an argument is a pronoun, we further use neuralcoref [2] for coreference resolution.

A.1.2 Predicate Mapping

The predicate mapping of Wikidata-OIE is constructed offline using the method in Sec. 3.4. In more detail, we randomly sampled a hold-out dataset including 2,000 documents from English Wikipedia for the bootstrapped predicate mapping construction based on the TAC KBP mapping (Angeli et al., 2015). To filter out the wrong predicate pairs, we manually check whether the top predicate phrases are true.

A.1.3 Gold Triples

For gold triples in Wikidata-OIE, we only preserve those triples describing predicates between arguments that can be linked to corresponding Wikipedia anchors. We rule out triples of attributes about arguments and triples of auxiliary predicates (such as *topic's main category.P901*) and finally result in 27,368,562 gold triple extractions.

A.1.4 Evaluation

Given the large number of source sentences and gold triples in Wikidata-OIE, a MongoDB database is maintained to store the gold triples to enable an efficient evaluation.

A.2 Zero-Shot Language Model Based Open Information Extraction

In this section, we introduce additional details about how we adapt pre-trained language models (LM) as zero-shot OIE systems.

A.2.1 Argument Extraction

We use spaCy noun chunker [3] to annotate the noun phrases in the sentences.

[2] https://github.com/huggingface/neuralcoref

[3] https://spacy.io/usage/linguistic-features/#noun-chunks

Algorithm 1 Beam search with attention scores.

Input: Argument pair (arg_0, arg_1), sentence s, attention matrix $\mathbf{A_s}$, action manager $\mathcal{O} = \{\texttt{START}, \texttt{YIELD}, \texttt{STOP}\}$, beam size k

Output: Triples $T_{(arg_0,arg_1)}$

```
T_(arg0,arg1) ← {START(arg0)}        ▷ Start by adding the first argument as a candidate in the beam
while ∃c ∈ T_(arg0,arg1)[O(c) = YIELD] do
    T~_(arg0,arg1) ← ∅                ▷ Initialize a new beam
    for each c ∈ T_(arg0,arg1) do
        if O(c) = YIELD then
            T~_(arg0,arg1) ← T~_(arg0,arg1) ∪ {YIELD(c, s, A_s)}   ▷ Yield a new candidate if not reached the second argument
        else
            T~_(arg0,arg1) ← T~_(arg0,arg1) ∪ {STOP(c, t)}   ▷ Stop then produce a valid triple if reached the second argument
        end if
    end for
    T_(arg0,arg1) ← TOP(k, T~_(arg0,arg1))   ▷ Maintain k-best candidates in the beam
end while
return T_(arg0,arg1)
```

A.2.2 Predicate Extraction

We first describe predicate extraction introduced in Sec. 2.2 in detail.

- **Beam Search**. The inputs of the search algorithm are an argument pair (arg_0, arg_1), a sentence s, an attention matrix $\mathbf{A}_s$ of s. Both arg_0 and arg_1 are identified by the noun chunker in s. $\mathbf{A}_s$ is the attention matrix associated with s from the forward pass of an LM without fine-tuning. The search gets started by adding the first argument arg_0 as the initial candidate in the beam. While there are still new candidates waiting to be yielded, the search continues, and the top k candidates sorted by the attention scores are maintained in the beam. The details of the proposed beam search are described in Algorithm 1. In practice, we implement an action manager $\mathcal{O}$ to decide which action to take at each step. Given a candidate c in the beam, $\mathcal{O}(c) = \texttt{START}$ always happens at the beginning of the search. If c has not reached the second argument arg_1 yet, $\mathcal{O}(c) = \texttt{YIELD}$. Otherwise, $\mathcal{O}(c) = \texttt{STOP}$.

- **Implementation Details**. For Wikidata-OIE, we randomly split the English Wikipedia data into 20 partitions, and map the data partitions to 20 distributed servers to run. Each server is configured with four Tesla K80 12Gs. We set the max sequence length to 256, and batch size as 32 for $\text{BERT}_{\text{LARGE}}$ and 4 for GPT-2_{XL}. We use implementations of pre-trained LMs in the

Transformers package [4]. We use spaCy sentencizer [5] to segment the documents into sentences. $\text{BERT}_{\text{LARGE}}$ takes approximately 48 hours, and GPT-2_{XL} costs around 96 hours. The resulting triples from the 20 servers are then reduced to a data server. The batch sizes of $\text{BERT}_{\text{BASE}}$, GPT-2, $\text{GPT-2}_{\text{MEDIUM}}$, $\text{GPT-2}_{\text{LARGE}}$ are 64, 32, 16, 8 respectively.

A.2.3 Parameter Settings

We then discuss the parameter setup of our OIE systems as below.

The parameter settings are shared across all OIE datasets. All the choices are based on the parameter study in Sec. 4.4. The beam size of Algorithm 1 is set to 6. The attention score threshold is set to 0.005, and the number of relation/predicate frequencies is set to 10. To generate the attention weight matrix $\mathbf{A}_s$ of a sentence, we reduce the weights of every attention head in the last layer of pre-trained LMs using the mean operator. We analyze the effects of various parameters below.

Figure 3(a) illustrates the effects of various beam sizes in Algorithm 1. We find that in general, the larger the beam size is, the better F1 the setting achieves. This is because our method is able to reserve more potentially correct triples when more candidates are allowed. However, F1 improvement gradually becomes subtle, while the computation costs increase more significantly. For efficiency consideration, we do not explore larger beam sizes. We set the beam size as 6.

Figure 3(b) compares the effect of different thresholds of the total score. We set the threshold as 0.005 since it achieves the best result. Note that the summed attention score is normalized by the length of the triple to penalize the cumbersome triples. The threshold is effective. This is mainly because of the relational information contained in the self-attention matrix: the score in the attention matrix is representing the chance of the triples to be the true triples based on the stored information. Figure 3(c) shows the impact of the predicate frequency threshold in identifying common predicates. The best result is achieved when it equals 10. This shows that while our method mostly identifies frequent predicates, it is also able to capture some rare predicates.

Figure 3(d) shows the comparison between the attention weights of the last layer and the mean of all layers. The attention weights of the last layer perform better. This is due to the attention weights in lower layers being low-level linguistic knowledge according to (Clark et al., 2019; Ramsauer et al., 2020), which are less relevant to the relational information. Figure 3(e) compares the impact of different attention reductions, i.e., mean, max, over the attention heads of the last layer. We find the "mean" performs better. The reason is that the token often intensively attends to several specific tokens in the sequence (Michel et al., 2019), and the "mean" operator is less sensitive to such biases.

A.3 The Number of Predicates of Standard and Factual OIE Datasets

Note that there are more predicates in standard OIE datasets than that in factual datasets. This is because, for standard OIE, predicates are open and not attached to a certain schema. These predicates were extracted from the input sentences and are usually natural language utterances. For factual OIE, the predicates are unified into a fixed KG schema. For example, for a person's birthplace, there are multiple natural language expressions like "was born in" or "gave birth" in standard OIE datasets, while only a single "birth_place" predicate exists in the factual OIE sets.

A.4 Predicate Mapping Examples

We show example predicate mappings in a dictionary below.

- `per:city_of_birth`: born in, born at, born, birth city, hometown.
- `org:founded_by`: established by, founded by, founded, founder, co-founder of.

where the keys are KG predicates, e.g., `per:city_of_birth` and `org:founded_by`. The values are the corresponding OIE relations.

A.5 Comparison Systems

We compare our zero-shot OIE systems with the following OIE systems.

A.5.1 Neural OIE Systems

The following neural network based systems are selected.

- SenseOIE (Roy et al., 2019)[6] learns to ensemble various previous unsupervised OIE systems' ex-

[4] https://github.com/huggingface/transformers
[5] https://spacy.io/api/sentencizer

[6] The code is not available.

tractions using supervised learning to combine their strengths.

- SpanOIE (Zhan and Zhao, 2020)[7] presents the Re-OIE2016 datasets for a more rigorous evaluation and a span-based (instead of sequence labeling) extraction model.

- RnnOIE (Stanovsky et al., 2018)[8] is one of the state-of-the-art OIE systems. It uses LSTM to model the OIE problem as a sequence tagging problem, and is trained on a large-scale OIE training set.

- NeuralOIE (Cui et al., 2018)[9] is an encoder-decoder based architecture that adopts the copy mechanism to conduct OIE.

- IMOJIE (Kolluru et al., 2020b)[10] is a sequence generation based OIE model that uses BERT at encoding time.

- Multi^{2}OIE (Ro et al., 2020)[11] models OIE as a sequence labeling problem that combines BERT with multi-head attention blocks.

- OpenIE6 (Kolluru et al., 2020a)[12] is one of the state-of-the-art OIE systems. It treats OIE as a 2-D grid labeling task, and trains a BERT family architecture for the task.

Note that while our methods are zero-shot without needing to use the specific training sets, all the neural OIE systems are supervised on corresponding training sets.

A.5.2 Linguistic OIE Systems

We also compare our systems with the following linguistic pattern based systems developed prior to the use of neural networks.

- MinIE (Gashteovski et al., 2017)[13] proposes to minimize facts in OIE by representing information by annotations rather than extraction and removing redundant specific information.

- ClausIE (Del Corro and Gemulla, 2013)[14] is a clause-based approach by first identifying linguistic structure and then their information and attributes.

- OLLIE (Schmitz et al., 2012)[15] uses contextual sentence decomposition to conduct OIE.

- PropS (Stanovsky et al., 2016)[16] proposes proposition structure which is implied from syntax using dependency trees.

- OpenIE4 (Christensen et al., 2011)[17] is the successor to OLLIE using similar argument and relation expansion heuristics to create OIE extractions from semantic role labeling frames.

- OpenIE5 (Saha et al., 2017, 2018)[18] is one of the state-of-the-art OIE systems, which is the successor to OLLIE, and it improves extractions from noun relations, numerical sentences, and conjunctive sentences depending on the linguistic patterns.

- Stanford OpenIE (Angeli et al., 2015)[19] leverages POS tag and dependency parser, and generates self-contained clauses from long sentences to extract the triples.

B The TAC KBP-OIE and Wikidata-OIE Datasets

We show samples of our zero-shot OIE extractions and the gold triples on both TAC KBP-OIE and Wikidata-OIE datasets.

B.1 TAC KBP-OIE

OIE Extractions and Gold Extractions We randomly sample 100 documents from the TAC KBP-OIE corpus, then sample sentences from those documents. The uncurated triples and the corresponding gold triples of the sampled sentences based on our best methods $\text{BERT}_{\text{LARGE}}$ and GPT-2_{XL} are shown in Figure 4 and Figure 5 respectively. We also randomly sample sentences in which $\text{BERT}_{\text{LARGE}}$ differs from GPT-2_{XL} in the resulting triples for comparison, which are illustrated in Figure 6. In each table, "**ID**" represents the document ID of a sampled sentence in the TAC KBP-OIE corpus. "**Sentence**" indicates the sampled sentence. "**Triples to gold triples**" column contains the extraction triples (on the left side of

[7] https://github.com/zhanjunlang/Span_OIE
[8] https://github.com/gabrielStanovsky/supervised-oie
[9] We use the BERT implementation available at https://github.com/dair-iitd/imojie.
[10] https://github.com/dair-iitd/imojie
[11] https://github.com/youngbin-ro/Multi2OIE
[12] https://github.com/dair-iitd/openie6
[13] https://github.com/uma-pi1/minie
[14] https://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/ambiverse-nlu/clausie
[15] https://github.com/knowitall/ollie
[16] https://github.com/gabrielStanovsky/props
[17] https://github.com/allenai/openie-standalone
[18] https://github.com/dair-iitd/OpenIE-standalone
[19] https://nlp.stanford.edu/software/openie.html

"→") and their corresponding gold triples (on the right side of "→").

B.2 Wikidata-OIE

OIE Extractions and Gold Extractions Similar to TAC KBP-OIE, we randomly sample 100 documents from the Wikidata-OIE corpus (i.e., English Wikipedia), then sample sentences from those documents. Similar to TAC KBP-OIE, Figure 7 and Figure 8 show the uncurated triples and the corresponding gold triples of the sampled sentences based on our zero-shot systems $\text{BERT}_{\text{LARGE}}$ and GPT-2_{XL} respectively. Figure 9 illustrates the randomly sampled sentences in which $\text{BERT}_{\text{LARGE}}$ extracts different triples compared to that from GPT-2_{XL}. In each table, "**ID**" represents the Wikipedia page's title of the sampled sentence. "**Sentence**" indicates the sampled sentence. "**Triples to gold triples**" column contains the triples (on the left side of "→") and their corresponding gold triples (on the right side of "→").

ID	Sentence	Triples to gold triples
SF13_ENG_001	Bashardost left Pakistan for France in 1981 .	(Bashardost, left, Pakistan) → (Ramazan Bashardost, per:countries_of_residence, Pakistan)
SF13_ENG_004	" Douglas Flint will succeed Stephen Green as Group Chairman and Stuart Gulliver will be appointed Group Chief Executive , following Michael Geoghegan ' s decision to retire early next year , " HSBC said in a statement .	(Douglas Flint, as, Group Chairman) → (Douglas Flint, per:title, Chairman)
SF13_ENG_007	Mohammed Tantawi , head of Al - Azhar University , told a schoolgirl to remove her niqab when he spotted her during a tour of an Al - Azhar affiliated school , the independent Al - Masry al - Youm newspaper reported this week .	(Mohammed Tantawi, head of, Al-Azhar University) → (Mohammed Sayed Tantawi, per:employee_or_member_of, Al-Azhar University)
SF13_ENG_009	He took office in 2006 by defeating longtime incumbent Anwar Chowdhry from Pakistan , who was later barred for alleged financial corruption .	(Longtime Incumbent Anwar Chowdhry, from, Pakistan) → (Anwar Chowdhry, per:origin, Pakistan)
SF13_ENG_010	In addition to his wife , Wendy , Dio is survived by son Daniel , grandchildren Julie and Joey , and father Pat .	(Dio, is survived by son, Daniel) → (Ronnie James Dio, per:children, Daniel)
SF13_ENG_011	Marshall is charged with grand larceny and fraud and faces up to 25 years in prison if convicted .	(Marshall, is charged, Fraud) → (Anthony Marshall, per:charges, fraud)
SF13_ENG_011	Marshall , a Tony Award - winning Broadway producer and former U . S . diplomat , sat stonelike as the jury forewoman read each verdict aloud , the word " guilty " clearly resonating in the otherwise silent courtroom .	(Marshall, ,, A Tony Award-Winning Broadway Producer) → (Anthony Marshall, per:title, producer)
SF13_ENG_011	The blueblood scion of one of America ' s most illustrious families appeared to listen impassively as verdicts finding him guilty on 14 counts of grand larceny , conspiracy and fraud were read to a packed courtroom .	(Him, guilty, Grand Larceny) → (Anthony Marshall, per:charges, grand larceny)
SF13_ENG_011	But Marshall ' s son , Philip , told a different story .	(Marshall'S Son, ,, Philip) → (Anthony Marshall, per:children, Philip)
SF13_ENG_012	Al - Qaeda ' s American spokesman Adam Gadahn , also known as Azzam the American , called on Muslims in the West on Sunday to carry out more attacks like the deadly shooting at the US base in Fort Hood , Texas .	(Adam Gadahn, also known, Azzam) → (Adam Gadahn, per:alternate_names, Azzam)
SF13_ENG_012	Gadahn , also known as Azzam the American , was born in 1978 .	(Gadahn, , also known as, Azzam) → (Adam Gadahn, per:alternate_names, Azzam)
SF13_ENG_012	Gadahn grew up in California and converted to Islam before he moved to Pakistan in 1998 and attended an al - Qaida training camp six years later , according to media reports .	(Gadahn, in, California) → (Adam Gadahn, per:statesorprovinces_of_residence, California)
SF13_ENG_012	Gadahn moved to Pakistan in 1998 , according to the FBI , and is said to have attended an al - Qaida training camp six years later , serving as a translator and consultant for the group .	(Gadahn, moved to, Pakistan) → (Adam Gadahn, per:countries_of_residence, Pakistan)
SF13_ENG_014	After the Munich attack , he lived in Lebanon , Jordan and several Eastern European countries , where he had close ties to Communist bloc intelligence agencies .	(He, lived in, Lebanon) → (Mohammed Oudeh, per:countries_of_residence, Lebanon)
SF13_ENG_015	Clifton attended Howard University but left before graduating to pursue poetry .	(Clifton, attended, Howard University) → (Lucille Clifton, per:schools_attended, Howard University)
SF13_ENG_017	" Alexander Haig devoted his career to serving our country , both as a soldier and as a diplomat , " Albright said . "	(Alexander Haig, devoted his, soldier) → (Alexander Haig, per:title, soldier)
SF13_ENG_017	In 1979 , he resigned and retired from the Army .	(He, and retired from, The Army) → (Alexander Haig, per:employee_or_member_of, Army)
SF13_ENG_019	McGregor is survived by his wife , Lori , and four children , daughters Jordan , Taylor and Landri , and a son , Logan .	(Mcgregor, his wife, Lori) → (Keli McGregor, per:spouse, Lori)
SF13_ENG_020	" Mike was a first - rate journalist , a valued member of our staff for 25 years and we will miss him , " Times Editor Russ Stanton said . "	(Mike, was, A First-Rate Journalist) → (Mike Penner, per:title, first-rate journalist)
SF13_ENG_020	Penner is survived by his brother , John , a copy editor at the Times , and his former wife , Times sportswriter Lisa Dillman .	(Penner, his brother ,, John) → (Mike Penner, per:siblings, John)
SF13_ENG_024	She was charged with theft in Beaumont , Texas , for allegedly failing to pay for $ 10 , 000 worth of dental work in 2006 .	(She, was charged, Theft) → (Crystal Taylor, per:charges, theft)
SF13_ENG_025	Michigan native Nancy Kissel was convicted of murder and sentenced in Hong Kong ' s High Court in September 2005 .	(Michigan Native Nancy Kissel, was convicted, Murder) → (Nancy Kissel, per:charges, murder)
SF13_ENG_026	Neal returned to New York and concentrated on stage work .	(Neal, returned to, New York) → (Patricia Neal, per:statesorprovinces_of_residence, New York)
SF13_ENG_026	In 1953 , she married Roald Dahl , the British writer famed for " Charlie and the Chocolate Factory , " " James and the Giant Peach " and other tales for children .	(She, married, Roald Dahl) → (Patricia Neal, per:spouse, Roald Dahl)
SF13_ENG_026	Oscar - winning actress Patricia Neal has died of lung cancer at her home on Martha ' s Vineyard , Massachusetts , on Sunday .	(Patricia, died of, Lung Cancer) → (Patricia Neal, per:cause_of_death, lung cancer)
SF13_ENG_027	Al - Hakim ' s son , Ammar al - Hakim , has been groomed for months to take his father ' s place .	(Al-Hakim'S Son, ,, Ammar Al-Hakim) → (Abdul Aziz Al-Hakim, per:children, Ammar al-Hakim)
SF13_ENG_027	Al - Hakim is the head of Supreme Iraqi Islamic Council (SIIC) , the largest Shiite party in Iraq .	(Al-Hakim, is, Supreme Iraqi Islamic Council) → (Abdul Aziz Al-Hakim, per:employee_or_member_of, Supreme Iraqi Islamic Council)
SF13_ENG_027	His former Shiite partners have gathered again to form their own group , the Iraqi National Alliance (INA) , which includes the influential Supreme Iraqi Islamic Council (SIIC) of Ammar al - Hakim , who succeeded his father Abdul Aziz al - Hakim , who died in a hospital in Iran last month after a long battle with cancer .	(Al-Hakim, , who died in, Iran) → (Abdul Aziz Al-Hakim, per:country_of_death, Iran)
SF13_ENG_028	" I ' d rather have Sully doing this than some stranger , or some hotshot trying to be the next Billy Mays , " said the guy who actually is the next Billy Mays , his son Billy Mays III .	(The Next Billy Mays, his son, Billy Mays Iii) → (Billy Mays, per:children, Billy Mays III)
SF13_ENG_029	Fignon continued cycling during and after a stint in the Army , and drew attention in the early 1980s when he managed to keep up with Hinault during a race in which amateurs and professionals rode together .	(Fignon, during and after a stint in, The Army) → (Laurent Fignon, per:employee_or_member_of, the Army)
SF13_ENG_029	Laurent Patrick Fignon was born in Paris on Aug . 12 , 1960 .	(Laurent Patrick Fignon, was born in, Paris) → (Laurent Fignon, per:city_of_birth, Paris)
SF13_ENG_030	Anderson became the Tigers ' manager in June 1979 and built on a foundation that included Alan Trammell at shortstop , Lou Whitaker at second base , Kirk Gibson in the outfield and Jack Morris on the pitching staff .	(Anderson, became, The Tigers' Manager) → (Sparky Anderson, per:title, manager)
SF13_ENG_030	In addition to his wife , Carol , Anderson is survived by his sons , Lee and Albert ; his daughter , Shirlee Englebrecht ; and many grandchildren .	(Anderson, is survived by his sons, Lee) → (Sparky Anderson, per:children, Albert)
SF13_ENG_031	Blake Edwards , a writer and director who was hailed as a Hollywood master of screwball farces and rude comedies like " Victor / Victoria " and the " Pink Panther " movies , died Wednesday night in Santa Monica , Calif . He was 88 .	(Blake Edwards, ,, writer) → (Blake Edwards, per:title, writer)
SF13_ENG_032	Hwang , who lives in Seoul under tight police security , has written books and delivered speeches condemning Kim ' s regime as authoritarian and dictatorial .	(Hwang, who lives, Seoul) → (Hwang Jang-Yop, per:cities_of_residence, Seoul)
SF13_ENG_035	Kaczynska , who was 67 , married Kaczynski in 1978 after meeting him in the northern Polish city of Gdansk , where they were both academics .	(Kaczynska, married, Kaczynski) → (Maria Kaczynska, per:spouse, Kaczynski)
SF13_ENG_036	Upon his release he went into exile in India , where he masterminded the 1973 hijacking of a Royal Nepal Airlines plane known to be carrying hundreds of thousands of dollars in cash to fund his banned Nepali Congress party .	(His Release, exile, India) → (Girija Prasad Koirala, per:countries_of_residence, India)
SF13_ENG_036	Koirala began his political career as a union organiser and was imprisoned for seven years in 1960 after a failed uprising against the monarchy .	(Koirala, began his, A Union Organiser) → (Girija Prasad Koirala, per:title, union organiser)
SF13_ENG_036	Koirala was born in 1925 in Bihar of India where his father Krishna Prasad Koirala and his family were living in exile .	(Koirala, was born in 1925 in, Bihar) → (Girija Prasad Koirala, per:city_of_birth, Bihar)
SF13_ENG_036	Koirala was born in 1925 in Bihar of India at the time when his father Krishna Prasad Koirala along with his family was exiled by Rana rulers .	(Koirala, was born in 1925 in, Bihar) → (Girija Prasad Koirala, per:city_of_birth, Bihar)
SF13_ENG_037	Chabrol ' s survivors also include his third wife , Aurore Pajot , who acted as his script supervisor on nearly all of his movies from 1968 on and whom he married in 1981 ; and Pajot ' s daughter , Cecile Maistre , who was an assistant director on his films and wrote the script with him for " The Girl Cut in Two " (2007) .	(Chabrol'S Survivors, third wife, Aurore Pajot) → (Claude Chabrol, per:spouse, Aurore Pajot)
SF13_ENG_038	The joint statement said Cunningham was " an inspiring performer and dancer into his 80s , and a visionary choreographer and dedicated teacher throughout his life , he led quietly and by example , " the statement said . "	(Cunningham, was ", An Inspiring Performer) → (Merce Cunningham, per:title, performer)
SF13_ENG_038	Merce Cunningham , the nonagenarian choreographer , is planning for a world without him .	(Merce Cunningham, ,, The Nonagenarian Choreographer) → (Merce Cunningham, per:title, choreographer)
SF13_ENG_039	A court on Monday cleared the widower of British reality television star Jade Goody , who died of cancer last year , of rape .	(British Reality Television Star Jade, who died, Cancer) → (Jade Goody, per:cause_of_death, cancer)
SF13_ENG_041	Don Hewitt , the CBS newsman who invented the highly popular TV newsmagazine " 60 Minutes " and produced it for 36 years , died Wednesday .	(Don Hewitt, ,, The Cbs Newsman) → (Don Hewitt, per:title, newsman)
SF13_ENG_041	" He was the consummate television newsman , " Don Hewitt , a longtime CBS News executive and creator of the long - running " 60 Minutes " news program , told Reuters . "	(Don Hewitt, ,, executive) → (Don Hewitt, per:title, executive)
SF13_ENG_041	Hewitt was already a highly respected TV newsman .	(Hewitt, was, A Highly Respected Tv Newsman) → (Don Hewitt, per:title, TV newsman)
SF13_ENG_041	Donald Shepard Hewitt was born in New York on Dec . 14 , 1922 , and grew up in the suburb of New Rochelle .	(Donald Shepard Hewitt, born, New York) → (Don Hewitt, per:stateorprovince_of_birth, New York)
SF13_ENG_043	Eleanor Louise Greenwich was born in Brooklyn on Oct . 23 , 1940 .	(Eleanor Louise Greenwich, was born, Oct.) → (Ellie Greenwich, per:date_of_birth, 1940-10-23)
SF13_ENG_044	A little more than a year after Dunne died from bladder cancer , the colorful remnants of his estate have been consigned by his family to Stair Galleries in Hudson , N . Y . , which will auction them Nov . 20 .	(Dunne, died, Bladder Cancer) → (Dominick Dunne, per:cause_of_death, bladder cancer)
SF13_ENG_048	Charles Gwathmey , an architect known for his influential modernist home designs and famous clients like director Steven Spielberg , has died .	(Charles Gwathmey, ,, architect) → (Charles Gwathmey, per:title, architect)
SF13_ENG_049	Besides his wife , Mandelbrot is survived by two sons , Laurent , of Paris , and Didier , of Newton , Mass . , and three grandchildren .	(Mandelbrot, is survived by two sons ,, Laurent) → (Benoit Mandelbrot, per:children, Laurent)
SF13_ENG_049	For years , he worked for IBM in New York .	(He, for, Ibm) → (Benoit Mandelbrot, per:employee_or_member_of, IBM)
SF13_ENG_049	After several years spent largely at the Centre National de la Recherche Scientifique in Paris , Mandelbrot was hired by IBM in 1958 to work at the Thomas J . Watson Research Center in Yorktown Heights , N . Y . Although he worked frequently with academic researchers and served as a visiting professor at Harvard and the Massachusetts Institute of Technology , it was not until 1987 that he began to teach at Yale , where he earned tenure in 1999 .	(Mandelbrot, was hired by, Ibm) → (Benoit Mandelbrot, per:employee_or_member_of, IBM)
SF13_ENG_056	" It ' s an issue for everybody in the state because peanuts are a big part of our economy , " said Don Koehler , executive director of the Georgia Peanut Commission . "	(The Georgia Peanut Commission, director, Don Koehler) → (Georgia Peanut Commission, org:top_members_employees, Don Koehler)
SF13_ENG_060	" We ' ll be meeting with scientists , university and science policy officials to explore practical opportunities for exchange and collaboration , " Agre , the AAAS president , was quoted as saying .	(The Aaas President, ,, Agre) → (American Association for the Advancement of Science, org:top_members_employees, Peter C. Agre)
SF13_ENG_060	However , Alan Leshner , chief executive officer of the American Association for the Advancement of Science , noted that Nobels are generally given for work that ' s a decade old or more , and that the U . S . mustn ' t become complacent .	(The American Association, chief executive, Alan Leshner) → (American Association for the Advancement of Science, org:top_members_employees, Alan Leshner)
SF13_ENG_062	" First of all , they never have enough funding , " said Andy Kunz , president of the U . S . High Speed Rail Association , a nonprofit that advocates a national high - speed rail network . "	(The U.S. High Speed Rail Association, ,, Andy Kunz) → (U.S. High Speed Rail Association, org:top_members_employees, Andy Kunz)
SF13_ENG_064	China ' s shock at NATO ' s military campaign in the former Yugoslavia helped prod Beijing into playing a bigger role in U . N . peacekeeping , said Bates Gill , director of the Stockholm International Peace Research Institute and co - author of a recent report on China ' s peacekeeping activities .	(The Stockholm International Peace Research Institute, director of, Bates Gill) → (Stockholm International Peace Research Institute, org:top_members_employees, Bates Gill)
SF13_ENG_064	" Non - state actors , for example , a small group of pirates off the coast of Somalia , Al Qaida and Taliban who operate across borders and have more and more sophisticated means of violence , are becoming bigger and bigger challenges to the international system , " said Bates Gill , director of the Stockholm International Peace Research Institute .	(The Stockholm International Peace Research Institute, director, Bates Gill) → (Stockholm International Peace Research Institute, org:top_members_employees, Bates Gill)
SF13_ENG_068	" It is not surprising that one primarily cosmetic business is trying to throw another under the bus by transferring a tax from rich doctors and their wealthy customers to struggling small businesses , " John Overstreet , director of the Indoor Tanning Association , said in a statement Saturday .	(The Indoor Tanning Association, director of, John Overstreet) → (Indoor Tanning Association, org:top_members_employees, John Overstreet)
SF13_ENG_076	The chairman of the Swiss Bankers Association , Patrick Odier , told weekly NZZ am Sonntag that Italy and France have shown interest in deals like ones Switzerland signed this week with Germany and Britain .	(The Swiss Bankers Association, ,, Patrick Odier) → (Swiss Bankers Association, org:top_members_employees, Patrick Odier)
SF13_ENG_076	The majority of voters in Switzerland , which manages more than 25 percent of the world ' s foreign - held private wealth , support banking secrecy , according to a survey published last month by the Swiss Bankers Association in Basel .	(The Swiss Bankers Association, in, Basel) → (Swiss Bankers Association, org:city_of_headquarters, Basel)
SF13_ENG_078	" Americans have a right to know the truth - - Islam is a religion of intolerance and violence , " said Richard Thompson , legal director of the Thomas More Law Center in Ann Arbor . "	(The Thomas More Law Center, in, Ann Arbor) → (Thomas More Law Center, org:city_of_headquarters, Ann Arbor)
SF13_ENG_082	New solutions may be enacted for these orphans , though , said Mary Robinson , CEO of the National Council for Adoption .	(The National Council, ,, Mary Robinson) → (National Council for Adoption, org:top_members_employees, Mary Robinson)
SF13_ENG_082	" When you close a country , you end up causing more problems than you prevented , " said Chuck Johnson , CEO of the National Council for Adoption . "	(Adoption, Council for, Chuck Johnson) → (National Council for Adoption, org:top_members_employees, Chuck Johnson)
SF13_ENG_084	" This is definitely a Goldilocks problem , " said Jason Grumet , president of the Bipartisan Policy Center and an energy adviser to the Obama campaign last year .	(The Bipartisan Policy Center, ,, Jason Grumet) → (Bipartisan Policy Center, org:top_members_employees, Jason Grumet)
SF13_ENG_085	" Banks are in strong need for the capital markets (to raise funds) , " Li Fuan , a director at the China Banking Regulatory Commission , was quoted as saying at a forum over the weekend .	(The China Banking Regulatory Commission, director at, Li Fuan) → (China Banking Regulatory Commission, org:top_members_employees, Li Fuan)
SF13_ENG_089	RIA Novosti and Interfax cite Anatoly Isaikin , head of Rosoboronexport , as saying Thursday " nothing is blocking the continuation of military - technical cooperation " with Iran .	(Rosoboronexport, head of, Anatoly Isaikin) → (Rosoboronexport, org:top_members_employees, Anatoly Isaikin)
SF13_ENG_089	Rosoboronexport is the only company in Russia that is allowed to export arms , dual - use products and military - related services .	(Rosoboronexport, is the only company in, Russia) → (Rosoboronexport, org:country_of_headquarters, Russia)
SF13_ENG_091	With his wife , Cornelie , Middelhoff invested money in 2000 and 2001 with Esch in funds that were formed to buy five properties from KarstadtQuelle , as Arcandor was then known , and leased back to the department store chain before Middelhoff joined the company , according to Middelhoff ' s spokesman .	(Arcandor, ,, Karstadtquelle) → (Arcandor, org:alternate_names, KarstadtQuelle)
SF13_ENG_091	Arcandor ' s mail - order subsidiary Quelle is in worse shape however , Klaus - Hubert Goerg told a press conference in the western city of Essen , where Arcandor is based .	(Arcandor, ,, Essen) → (Arcandor, org:city_of_headquarters, Essen)

Figure 4: $\text{BERT}_{\text{LARGE}}$ on TAC KBP-OIE.

ID	Sentence	Triples to gold triples
SF13_ENG_001	But in a sign of just how disenchanted some Afghans have become with their government, Bashardost, a doctoral scholar who lived in France for two decades, is widely believed to be at least fourth in popularity among 42 candidates in the August elections.	(Bashardost, ,, A Doctoral Scholar) → (Ramazan Bashardost, per:title, scholar)
SF13_ENG_001	Bashardost left Pakistan for France in 1981.	(Bashardost, left, Pakistan) → (Ramazan Bashardost, per:countries_of_residence, Pakistan)
SF13_ENG_007	Mohammed Tantawi, head of Al-Azhar University, told a schoolgirl to remove her niqab when he spotted her during a tour of an Al-Azhar affiliated school, the independent Al-Masry al-Youm newspaper reported this week.	(Mohammed Tantawi, head, Al-Azhar University) → (Mohammed Sayed Tantawi, per:employee_or_member_of, Al-Azhar University)
SF13_ENG_009	He took office in 2006 by defeating longtime incumbent Anwar Chowdhry from Pakistan, who was later barred for alleged financial corruption.	(Longtime Incumbent Anwar Chowdhry, from, Pakistan) → (Anwar Chowdhry, per:origin, Pakistan)
SF13_ENG_010	In addition to his wife, Wendy, Dio is survived by son Daniel, grandchildren Julie and Joey, and father Pat.	(Dio, is survived by son, Daniel) → (Ronnie James Dio, per:children, Daniel)
SF13_ENG_011	Marshall is charged with grand larceny and fraud and faces up to 25 years in prison if convicted.	(Marshall, with, Fraud) → (Anthony Marshall, per:charges, fraud)
SF13_ENG_011	Marshall, a Tony Award-winning Broadway producer and former U.S. diplomat, sat stonelike as the jury forewoman read each verdict aloud, the word "guilty" clearly resonating in the otherwise silent courtroom.	(Marshall, ,, producer) → (Anthony Marshall, per:title, producer)
SF13_ENG_011	The blueblood scion of one of America's most illustrious families appeared to listen impassively as verdicts finding him guilty on 14 counts of grand larceny, conspiracy and fraud were read to a packed courtroom.	(Him, guilty on, Grand Larceny) → (Anthony Marshall, per:charges, grand larceny)
SF13_ENG_011	But Marshall's son, Philip, told a different story.	(Marshall'S Son, ,, Philip) → (Anthony Marshall, per:children, Philip)
SF13_ENG_012	Al-Qaeda's American spokesman Adam Gadahn, also known as Azzam the American, called on Muslims in the West on Sunday to carry out more attacks like the deadly shooting at the US base in Fort Hood, Texas.	(Adam Gadahn, , also known, Azzam) → (Adam Gadahn, per:alternate_names, Azzam)
SF13_ENG_012	Gadahn, also known as Azzam the American, was born in 1978.	(Gadahn, also known as, Azzam) → (Adam Gadahn, per:alternate_names, Azzam)
SF13_ENG_012	Gadahn grew up in California and converted to Islam before he moved to Pakistan in 1998 and attended an al-Qaida training camp six years later, according to media reports.	(Gadahn, in, California) → (Adam Gadahn, per:statesorprovinces_of_residence, California)
SF13_ENG_012	Gadahn moved to Pakistan in 1998, according to the FBI, and is said to have attended an al-Qaida training camp six years later, serving as a translator and consultant for the group.	(Gadahn, moved to, Pakistan) → (Adam Gadahn, per:countries_of_residence, Pakistan)
SF13_ENG_014	After the Munich attack, he lived in Lebanon, Jordan and several Eastern European countries, where he had close ties to Communist bloc intelligence agencies.	(He, lived in, Lebanon) → (Mohammed Oudeh, per:countries_of_residence, Lebanon)
SF13_ENG_015	Clifton attended Howard University but left before graduating to pursue poetry.	(Clifton, attended, Howard University) → (Lucille Clifton, per:schools_attended, Howard University)
SF13_ENG_017	"Alexander Haig devoted his career to serving our country, both as a soldier and as a diplomat," Albright said. "	(Alexander Haig, as, soldier) → (Alexander Haig, per:title, soldier)
SF13_ENG_017	In 1979, he resigned and retired from the Army.	(He, resigned and, The Army) → (Alexander Haig, per:employee_or_member_of, Army)
SF13_ENG_019	McGregor is survived by his wife, Lori, and four children, daughters Jordan, Taylor and Landri, and a son, Logan.	(Mcgregor, by his wife, Lori) → (Keli McGregor, per:spouse, Lori)
SF13_ENG_020	"Mike was a first-rate journalist, a valued member of our staff for 25 years and we will miss him," Times Editor Russ Stanton said. "	(Mike, was, A First-Rate Journalist) → (Mike Penner, per:title, first-rate journalist)
SF13_ENG_020	Penner is survived by his brother, John, a copy editor at the Times, and his former wife, Times sportswriter Lisa Dillman.	(Penner, by his brother, John) → (Mike Penner, per:siblings, John)
SF13_ENG_024	She was charged with theft in Beaumont, Texas, for allegedly failing to pay for $10,000 worth of dental work in 2006.	(She, was charged with, Theft) → (Crystal Taylor, per:charges, theft)
SF13_ENG_025	Michigan native Nancy Kissel was convicted of murder and sentenced in Hong Kong's High Court in September 2005.	(Michigan Native Nancy Kissel, convicted of, Murder) → (Nancy Kissel, per:charges, murder)
SF13_ENG_026	Neal returned to New York and concentrated on stage work.	(Neal, returned, New York) → (Patricia Neal, per:statesorprovinces_of_residence, New York)
SF13_ENG_026	In 1953, she married Roald Dahl, the British writer famed for "Charlie and the Chocolate Factory," "James and the Giant Peach" and other tales for children.	(She, married, Roald Dahl) → (Patricia Neal, per:spouse, Roald Dahl)
SF13_ENG_026	Oscar-winning actress Patricia Neal has died of lung cancer at her home on Martha's Vineyard, Massachusetts, on Sunday.	(Neal, has died of, Lung Cancer) → (Patricia Neal, per:cause_of_death, lung cancer)
SF13_ENG_027	Al-Hakim's son, Ammar al-Hakim, has been groomed for months to take his father's place.	(Al-Hakim'S Son, ,, Ammar Al-Hakim) → (Abdul Aziz Al-Hakim, per:children, Ammar al-Hakim)
SF13_ENG_027	Al-Hakim is the head of Supreme Iraqi Islamic Council (SIIC), the largest Shiite party in Iraq.	(Al-Hakim, of, Supreme Iraqi Islamic Council) → (Abdul Aziz Al-Hakim, per:employee_or_member_of, Supreme Iraqi Islamic Council)
SF13_ENG_027	His former Shiite partners have gathered again to form their own group, the Iraqi National Alliance (INA), which includes the influential Supreme Iraqi Islamic Council (SIIC) of Ammar al- Hakim, who succeeded his father Abdul Aziz al-Hakim, who died in a hospital in Iran last month after a long battle with cancer.	(Al-Hakim, died in a, Iran) → (Abdul Aziz Al-Hakim, per:country_of_death, Iran)
SF13_ENG_028	"I'd rather have Sully doing this than some stranger, or some hotshot trying to be the next Billy Mays," said the guy who actually is the next Billy Mays, his son Billy Mays III.	(The Next Billy Mays, , his son, Billy Mays Iii) → (Billy Mays, per:children, Billy Mays III)
SF13_ENG_029	Fignon continued cycling during and after a stint in the Army, and drew attention in the early 1980s when he managed to keep up with Hinault during a race in which amateurs and professionals rode together.	(Fignon, during and after a stint in, The Army) → (Laurent Fignon, per:employee_or_member_of, the Army)
SF13_ENG_029	Laurent Patrick Fignon was born in Paris on Aug. 12, 1960.	(Laurent Patrick Fignon, was born in, Paris) → (Laurent Fignon, per:city_of_birth, Paris)
SF13_ENG_030	Anderson became the Tigers' manager in June 1979 and built on a foundation that included Alan Trammell at shortstop, Lou Whitaker at second base, Kirk Gibson in the outfield and Jack Morris on the pitching staff.	(Anderson, became, The Tigers' Manager) → (Sparky Anderson, per:title, manager)
SF13_ENG_030	In addition to his wife, Carol, Anderson is survived by his sons, Lee and Albert; his daughter, Shirlee Englebrecht; and many grandchildren.	(Anderson, is survived by his sons, Lee) → (Sparky Anderson, per:children, Lee)
SF13_ENG_031	Blake Edwards, a writer and director who was hailed as a Hollywood master of screwball farces and rude comedies like "Victor/Victoria" and the "Pink Panther" movies, died Wednesday night in Santa Monica, Calif. He was 88.	(Blake Edwards, ,, writer) → (Blake Edwards, per:title, writer)
SF13_ENG_032	Hwang, who lives in Seoul under tight police security, has written books and delivered speeches condemning Kim's regime as authoritarian and dictatorial.	(Hwang, , who lives, Seoul) → (Hwang Jang-Yop, per:cities_of_residence, Seoul)
SF13_ENG_035	Kaczynska, who was 67, married Kaczynski in 1978 after meeting him in the northern Polish city of Gdansk, where they were both academics.	(Kaczynska, married, Kaczynski) → (Maria Kaczynska, per:spouse, Kaczynski)
SF13_ENG_036	Upon his release he went into exile in India, where he masterminded the 1973 hijacking of a Royal Nepal Airlines plane known to be carrying hundreds of thousands of dollars in cash to fund his banned Nepali Congress party.	(His Release, he went into, India) → (Girija Prasad Koirala, per:countries_of_residence, India)
SF13_ENG_036	Koirala began his political career as a union organiser and was imprisoned for seven years in 1960 after a failed uprising against the monarchy.	(Koirala, career as, A Union Organiser) → (Girija Prasad Koirala, per:title, union organiser)
SF13_ENG_036	Koirala was born in 1925 in Bihar of India where his father Krishna Prasad Koirala and his family were living in exile.	(Koirala, was born in 1925 in, Bihar) → (Girija Prasad Koirala, per:city_of_birth, Bihar)
SF13_ENG_036	Koirala was born in 1925 in Bihar of India at the time when his father Krishna Prasad Koirala along with his family was exiled by Rana rulers.	(Koirala, born in 1925, Bihar) → (Girija Prasad Koirala, per:city_of_birth, Bihar)
SF13_ENG_037	Chabrol's survivors also include his third wife, Aurore Pajot, who acted as his script supervisor on nearly all of his movies from 1968 on and whom he married in 1981; and Pajot's daughter, Cecile Maistre, who was an assistant director on his films and wrote the script with him for "The Girl Cut in Two" (2007).	(Chabrol'S Survivors, wife, Aurore Pajot) → (Claude Chabrol, per:spouse, Aurore Pajot)
SF13_ENG_038	The joint statement said Cunningham was "an inspiring performer and dancer into his 80s, and a visionary choreographer and dedicated teacher throughout his life, he led quietly and by example," the statement said. "	(Cunningham, was, An Inspiring Performer) → (Merce Cunningham, per:title, performer)
SF13_ENG_038	Merce Cunningham, the nonagenarian choreographer, is planning for a world without him.	(Merce Cunningham, ,, The Nonagenarian Choreographer) → (Merce Cunningham, per:title, choreographer)
SF13_ENG_039	A court on Monday cleared the widower of British reality television star Jade Goody, who died of cancer last year, of rape.	(Goody, ,, Cancer) → (Jade Goody, per:cause_of_death, cancer)
SF13_ENG_041	Don Hewitt, the CBS newsman who invented the highly popular TV newsmagazine "60 Minutes" and produced it for 36 years, died Wednesday.	(Don Hewitt, ,, The Cbs Newsman) → (Don Hewitt, per:title, newsman)
SF13_ENG_041	"He was the consummate television newsman," Don Hewitt, a longtime CBS News executive and creator of the long-running "60 Minutes" news program, told Reuters. "	(Don Hewitt, ,, executive) → (Don Hewitt, per:title, executive)
SF13_ENG_041	Hewitt was already a highly respected TV newsman.	(Hewitt, was, A Highly Respected Tv Newsman) → (Don Hewitt, per:title, TV newsman)
SF13_ENG_041	Donald Shepard Hewitt was born in New York on Dec. 14, 1922, and grew up in the suburb of New Rochelle.	(Donald Shepard Hewitt, born, New York) → (Don Hewitt, per:stateorprovince_of_birth, New York)
SF13_ENG_043	Eleanor Louise Greenwich was born in Brooklyn on Oct. 23, 1940.	(Eleanor Louise Greenwich, born in, Oct.) → (Ellie Greenwich, per:date_of_birth, 1940-10-23)
SF13_ENG_044	A little more than a year after Dunne died from bladder cancer, the colorful remnants of his estate have been consigned by his family to Stair Galleries in Hudson, N.Y., which will auction them Nov. 20.	(Dunne, died, Bladder Cancer) → (Dominick Dunne, per:cause_of_death, bladder cancer)
SF13_ENG_048	Charles Gwathmey, an architect known for his influential modernist home designs and famous clients like director Steven Spielberg, has died.	(Charles Gwathmey, ,, architect) → (Charles Gwathmey, per:title, architect)
SF13_ENG_049	Besides his wife, Mandelbrot is survived by two sons, Laurent, of Paris, and Didier, of Newton, Mass., and three grandchildren.	(Mandelbrot, two sons, Laurent) → (Benoit Mandelbrot, per:children, Laurent)
SF13_ENG_049	For years, he worked for IBM in New York.	(He, worked, Ibm) → (Benoit Mandelbrot, per:employee_or_member_of, IBM)
SF13_ENG_049	After several years spent largely at the Centre National de la Recherche Scientifique in Paris, Mandelbrot was hired by IBM in 1958 to work at the Thomas J. Watson Research Center in Yorktown Heights, N.Y. Although he worked frequently with academic researchers and served as a visiting professor at Harvard and the Massachusetts Institute of Technology, it was not until 1987 that he began to teach at Yale, where he earned tenure in 1999.	(Mandelbrot, hired by, Ibm) → (Benoit Mandelbrot, per:employee_or_member_of, IBM)
SF13_ENG_056	"It's an issue for everybody in the state because peanuts are a big part of our economy," said Don Koehler, executive director of the Georgia Peanut Commission. "	(The Georgia Peanut Commission, director of, Don Koehler) → (Georgia Peanut Commission, org:top_members_employees, Don Koehler)
SF13_ENG_060	"We'll be meeting with scientists, university and science policy officials to explore practical opportunities for exchange and collaboration," Agre, the AAAS president, was quoted as saying.	(The Aaas President, ,, Agre) → (American Association for the Advancement of Science, org:top_members_employees, Peter C. Agre)
SF13_ENG_060	However, Alan Leshner, chief executive officer of the American Association for the Advancement of Science, noted that Nobels are generally given for work that's a decade old or more, and that the U.S. mustn't become complacent.	(The American Association, ,, Alan Leshner) → (American Association for the Advancement of Science, org:top_members_employees, Alan Leshner)
SF13_ENG_060	Norman Augustine, the former chairman and chief executive of the Lockheed Martin Corporation; Patricia Goldman, former vice chairman of the National Transportation Safety Board; Mary Good, a former president of the American Association for the Advancement of Science; Roger Martin, dean of the Rotman School of Management; Brian O'Neill, a former president of the Insurance Institute for Highway Safety; and Sheila Widnall, a professor at the Massachusetts Institute of Technology.	(The American Association, of, Mary Good) → (American Association for the Advancement of Science, org:top_members_employees, Mary Good)
SF13_ENG_062	"First of all, they never have enough funding," said Andy Kunz, president of the U.S. High Speed Rail Association, a nonprofit that advocates a national high- speed rail network. "	(The U.S. High Speed Rail Association, of, Andy Kunz) → (U.S. High Speed Rail Association, org:top_members_employees, Andy Kunz)
SF13_ENG_064	China's shock at NATO's military campaign in the former Yugoslavia helped prod Beijing into playing a bigger role in U.N. peacekeeping, said Bates Gill, director of the Stockholm International Peace Research Institute and co-author of a recent report on China's peacekeeping activities.	(The Stockholm International Peace Research Institute, ,, Bates Gill) → (Stockholm International Peace Research Institute, org:top_members_employees, Bates Gill)
SF13_ENG_064	"Non-state actors, for example, a small group of pirates off the coast of Somalia, Al Qaida and Taliban who operate across borders and have more and more sophisticated means of violence, are becoming bigger and bigger challenges to the international system," said Bates Gill, director of the Stockholm International Peace Research Institute.	(The Stockholm International Peace Research Institute, , director of, Bates Gill) → (Stockholm International Peace Research Institute, org:top_members_employees, Bates Gill)
SF13_ENG_068	"It is not surprising that one primarily cosmetic business is trying to throw another under the bus by transferring a tax from rich doctors and their wealthy customers to struggling small businesses," John Overstreet, director of the Indoor Tanning Association, said in a statement Saturday.	(The Indoor Tanning Association, director of, John Overstreet) → (Indoor Tanning Association, org:top_members_employees, John Overstreet)
SF13_ENG_076	The chairman of the Swiss Bankers Association, Patrick Odier, told weekly NZZ am Sonntag that Italy and France have shown interest in deals like ones Switzerland signed this week with Germany and Britain.	(The Swiss Bankers Association, ,, Patrick Odier) → (Swiss Bankers Association, org:top_members_employees, Patrick Odier)
SF13_ENG_076	The majority of voters in Switzerland, which manages more than 25 percent of the world's foreign-held private wealth, support banking secrecy, according to a survey published last month by the Swiss Bankers Association in Basel.	(The Swiss Bankers Association, in, Basel) → (Swiss Bankers Association, org:city_of_headquarters, Basel)
SF13_ENG_078	"Americans have a right to know the truth – Islam is a religion of intolerance and violence," said Richard Thompson, legal director of the Thomas More Law Center in Ann Arbor. "	(The Thomas More Law Center, in, Ann Arbor) → (Thomas More Law Center, org:city_of_headquarters, Ann Arbor)
SF13_ENG_082	New solutions may be enacted for these orphans, though, said Mary Robinson, CEO of the National Council for Adoption.	(The National Council, of, Mary Robinson) → (National Council for Adoption, org:top_members_employees, Mary Robinson)
SF13_ENG_082	"When you close a country, you end up causing more problems than you prevented," said Chuck Johnson, CEO of the National Council for Adoption. "	(The National Council, CEO of, Chuck Johnson) → (National Council for Adoption, org:top_members_employees, Chuck Johnson)
SF13_ENG_084	"This is definitely a Goldilocks problem," said Jason Grumet, president of the Bipartisan Policy Center and an energy adviser to the Obama campaign last year.	(The Bipartisan Policy Center, ,, Jason Grumet) → (Bipartisan Policy Center, org:top_members_employees, Jason Grumet)
SF13_ENG_085	"Banks are in strong need for the capital markets (to raise funds)," Li Fuan, a director at the China Banking Regulatory Commission, was quoted as saying at a forum over the weekend.	(The China Banking Regulatory Commission, a director, Li Fuan) → (China Banking Regulatory Commission, org:top_members_employees, Li Fuan)
SF13_ENG_089	RIA Novosti and Interfax cite Anatoly Isaikin, head of Rosoboronexport, as saying Thursday "nothing is blocking the continuation of military-technical cooperation" with Iran.	(Rosoboronexport, , head of, Anatoly Isaikin) → (Rosoboronexport, org:top_members_employees, Anatoly Isaikin)
SF13_ENG_089	Rosoboronexport is the only company in Russia that is allowed to export arms, dual-use products and military-related services.	(Rosoboronexport, is the only company in, Russia) → (Rosoboronexport, org:country_of_headquarters, Russia)
SF13_ENG_091	With his wife, Cornelie, Middelhoff invested money in 2000 and 2001 with Esch in funds that were formed to buy five properties from KarstadtQuelle, as Arcandor was then known, and leased back to the department store chain before Middelhoff joined the company, according to Middelhoff's spokesman.	(Arcandor, ,, Karstadtquelle) → (Arcandor, org:alternate_names, KarstadtQuelle)
SF13_ENG_091	Arcandor's mail-order subsidiary Quelle is in worse shape however, Klaus-Hubert Goerg told a press conference in the western city of Essen, where Arcandor is based.	(Arcandor, ,, Essen) → (Arcandor, org:city_of_headquarters, Essen)

Figure 5: GPT-2_{XL} on TAC KBP-OIE.

ID	Sentence	Triples to gold triples BERT$_{\text{LARGE}}$	GPT-2$_{\text{XL}}$
SF13_ENG_005	The men , who live in the Chicago area , were identified as David Coleman Headley , 49 , a US citizen , and Tahawwur Hussain Rana , 48 , Rana , a native of Pakistan and citizen of Canada .	(Rana, , a native of, Canada) $\rightarrow$ (Tahawwur Hussain Rana, per:origin, Canada)	-
SF13_ENG_006	Khan died at his music centre near San Francisco in the United States after a prolonged kidney ailment , the news agency said .	(Khan, died at his, The United States) $\rightarrow$ (Ali Akbar Khan, per:country_of_death, United States)	-
SF13_ENG_006	By his early 20s he was music director of All-India Radio in Lucknow, broadcasting as a solo artist and composing for the radio's orchestra.	-	(He, of, All-India Radio) $\rightarrow$ (Ali Akbar Khan, per:employee_or_member_of, All-India Radio)
SF13_ENG_006	Defying his father , Khan moved to Bombay and began writing scores for films , including Chetan Anand ' s " Aandhiyan " (1952) , Satyajit Ray ' s " Devi " (1960) and Tapan Sinha ' s " Hungry Stones " (1960) .	(Khan, moved, Bombay) $\rightarrow$ (Ali Akbar Khan, per:cities_of_residence, Bombay)	-
SF13_ENG_007	In 1966 he graduated from the faculty of theology at Al-Azhar, the 10th century university that has trained the majority of Sunni Muslim clerics from Africa and Asia.	-	(He, graduated from, Al-Azhar) $\rightarrow$ (Mohammed Sayed Tantawi, per:schools_attended, Al-Azhar)
SF13_ENG_009	He became AIBA president in 2006 by defeating longtime incumbent Anwar Chowdhry of Pakistan, who was later barred for alleged financial corruption.	-	(Longtime Incumbent Anwar Chowdhry, by, Aiba President) $\rightarrow$ (Anwar Chowdhry, per:employee_or_member_of, AIBA)
SF13_ENG_011	Marshall could get a lengthy prison term if convicted of grand larceny.	-	(Marshall, of, Grand Larceny) $\rightarrow$ (Anthony Marshall, per:charges, larceny)
SF13_ENG_012	Gadahn , also known as Azzam al - Amriki , was dressed in white robes and wearing a white turban as he called for attacks on what he described as " high - value targets . "	(Gadahn, known as, Azzam Al-Amriki) $\rightarrow$ (Adam Gadahn, per:alternate_names, Azzam al-Amriki)	-
SF13_ENG_012	Gadahn grew up on a farm in California and converted to Islam before moving to Pakistan in 1998 and reportedly attending an al - Qaida training camp .	(Gadahn, to, Pakistan) $\rightarrow$ (Adam Gadahn, per:countries_of_residence, Pakistan)	-
SF13_ENG_012	He was born in Oregon on Sept. 1, 1978 to a Jewish family.	-	(He, born in, Sept.) $\rightarrow$ (Adam Gadahn, per:date_of_birth, 1978-09-01)
SF13_ENG_015	Besides her sister , Philip , survivors include a son , Graham ; three daughters , Sidney , Gillian and Alexia ; and three grandchildren .	(Her Sister, son ,, Graham) $\rightarrow$ (Lucille Clifton, per:children, Graham)	-
SF13_ENG_015	In addition to her daughter , of Columbia , other survivors include two daughters , Sidney Clifton of Los Angeles and Gillian Clifton - Monell of Fort Washington , Md . , ; a son , Graham Clifton of Wichita ; a sister ; and three grandchildren .	(Her Daughter, ,, Sidney Clifton) $\rightarrow$ (Lucille Clifton, per:children, Sidney Clifton)	-
SF13_ENG_017	Haig , 85 , died early Saturday in a Baltimore , Maryland , hospital .	(Haig, died early, A Baltimore, Maryland, Hospital) $\rightarrow$ (Alexander Haig, per:stateorprovince_of_death, Maryland)	-
SF13_ENG_019	McGregor was born in Iowa , but grew up in Colorado where his father , Brian , was a coaching legend at Arvada West .	(Mcgregor, but grew up, Brian) $\rightarrow$ (Keli McGregor, per:parents, Brian)	-
SF13_ENG_025	Kissel was born in Adrian, Michigan, but her family had also lived in Minneapolis.	-	(Kissel, in, Minneapolis) $\rightarrow$ (Nancy Kissel, per:cities_of_residence, Minneapolis)
SF13_ENG_025	Kissel, 46, of Adrian, Michigan, was convicted of drugging then bashing her husband Robert to death in a luxury Hong Kong apartment and sentenced to life in prison in September 2005.	-	(Kissel, of, Michigan) $\rightarrow$ (Nancy Kissel, per:statesorprovinces_of_residence, Michigan)
SF13_ENG_026	Neal had lung cancer and died at her home in Edgartown , Massachusetts on Martha ' s Vineyard , said longtime friend Bud Albers of Knoxville .	(Neal, died at her home in, Edgartown) $\rightarrow$ (Patricia Neal, per:city_of_death, Edgartown)	-
SF13_ENG_026	She grew up in Knoxville, Tenn., where Neal was turned onto acting at 10 after seeing an impassioned speech about "demon rum" at the local temperance union speaking contest.	-	(She, in, Tenn.) $\rightarrow$ (Patricia Neal, per:statesorprovinces_of_residence, Tennessee)
SF13_ENG_027	Al-Hakim, who died Wednesday of lung cancer in Tehran, was a symbol for many of the re-emergence of Iraq's Shiite political majority after decades of oppression under Saddam Hussein's Sunni-led regime.	-	(Al-Hakim, who died, Tehran) $\rightarrow$ (Abdul Aziz Al-Hakim, per:city_of_death, Tehran)
SF13_ENG_028	TAMPA Heart disease , exacerbated by cocaine use , killed celebrated TV pitchman Billy Mays , according to the Hillsborough County medical examiner ' s final autopsy report released Friday .	(Celebrated Tv Pitchman Billy Mays, killed, Tampa) $\rightarrow$ (Billy Mays, per:cause_of_death, TAMPA Heart disease)	-
SF13_ENG_030	Family spokesman Dan Ewald said Anderson died from complications from dementia .	(Anderson, died from, Dementia) $\rightarrow$ (Sparky Anderson, per:cause_of_death, dementia)	-
SF13_ENG_031	He served in the Coast Guard during World War II and had a rare leading role in the low - budget thriller " Strangler of the Swamp " (1946) .	(He, in, The Coast Guard) $\rightarrow$ (Blake Edwards, per:employee_or_member_of, Coast Guard)	-
SF13_ENG_031	His agent Lou Pitt confirmed he had died , while the " Entertainment Tonight " news show and website reported that he died in Santa Monica on Wednesday night from complications of pneumonia , with Andrews at his side .	(He, died, Complications) $\rightarrow$ (Blake Edwards, per:cause_of_death, complications)	-
SF13_ENG_033	In a career that spanned seven decades, Ginzburg authored several groundbreaking studies in various fields – such as quantum theory, astrophysics, radio-astronomy and diffusion of cosmic radiation in the Earth's atmosphere – that were of "Nobel Prize caliber," said Gennady Mesyats, the director of the Lebedev Physics Institute in Moscow, where Ginzburg worked.	-	(Ginzburg, in, The Lebedev Physics Institute) $\rightarrow$ (Vitaly Ginzburg, per:employee_or_member_of, Lebedev Physics Institute)
SF13_ENG_035	Kaczynska , who died aged 59 , married Kaczynski in 1978 .	(Kaczynska, married, Kaczynski) $\rightarrow$ (Maria Kaczynska, per:spouse, Kaczynski)	-
SF13_ENG_037	Prime Minister Francois Fillon called Chabrol a " great director , producer and screenwriter (who) was one of the grand figures of the ' Nouvelle vague , ' which revolutionized the style and techniques of cinema by looking at real experience , true life , that which is indiscreet and subtle . "	(Afp Chabrol, was, director) $\rightarrow$ (Claude Chabrol, per:title, film director); (Chabrol, director, Producer) $\rightarrow$ (Claude Chabrol, per:title, producer)	-
SF13_ENG_038	Cunningham's survivors include a brother, Jack Cunningham of Centralia.	-	(Cunningham'S Survivors, brother, Jack Cunningham) $\rightarrow$ (Merce Cunningham, per:siblings, Jack Cunningham)
SF13_ENG_038	Mercier Philip Cunningham was born in Centralia, Wash., on April 16, 1919.	-	(Mercier Philip Cunningham, was born in, April) $\rightarrow$ (Merce Cunningham, per:date_of_birth, 1919-04-16)
SF13_ENG_041	A year after enrolling in New York University on a track scholarship , Hewitt dropped out .	(Hewitt, ,, New York University) $\rightarrow$ (Don Hewitt, per:schools_attended, New York University)	-
SF13_ENG_041	By the time Hewitt was 6, his family was living in New Rochelle, N.Y.	-	(Hewitt, in, New Rochelle) $\rightarrow$ (Don Hewitt, per:cities_of_residence, New Rochelle)
SF13_ENG_043	Greenwich also worked as an arranger and singer , a role that saw her working with artists including Frank Sinatra and Ella Fitzgerald .	(Greenwich, worked as an, Singer) $\rightarrow$ (Ellie Greenwich, per:title, singer)	-
SF13_ENG_044	Dunne was born in 1925 in Hartford, Connecticut, to a wealthy Roman Catholic family and grew up in some of the same social circles as the Kennedys.	-	(Dunne, was born in 1925 in, Connecticut) $\rightarrow$ (Dominick Dunne, per:stateorprovince_of_birth, Connecticut)
SF13_ENG_044	In 1957 , Dunne moved to Los Angeles to work on the CBS showcase " Playhouse 90 . "	(Dunne, moved to, Los Angeles) $\rightarrow$ (Dominick Dunne, per:cities_of_residence, Los Angeles)	-
SF13_ENG_044	Dunne found his greatest prominence as a celebrity journalist while covering the 1995 murder trial of football star and actor O.J. Simpson, who had been accused of killing his ex-wife Nicole Brown Simpson and her friend Ronald Goldman.	-	(Dunne, as, A Celebrity Journalist) $\rightarrow$ (Dominick Dunne, per:title, celebrity journalist)
SF13_ENG_045	Goldstein died in New York on Aug . 28 .	(Goldstein, died, Aug.) $\rightarrow$ (Adam Goldstein, per:date_of_death, XXXX-08-28)	-
SF13_ENG_046	After acting in movies and television , Meredith receded into a quiet life in Santa Fe , writing , painting , golfing and acting in a stage production of " The Odd Couple . "	(Meredith, in, Santa Fe) $\rightarrow$ (Don Meredith, per:cities_of_residence, Santa Fe)	-
SF13_ENG_046	In high school and at Southern Methodist University , where , already known as Dandy Don (a nickname bestowed on him by his brother) , Meredith became an all - American .	(Dandy Don, where ,, Southern Methodist University) $\rightarrow$ (Don Meredith, per:schools_attended, Southern Methodist University)	-
SF13_ENG_046	He spent much of his life backing away from the nickname Dandy Don , particularly during his secluded later decades in New Mexico .	(He, the nickname, Dandy Don) $\rightarrow$ (Don Meredith, per:alternate_names, Dandy Don)	-
SF13_ENG_048	The group, known variously as the Five or the Whites (for the color of most of their buildings) or the New York School, consisted of Gwathmey, Michael Graves, Eisenman, John Hejduk and Richard Meier.	-	(Gwathmey, of, The New York School) $\rightarrow$ (Charles Gwathmey, per:employee_or_member_of, New York School)
SF13_ENG_048	Meier , who said he had known Gwathmey for 50 years , has particularly fond memories from the time when Gwathmey was first courting his second wife , Bette - Ann Damson , and they all picked corn for dinner in a field adjacent to a barn Meier was renting on the East End of Long Island .	(Gwathmey, his second wife ,, Bette-Ann Damson) $\rightarrow$ (Charles Gwathmey, per:spouse, Bette-Ann Damson)	-
SF13_ENG_057	By 2013, automakers will have dozens of plug-in electric hybrid vehicles and fully electric vehicles, said Jason Forcier, a vice president at battery maker A123 Systems Inc.	-	(A123 Systems Inc., a vice president, Jason Forcier) $\rightarrow$ (A123 Systems Inc., org:top_members_employees, Jason Forcier)
SF13_ENG_058	Access Industries , a privately held company founded in 1986 by Len Blavatnik , has a diverse portfolio of investments in industry , real estate , media and telecommunications .	(Access Industries, founded in 1986 by, Len Blavatnik) $\rightarrow$ (Access Industries, org:founded_by, Len Blavatnik)	-
SF13_ENG_069	"The indications are positive," said Vincent Cogliano, director of the Monographs program at IARC, which decides on carcinogen classifications.	-	(Iarc, , director of the, Vincent Cogliano) $\rightarrow$ (International Agency for Research on Cancer, org:top_members_employees, Vincent Cogliano)
SF13_ENG_082	"There hasn't been a concerted push to open doors for Muslim orphans because the expectation would be that those efforts would fall flat," said Chuck Johnson, chief executive of the National Council for Adoption, a policy group in Alexandria, Va.	-	(Adoption, in, Alexandria) $\rightarrow$ (National Council for Adoption, org:city_of_headquarters, Alexandria)
SF13_ENG_084	" The scenario itself is secret , " said Eileen McMenamin , vice president of communications for the Bipartisan Policy Center (BPC) , which is hosting the event dubbed " Cyber ShockWave . "	(The Bipartisan Policy Center, , vice president of, Eileen Mcmenamin) $\rightarrow$ (Bipartisan Policy Center, org:top_members_employees, Eileen McMenamin)	-
SF13_ENG_088	Bank Julius Baer Co., based in Basel, Switzerland, sued because WikiLeaks posted accountholder information from its Cayman outpost amid allegations of money laundering and tax evasion.	-	(Bank Julius Baer Co., based, Basel) $\rightarrow$ (Bank Julius Baer, org:city_of_headquarters, Basel)
SF13_ENG_094	The other manufacturers are two US companies, L-3 Communications and Rapiscan Systems, a unit of OSI Systems, and British rival Smiths Detection.	-	(Rapiscan, ,, Two Us Companies) $\rightarrow$ (Rapiscan Systems, org:country_of_headquarters, US)
SF13_ENG_096	Freedom was founded in the 1930s by R.C. Hoiles and is still majority owned by the Hoiles family.	-	(Freedom, was founded in the, R.C. Hoiles) $\rightarrow$ (Freedom Communications, org:founded_by, R.C. Hoiles)
SF13_ENG_098	Shares of Millipore have more than doubled since Martin D . Madaus took over as chief executive in January 2005 .	(Millipore, have, Martin D.) $\rightarrow$ (Millipore Corp., org:top_members_employees, Martin D. Madaus)	-

Figure 6: BERT$_{\text{LARGE}}$ vs. GPT-2$_{\text{XL}}$ on TAC KBP-OIE.

ID	Sentence	Triples to gold triples
Appropriate_Rural_Technology_Institute	Appropriate Rural Technology Institute Appropriate Rural Technology Institute is an Indian non - governmental organization founded by a group of about 20 scientists and technologists to develop innovative and environmentally friendly rural technologies based on modern scientific knowledge in 1996 .	(Appropriate Rural Technology Institute, is, An Indian Non-Governmental Organization) → (appropriate_rural_technology_institute.Q16245764, country.P17, india.Q668)
Ayatollah_(Record_Producer)	Ayatollah Lamont Dorrell , known as Ayatollah , is a hip - hop record producer from Queens , New York who has produced for predominantly New York - based rappers including Mos Def , Talib Kweli , R . A . The Rugged Man , Tragedy Khadafi , Wordsworth , Vast Aire , Afu - Ra , Guru , M . O . P . , Inspectah Deck , Cormega , Ghostface Killah as well as many others .	(Ayatollah, is, A Hip-Hop Record Producer) → (ayatollah_(record_producer).Q4830984, occupation.P106, record_producer.Q183945)
Benjamin_Darbelet	Benjamin Darbelet Benjamin Darbelet is a French judoka .	(Benjamin Darbelet, is, A French Judoka) → (benjamin_darbelet.Q2456207, country_of_citizenship.P27, france.Q142); (Benjamin Darbelet, is, A French Judoka) → (benjamin_darbelet.Q2456207, sport.P641, judo.Q11420)
Bharat_Khawas	He plays for club Tribhuvan Army in Martyr ' s Memorial A - Division League as a striker .	(He, plays, Club Tribhuvan Army) → (bharat_khawas.Q4901128, member_of_sports_team.P54, nepal_army_club.Q6994327); (He, plays, Club Tribhuvan Army) → (bharat_khawas.Q4901128, sport.P641, association_football.Q2736)
Bill_White_(Shortstop)	White was born on May 1 , 1860 in Bridgeport , Ohio , and he began his professional baseball career in 1883 with the Pottsville Anthracites of the International Association .	(White, in, Bridgeport) → (bill_white_(shortstop).Q4911357, place_of_birth.P19, bridgeport,_ohio.Q2062947)
Bob_Dylan.Q392	Dylan married Sara Lownds , who had worked as a model and a secretary to Drew Associates , on November 22 , 1965 .	(Dylan, married, Sara Lownds) → (Bob_Dylan.Q392, spouse.P26, Sara_Lownds.Q457433)
Bob_Dylan.Q392	In 1991 , Dylan received a Grammy Lifetime Achievement Award from American actor Jack Nicholson .	(Dylan, received, Grammy Lifetime Achievement Award) → (Bob_Dylan.Q392, award_received.P166, Grammy_Lifetime_Achievement_Award.Q935843)
Boog_Powell	He also played for the Cleveland Indians and the Los Angeles Dodgers .	(He, also played for the, Cleveland Indians) → (boog_powell.Q4942810, member_of_sports_team.P54, cleveland_indians.Q642553)
Boog_Powell	He played in Major League Baseball as a first baseman and left fielder from through , most notably as a member of the Baltimore Orioles dynasty that won four American League pennants and two World Series championships between 1966 and 1971 .	(He, played, A First Baseman) → (boog_powell.Q4942810, position_played_on_team_/_speciality.P413, first_baseman.Q1326154)
Bouchraya_Hammoudi_Bayoun	He studied Economics at the University of Havana , Cuba .	(He, studied Economics, Havana) → (bouchraya_hammoudi_bayoun.Q4949324, educated_at.P69, university_of_havana.Q837320)
Cezaro_Rossetti	Of Italian - Swiss derivation , he was born in Glasgow and lived in Britain .	(He, was, Glasgow) → (cezaro_rossetti.Q2947206, place_of_birth.P19, glasgow.Q4093)
Chan_Sau_Ying	Chan Sau Ying Chan Sau Ying is a retired athlete from Hong Kong who specialised in the 100 metres hurdles .	(Chan Sau Ying, is, A Retired Athlete) → (chan_sau_ying.Q21858345, sport.P641, sport_of_athletics.Q542)
Christian_Mcgrath	He was born in the Bronx in 1972 and became interested in art and illustration as a small child .	(He, was born in, The Bronx) → (christian_mcgrath.Q5109897, place_of_birth.P19, the_bronx.Q18426)
Cliff_Michelmore	He was a regular presenter on BBC1 ' s " Holiday " programme from 1969 to 1986 , and presented other shows for BBC TV , ITV and BBC Radio .	(He, was, A Regular Presenter) → (cliff_michelmore.Q5132704, occupation.P106, television_presenter.Q947873)
Crooked_I	Crooked I Dominick Antron Wickliffe , better known by his stage name Crooked I or Kxng Crooked , is an American rapper from Long Beach , California .	(Kxng Crooked, ,, An American Rapper) →(crooked_i.Q253695, country_of_citizenship.P27, united_states.Q30)
Denise_Johns	She graduated in 2002 , and moved to Atlanta , playing amateur beach volleyball on the east coast of America .	(She, , playing, Amateur Beach Volleyball) → (denise_johns.Q5257690, sport.P641, beach_volleyball.Q4543)
Diana_Rast	She also competed at the 1996 Summer Olympics .	(She, competed at, The 1996 Summer Olympics) → (diana_rast.Q2491787, participant_of.P1344, 1996_summer_olympics.Q8531)
Eduard_Friedrich_Poeppig	Eduard Friedrich Poeppig Eduard Friedrich Poeppig was a German botanist , zoologist and explorer .	(Eduard Friedrich, was, A German Botanist) → (eduard_friedrich_poeppig.Q61501, field_of_work.P101, botany.Q441)
Edward_L._Romero	After the Vietnam war , he became a county chairman of Democratic Party .	(A County Chairman, of, Democratic Party) → (edward_l._romero.Q5344025, member_of_political_party.P102, democratic_party_(united_states).Q29552)
Endre_Tilli	Endre Tilli Endre Tilli was a Hungarian fencer .	(Endre Tilli, was, A Hungarian Fencer) → (endre_tilli.Q1642140, country_of_citizenship.P27, hungary.Q28); (Endre Tilli, was, A Hungarian Fencer) → (endre_tilli.Q1642140, sport.P641, fencing.Q12100)
Enrico_Porro	Porro competed at the 1908 Summer Olympics in London where he won the gold medal in Greco - Roman wrestling , the lightweight class .	(Porro, competed, The 1908 Summer Olympics) → (enrico_porro.Q1343905, participant_of.P1344, 1908_summer_olympics.Q8111)
Erfan_Zeneli	Erfan Zeneli Erfan Zeneli is a Finnish footballer of Kosovar Albanian descent who currently plays for HJK Helsinki as a striker or left winger .	(Erfan Zeneli, is, A Finnish Footballer) → (erfan_zeneli.Q2053367, country_of_citizenship.P27, finland.Q33); (Erfan Zeneli, is, A Finnish Footballer) → (erfan_zeneli.Q2053367, sport.P641, association_football.Q2736); (Erfan Zeneli, is, A Finnish Footballer) → (erfan_zeneli.Q2053367, country_for_sport.P1532, finland.Q33)
Erminio_Favalli	He then moved to Mantova before to join Palermo , where he spent seven seasons as a " rosanero " mainstay .	(He, then moved to, Mantova) → (erminio_favalli.Q1405871, member_of_sports_team.P54, mantova_1911_s.s.d..Q430993)
George_Dowty	Dowty was born in Pershore , Worcestershire in 1901 .	(Dowty, was born in, Pershore) → (george_dowty.Q1507239, place_of_birth.P19, pershore.Q767858)
Harald_Fairhair	In the Saga of Harald Fairhair in " Heimskringla " , which is the most elaborate although not the oldest or most reliable source to the life of Harald , it is written that Harald succeeded , on the death of his father Halfdan the Black Gudrödarson , to the sovereignty of several small , and somewhat scattered kingdoms in Vestfold , which had come into his father ' s hands through conquest and inheritance .	(Harald Fairhair, of his father, Halfdan the Black) → (harald_fairhair.Q206123, father.P22, halfdan_the_black.Q504932)
Hector_Dyer	Born in Los Angeles , California , Hector Dyer enrolled at the Stanford University and won the IC4A championships in in 1930 .	(Hector Dyer, enrolled at, The Stanford University) → (hector_dyer.Q1364518, educated_at.P69, stanford_university.Q41506)
Henny_Magnussen	Thanks to the support she received , she was able to practise in all the courts of Denmark from 1909 , becoming the first woman to do so .	(The First Woman, ,, Denmark) → (henny_magnussen.Q56577333, country_of_citizenship.P27, denmark.Q35)
Hina_Pervaiz_Butt	Hina Pervaiz Butt Hina Pervaiz Butt is a Pakistani politician who was a Member of the Provincial Assembly of the Punjab , since May 2013 .	(Hina Pervaiz Butt, is, A Pakistani Politician) → (hina_pervaiz_butt.Q47918210, occupation.P106, politician.Q82955)
Isabella_Of_Valois,_Duchess_Of_Bourbon	She was the wife of Peter I , Duke of Bourbon .	(She, was the wife of, Peter I) → (isabella_of_valois,_duchess_of_bourbon.Q2363495, spouse.P26, peter_i,_duke_of_bourbon.Q444668)
Jahir_Butrón	Then from 1995 , he played for Peruvian Second Division side Guardia Republicana .	(He, played for, Peruvian Second Division Side Guardia Republicana) → (jahir_butrón.Q6123259, member_of_sports_team.P54, guardia_republicana.Q5614012)
Jahir_Butrón	Jahir Butrón Jahir Butrón Gotuzzo is a Peruvian footballer who plays as a center back .	(Jahir Butrón Gotuzzo, is, A Peruvian Footballer) → (jahir_butrón.Q6123259, country_of_citizenship.P27, peru.Q419); (Jahir Butrón Gotuzzo, is, A Peruvian Footballer) → (jahir_butrón.Q6123259, sport.P641, association_football.Q2736)
Joan_Of_France,_Duchess_Of_Berry	She was the second daughter of King Louis XI of France and of his second wife Charlotte of Savoy ; her surviving siblings were King Charles VIII of France and Anne of France .	(She, was the second daughter, King Louis Xi) → (joan_of_france,_duchess_of_berry.Q236220, father.P22, louis_xi_of_france.Q8058)
John_Ogilvie_(Footballer)	He played for Hibs in their most successful era renowned for their front five known as The Famous Five .	(He, played for, Hibs) → (john_ogilvie_(footballer).Q6251147, member_of_sports_team.P54, hibernian_f.c..Q192597)
Johnedel_Cardel	He moved to Sta .	(He, to, Sta) → (johnedel_cardel.Q56599981, member_of_sports_team.P54, sta._lucia_realtors.Q3547751)
Joseph_W._Nega	Joseph W . Nega Joseph W . Nega is a Judge of the United States Tax Court .	(Joseph W. Nega, is, A Judge) → (joseph_w._nega.Q16733242, occupation.P106, judge.Q16533)
José_Alejandro_Aguilar_López	José Alejandro Aguilar López José Alejandro Aguilar López is a Mexican politician from the National Action Party .	(José Alejandro Aguilar López, is, A Mexican Politician) → (josé_alejandro_aguilar_lópez.Q18385891, country_of_citizenship.P27, mexico.Q96); (José Alejandro Aguilar López, is, A Mexican Politician) → (josé_alejandro_aguilar_lópez.Q18385891, occupation.P106, politician.Q82955); (José Alejandro Aguilar López, Mexican politician, The National Action Party) → (josé_alejandro_aguilar_lópez.Q18385891, member_of_political_party.P102, national_action_party_(mexico).Q851087)
Juan_Bautista_Azopardo	Juan Bautista Azopardo was born in Senglea , Malta , the son of Rosina and Salvatore Azopardo .	(Juan Bautista Azopardo, ,, Malta) → (juan_bautista_azopardo.Q3805864, country_of_citizenship.P27, malta.Q233); (Juan Bautista Azopardo, was born, Senglea) → (juan_bautista_azopardo.Q3805864, place_of_birth.P19, senglea.Q846593)
Kevin_Kerr_(Scottish_Footballer)	He signed for AGOVV Apeldoorn of Netherlands ' Eerste Divisie in January 2012 after his contract in Bielefeld had expired the previous summer .	(He, signed for, Agovv Apeldoorn) → (kevin_kerr_(scottish_footballer).Q326761, member_of_sports_team.P54, agovv_apeldoorn.Q292618)
Lachezar_Kotev	Lachezar Kotev Lachezar Kotev is a Bulgarian footballer who plays as a midfielder for Vitosha Bistritsa .	(Lachezar Kotev, is, A Bulgarian Footballer) → (lachezar_kotev.Q49560773, sport.P641, association_football.Q2736); (Lachezar Kotev, is, A Midfielder) → (lachezar_kotev.Q49560773, position_played_on_team_/_speciality.P413, midfielder.Q193592)
Marc_Zeno	In 1988 , Zeno was drafted in the 7th round by the Pittsburgh Steelers .	(Zeno, was drafted in the, The Pittsburgh Steelers) → (marc_zeno.Q16196353, member_of_sports_team.P54, pittsburgh_steelers.Q191477)
Marcel_Dheere	Marcel Dheere Marcel Albert " Ching " Dheere was a Canadian professional ice hockey forward who played 11 games in the National Hockey League for the Montreal Canadiens .	(Marcel Albert "Ching, in, A Canadian Professional Ice Hockey) → (marcel_dheere.Q6756083, country_of_citizenship.P27, canada.Q16); (The National Hockey League, for, The Montreal Canadiens) → (marcel_dheere.Q6756083, member_of_sports_team.P54, montreal_canadiens.Q188143); (A Canadian Professional Ice Hockey, played, The National Hockey League) → (marcel_dheere.Q6756083, league.P118, national_hockey_league.Q1215892); (Marcel Albert "Ching, played, A Canadian Professional Ice Hockey) → (marcel_dheere.Q6756083, sport.P641, ice_hockey.Q41466)
Mark_Schultz_(Wrestler)	Mark Schultz was born in 1960 in Palo Alto , California to Dorothy Jean St . Germain and Philip Gary Schultz .	(Mark Schultz, was born in 1960 in, Dorothy Jean St. Germain) → (mark_schultz_(wrestler).Q3487332, place_of_birth.P19, palo_alto,_california.Q47265)
Michael_Jones_(Footballer)	Michael Jones Michael Jones is an English footballer , who plays for Cefn Druids .	(Michael Jones, is, An English Footballer) → (michael_jones_(footballer).Q6831707, sport.P641, association_football.Q2736)
Mike_Bellotti	He is an alumnus of the Delta Sigma Phi fraternity .	(He, is an alumnus of, The Delta Sigma Phi Fraternity) → (mike_bellotti.Q14950800, educated_at.P69, california_state_university,_east_bay.Q1026916)
Olga_Semenova_Tyan-Shanskaya	She learned to play chess at the age of 18 ; four years later she won the Leningrad city women ' s chess championship .	(She, to play, Chess) → (olga_semenova_tyan-shanskaya.Q2392240, sport.P641, chess.Q718)
Owen_Paterson	Owen Paterson Owen William Paterson is a British Conservative Party politician who was the Secretary of State for Environment , Food and Rural Affairs from 2012 to 2014 .	(Owen Paterson, is, A British Conservative Party Politician) → (owen_paterson.Q197894, occupation.P106, politician.Q82955)
Pan_Gongsheng	Pan Gongsheng Pan Gongsheng is a Chinese economist , banker , reformist and bureaucrat .	(Pan Gongsheng, is, A Chinese Economist) → (pan_gongsheng.Q9309080, country_of_citizenship.P27, china.Q148)
Paul Carrington (American Football)	Paul Carrington Paul Carrington is a former American football defensive end .	(Paul Carrington, is, A Former American Football Defensive End) → (paul_carrington_(american_football).Q9057011, sport.P641, american_football.Q41323)
Peter_Jebsen	Peter Jebsen Peter Jebsen was a Norwegian businessperson and politician .	(Peter Jebsen, was, A Norwegian Businessperson) → (peter_jebsen.Q1774966, country_of_citizenship.P27, norway.Q20); (Peter Jebsen, was, A Norwegian Businessperson) → (peter_jebsen.Q1774966, occupation.P106, businessperson.Q43845); (Peter Jebsen, was, Politician) → (peter_jebsen.Q1774966, occupation.P106, politician.Q82955)
Pina_Piovani	Pina Piovani Pina Piovani was an Italian stage and film actress .	(Pina Piovani, was, An Italian Stage And Film Actress) → (pina_piovani.Q3388583, country_of_citizenship.P27, italy.Q38)
Robert_G._Cole	Rapidly advancing through the ranks at Fort Benning as the parachute infantry battalions were expanded to regiments , he was a lieutenant colonel commanding the 3rd Battalion of the 502nd Parachute Infantry Regiment on June 6 , 1944 , the date of his unit ' s first combat jump .	(He, was, A Lieutenant Colonel) → (robert_g._cole.Q969944, military_rank.P410, lieutenant_colonel.Q493898)
Robert_Holt_Carpenter	Robert Holt Carpenter Robert Holt Carpenter was a New Zealand bookbinder , local politician , bookseller and character .	(Robert Holt Carpenter, was, A New Zealand Bookbinder) → (robert_holt_carpenter.Q7345593, country_of_citizenship.P27, new_zealand.Q664)
Robert_Van_Boxel	He joined amateur side FC Lisse in summer 2015 from Sparta .	(He, joined, Amateur Side Fc Lisse) → (robert_van_boxel.Q2017870, member_of_sports_team.P54, fc_lisse.Q978325)
Rolf_Maurer	Rolf Maurer Rolf Maurer was a Swiss road racing cyclist who competed professionally between 1960 and 1969 .	(Rolf Maurer, was, A Swiss Road Racing Cyclist) → (rolf_maurer.Q870132, country_of_citizenship.P27, switzerland.Q39); (Rolf Maurer, was, A Swiss Road Racing Cyclist) → (rolf_maurer.Q870132, sport.P641, road_bicycle_racing.Q3609)
Ron_Maudsley	Maudsley was born in Lostock Gralam , Cheshire , and educated at Malvern College and Birmingham University , then served in the Royal Army Service Corps during World War II .	(Maudsley, was born in, Lostock Gralam) → (ron_maudsley.Q16009653, place_of_birth.P19, lostock_gralam.Q4640202)
Safa_Al-Safi	Safa al - Safi Safa al - Din Mohammed al - Safi is an Iraqi politician and former Justice Minister who is currently Minister of State for the Council of Representatives .	(Mohammed Al-Safi, is, An Iraqi Politician) → (safa_al-safi.Q7398343, occupation.P106, politician.Q82955)
Salvador_Luria	In 1959 , he became chair of Microbiology at the Massachusetts Institute of Technology .	(He, became chair of, Microbiology) → (salvador_luria.Q205667, field_of_work.P101, microbiology.Q7193)
Sanele_Vavae_Tuilagi	He now plays for US Carcassonne .	(He, now plays, Us Carcassonne) → (sanele_vavae_tuilagi.Q3472139, member_of_sports_team.P54, us_carcassonne.Q3550526)
Seth_Plum	Seth Plum Seth Lewis Plum was an English international footballer who played as a wing half .	(Seth Lewis Plum, was, An English International Footballer) → (seth_plum.Q7456581, sport.P641, association_football.Q2736); (Seth Lewis Plum, was, An English International Footballer) → (seth_plum.Q7456581, member_of_sports_team.P54, england_national_football_team.Q47762)
Sofija_Korkutytė	Sofija Korkutytė Sofija Korkutytė was a Lithuanian rower who won three European titles in the eights event in 1963 , 1965 and 1967 ; she finished second in 1964 and 1966 .	(Sofija Korkutytė, was, A Lithuanian Rower) → (sofija_korkutytė.Q16014909, sport.P641, rowing_(sport).Q159354)
Thomas_Rogne	He made his first start for Celtic at home against Dundee United in a 1 – 0 victory on 20 February 2010 , he played very well and impressed the coach Tony Mowbray .	(He, at, Celtic) → (thomas_rogne.Q609801, member_of_sports_team.P54, celtic_f.c..Q19593)
Uri_Avnery	Avnery was born in Beckum , near Münster in Westphalia , as Helmut Ostermann , the youngest of four children , to a well - established German Jewish family , his father being a private banker in the town .	(Westphalia, in, Beckum) → (uri_avnery.Q325679, place_of_birth.P19, beckum,_germany.Q2707)
Ursula_Lehr	She was a member of the Christian Democratic Union .	(She, of, The Christian Democratic Union) → (ursula_lehr.Q290708, member_of_political_party.P102, christian_democratic_union_of_germany.Q49762)
Valdívia_(Brazilian_Footballer)	Valdívia Wanderson Ferreira de Oliveira , commonly known as Valdívia , is a Brazilian footballer who plays as an attacking midfielder for Avaí , on loan from Internacional .	(A Brazilian Footballer, who plays as, An Attacking Midfielder) → (valdívia_(brazilian_footballer).Q18176477, position_played_on_team_/_speciality.P413, midfielder.Q193592)
Victor-François,_2Nd_Duc_De_Broglie	His victory over Prince Ferdinand at Bergen won him the rank of marshal of France from the French King Louis XV and the title of Prince of the Holy Roman Empire " Reichsfürst " from Holy Roman Emperor Francis I . In 1759 , he won the Battle of Bergen and followed that with the capture the city of Minden , later fighting at the Battle of Minden under the command of Contades , whom he would succeed in command .	(His victory, won him the, Marshal) → (victor-françois,_2nd_duc_de_broglie.Q695929, award_received.P166, marshals_of_france.Q474485)

Figure 7: $\text{BERT}_{\text{LARGE}}$ on Wikidata-OIE.

ID	Sentence	Triples to gold triples
Appropriate_Rural_Technology_Institute	Appropriate Rural Technology Institute Appropriate Rural Technology Institute is an Indian non-governmental organization founded by a group of about 20 scientists and technologists to develop innovative and environmentally friendly rural technologies based on modern scientific knowledge in 1996.	(Appropriate Rural Technology Institute, is, An Indian Non-Governmental Organization) → (appropriate_rural_technology_institute.Q16245764, country.P17, india.Q668)
Ayatollah_(Record_Producer)	Ayatollah Lamont Dorrell, known as Ayatollah, is a hip-hop record producer from Queens, New York who has produced for predominantly New York-based rappers including Mos Def, Talib Kweli, R.A. The Rugged Man, Tragedy Khadafi, Wordsworth, Vast Aire, Afu-Ra, Guru, M.O.P., Inspectah Deck, Cormega, Ghostface Killah as well as many others.	(Ayatollah, is, A Hip-Hop Record Producer) → (ayatollah_(record_producer).Q4830984, occupation.P106, record_producer.Q183945)
Benjamin_Darbelet	Benjamin Darbelet Benjamin Darbelet is a French judoka.	(Benjamin Darbelet, is, A French Judoka) → (benjamin_darbelet.Q2456207, country_of_citizenship.P27, france.Q142); (Benjamin Darbelet, is, A French Judoka) → (benjamin_darbelet.Q2456207, sport.P641, judo.Q11420)
Bharat_Khawas	He plays for club Tribhuvan Army in Martyr's Memorial A-Division League as a striker.	(He, plays for, Club Tribhuvan Army) → (bharat_khawas.Q4901128, member_of_sports_team.P54, nepal_army_club.Q6994327); (He, plays for, Club Tribhuvan Army) → (bharat_khawas.Q4901128, sport.P641, association_football.Q2736)
Bill_White_(Shortstop)	White was born on May 1, 1860 in Bridgeport, Ohio, and he began his professional baseball career in 1883 with the Pottsville Antharcites of the International Association.	(White, in, Bridgeport) → (bill_white_(shortstop).Q4911357, place_of_birth.P19, bridgeport,_ohio.Q2062947)
Bob_Dylan.Q392	Dylan married Sara Lownds , who had worked as a model and a secretary to Drew Associates , on November 22 , 1965 .	(Dylan, married, Sara Lownds) → (Bob_Dylan.Q392, spouse.P26, Sara_Lownds.Q457433)
Bob_Dylan.Q392	In 1991 , Dylan received a Grammy Lifetime Achievement Award from American actor Jack Nicholson .	(Dylan, received, A Grammy Lifetime Achievement Award) → (Bob_Dylan.Q392, award_received.P166, Grammy_Lifetime_Achievement_Award.Q935843)
Boog_Powell	He also played for the Cleveland Indians and the Los Angeles Dodgers.	(He, also played for, The Cleveland Indians) → (boog_powell.Q4942810, member_of_sports_team.P54, cleveland_indians.Q642553)
Boog_Powell	He played in Major League Baseball as a first baseman and left fielder from through , most notably as a member of the Baltimore Orioles dynasty that won four American League pennants and two World Series championships between 1966 and 1971.	(He, played in, A First Baseman) → (boog_powell.Q4942810, position_played_on_team_/_speciality.P413, first_baseman.Q1326154)
Bouchraya_Hammoudi_Bayoun	He studied Economics at the University of Havana, Cuba.	(He, studied, University of Havana) → (bouchraya_hammoudi_bayoun.Q4949324, educated_at.P69, university_of_havana.Q837320)
Cezaro_Rossetti	Of Italian-Swiss derivation, he was born in Glasgow and lived in Britain.	(He, was born in, Glasgow) → (cezaro_rossetti.Q2947206, place_of_birth.P19, glasgow.Q4093)
Chan_Sau_Ying	Chan Sau Ying Chan Sau Ying is a retired athlete from Hong Kong who specialised in the 100 metres hurdles.	(Sau Ying, is, A Retired Athlete) → (chan_sau_ying.Q21858345, sport.P641, sport_of_athletics.Q542)
Christian_Mcgrath	He was born in the Bronx in 1972 and became interested in art and illustration as a small child.	(He, was born in, The Bronx) → (christian_mcgrath.Q5109897, place_of_birth.P19, the_bronx.Q18426)
Cliff_Michelmore	He was a regular presenter on BBC1's "Holiday" programme from 1969 to 1986, and presented other shows for BBC TV, ITV and BBC Radio.	(He, was, A Regular Presenter) → (cliff_michelmore.Q5132704, occupation.P106, television_presenter.Q947873)
Crooked_I	Crooked I Dominick Antron Wickliffe , better known by his stage name Crooked I or Kxng Crooked, is an American rapper from Long Beach, California.	(Crooked I, is, An American Rapper) → (crooked_i.Q253695, country_of_citizenship.P27, united_states.Q30)
Denise_Johns	She graduated in 2002, and moved to Atlanta, playing amateur beach volleyball on the east coast of America.	(She, playing, Amateur Beach Volleyball) → (denise_johns.Q5257690, sport.P641, beach_volleyball.Q4543)
Diana_M._Fennell	Fennell was born in Emporia, Virginia and attended the public schools there.	(Fennell, was born in, Emporia) → (diana_m._fennell.Q28037279, place_of_birth.P19, emporia,_virginia.Q177678)
Diana_Rast	She also competed at the 1996 Summer Olympics.	(She, also competed at, The 1996 Summer Olympics) → (diana_rast.Q2491787, participant_of.P1344, 1996_summer_olympics.Q8531)
Eduard_Friedrich_Poeppig	Eduard Friedrich Poeppig Eduard Friedrich Poeppig was a German botanist, zoologist and explorer.	(Eduard Friedrich, was, A German Botanist) → (eduard_friedrich_poeppig.Q61501, field_of_work.P101, botany.Q441)
Edward_L._Romero	After the Vietnam war, he became a county chairman of Democratic Party.	(He, of, Democratic Party) → (edward_l._romero.Q5344025, member_of_political_party.P102, democratic_party_(united_states).Q29552)
Endre_Tilli	Endre Tilli Endre Tilli was a Hungarian fencer.	(Endre Tilli, was, A Hungarian Fencer) → (endre_tilli.Q1642140, country_of_citizenship.P27, hungary.Q28); (Endre Tilli, was, A Hungarian Fencer) → (endre_tilli.Q1642140, sport.P641, fencing.Q12100)
Enrico_Porro	Porro competed at the 1908 Summer Olympics in London where he won the gold medal in Greco-Roman wrestling, the lightweight class.	(Porro, competed at, The 1908 Summer Olympics) → (enrico_porro.Q1343905, participant_of.P1344, 1908_summer_olympics.Q8111)
Erfan_Zeneli	Erfan Zeneli Erfan Zeneli is a Finnish footballer of Kosovar Albanian descent who currently plays for HJK Helsinki as a striker or left winger.	(Erfan Zeneli, is, A Finnish Footballer) → (erfan_zeneli.Q2053367, country_of_citizenship.P27, finland.Q33); (Erfan Zeneli, is, A Finnish Footballer) → (erfan_zeneli.Q2053367, sport.P641, association_football.Q2736); (Erfan Zeneli, is, A Finnish Footballer) → (erfan_zeneli.Q2053367, country_for_sport.P1532, finland.Q33)
Erminio_Favalli	He then moved to Mantova before to join Palermo, where he spent seven seasons as a "rosanero" mainstay.	(He, then moved to, Mantova) → (erminio_favalli.Q1405871, member_of_sports_team.P54, mantova_1911_s.s.d..Q430993)
George_Dowty	Dowty was born in Pershore, Worcestershire in 1901.	(Dowty, was born in, Pershore) → (george_dowty.Q1507239, place_of_birth.P19, pershore.Q767858)
Harald_Fairhair	In the Saga of Harald Fairhair in "Heimskringla", which is the most elaborate although not the oldest or most reliable source to the life of Harald, it is written that Harald succeeded, on the death of his father Halfdan the Black Gudrödarson, to the sovereignty of several small, and somewhat scattered kingdoms in Vestfold, which had come into his father's hands through conquest and inheritance.	(Harald, of his father, Halfdan the Black) → (harald_fairhair.Q206123, father.P22, halfdan_the_black.Q504932)
Hector_Dyer	Born in Los Angeles, California, Hector Dyer enrolled at the Stanford University and won the IC4A championships in in 1930.	(Hector Dyer, at, The Stanford University) → (hector_dyer.Q1364518, educated_at.P69, stanford_university.Q41506)
Henny_Magnussen	Thanks to the support she received, she was able to practise in all the courts of Denmark from 1909, becoming the first woman to do so.	(She, , she, Denmark) → (henny_magnussen.Q56577333, country_of_citizenship.P27, denmark.Q35)
Hina_Pervaiz_Butt	Hina Pervaiz Butt Hina Pervaiz Butt is a Pakistani politician who was a Member of the Provincial Assembly of the Punjab, since May 2013.	(Hina Pervaiz Butt, is, A Pakistani Politician) → (hina_pervaiz_butt.Q47918210, occupation.P106, politician.Q82955)
Isabella_Of_Valois,_Duchess_Of_Bourbon	She was the wife of Peter I, Duke of Bourbon.	(She, was the wife of, Peter I) → (isabella_of_valois,_duchess_of_bourbon.Q2363495, spouse.P26, peter_i,_duke_of_bourbon.Q444668)
Jahir_Butrón	Then from 1995, he played for Peruvian Second Division side Guardia Republicana.	(He, played for, Peruvian Second Division Side Guardia Republicana) → (jahir_butrón.Q6123259, member_of_sports_team.P54, guardia_republicana.Q5614012)
Jahir_Butrón	Jahir Butrón Jahir Butrón Gotuzzo is a Peruvian footballer who plays as a center back.	(Jahir Butrón Gotuzzo, is, A Peruvian Footballer) → (jahir_butrón.Q6123259, country_of_citizenship.P27, peru.Q419); (Jahir Butrón Gotuzzo, is, A Peruvian Footballer) → (jahir_butrón.Q6123259, sport.P641, association_football.Q2736)
Janina_Altman	In 1950, Hescheles emigrated to Israel, where she eventually earned a doctorate in chemistry at Technion .	(She, earned a doctorate in, Technion) → (janina_altman.Q22574794, educated_at.P69, technion_–_israel_institute_of_technology.Q333705)
Joan_Of_France,_Duchess_Of_Berry	She was the second daughter of King Louis XI of France and of his second wife Charlotte of Savoy; her surviving siblings were King Charles VIII of France and Anne of France.	(She, was the second daughter of, King Louis Xi) → (joan_of_france,_duchess_of_berry.Q236220, father.P22, louis_xi_of_france.Q8058)
John_Ogilvie_(Footballer)	He played for Hibs in their most successful era renowned for their front five known as The Famous Five.	(He, played for, Hibs) → (john_ogilvie_(footballer).Q6251147, member_of_sports_team.P54, hibernian_f.c..Q192597)
Johnedel_Cardel	He moved to Sta.	(He, moved to, Sta) → (johnedel_cardel.Q56599981, member_of_sports_team.P54, sta._lucia_realtors.Q3547751)
Joseph_W._Nega	Joseph W. Nega Joseph W. Nega is a Judge of the United States Tax Court.	(Joseph W. Nega, is, A Judge) → (joseph_w._nega.Q16733242, occupation.P106, judge.Q16533)
José_Alejandro_Aguilar_López	José Alejandro Aguilar López José Alejandro Aguilar López is a Mexican politician from the National Action Party.	(José Alejandro Aguilar López, is a Mexican politician, The National Action Party) → (josé_alejandro_aguilar_lópez.Q18385891, member_of_political_party.P102, national_action_party_(mexico).Q851087); (José Alejandro Aguilar López, is, A Mexican Politician) → (josé_alejandro_aguilar_lópez.Q18385891, country_of_citizenship.P27, mexico.Q96); (José Alejandro Aguilar López, is, A Mexican Politician) → (josé_alejandro_aguilar_lópez.Q18385891, occupation.P106, politician.Q82955)
Juan_Bautista_Azopardo	Juan Bautista Azopardo was born in Senglea, Malta, the son of Rosina and Salvatore Azopardo.	(Juan Bautista Azopardo, was born in, Senglea) → (juan_bautista_azopardo.Q3805864, place_of_birth.P19, senglea.Q846593); (Juan Bautista Azopardo, was born in, Malta) → (juan_bautista_azopardo.Q3805864, country_of_citizenship.P27, malta.Q233)
Kevin_Kerr_(Scottish_Footballer)	He signed for AGOVV Apeldoorn of Netherlands' Eerste Divisie in January 2012 after his contract in Bielefeld had expired the previous summer.	(He, signed for, Agovv Apeldoorn) → (kevin_kerr_(scottish_footballer).Q326761, member_of_sports_team.P54, agovv_apeldoorn.Q292618)
Lachezar_Kotev	Lachezar Kotev Lachezar Kotev is a Bulgarian footballer who plays as a midfielder for Vitosha Bistritsa.	(Lachezar Kotev, is, A Bulgarian Footballer) → (lachezar_kotev.Q49560773, sport.P641, association_football.Q2736)
Marc_Zeno	In 1988, Zeno was drafted in the 7th round by the Pittsburgh Steelers.	(Zeno, by, The Pittsburgh Steelers) → (marc_zeno.Q16196353, member_of_sports_team.P54, pittsburgh_steelers.Q191477)
Marcel_Dheere	Marcel Dheere Marcel Albert "Ching" Dheere was a Canadian professional ice hockey forward who played 11 games in the National Hockey League for the Montreal Canadiens.	(Marcel Albert "Ching, was, A Canadian Professional Ice Hockey) → (marcel_dheere.Q6756083, country_of_citizenship.P27, canada.Q16); (Marcel Albert "Ching, was, A Canadian Professional Ice Hockey) → (marcel_dheere.Q6756083, sport.P641, ice_hockey.Q41466); (Marcel Albert "Ching, for, The Montreal Canadiens) → (marcel_dheere.Q6756083, member_of_sports_team.P54, montreal_canadiens.Q188143)
Mark_Schultz_(Wrestler)	Mark Schultz was born in 1960 in Palo Alto, California to Dorothy Jean St. Germain and Philip Gary Schultz.	(Mark Schultz, was born in 1960 in, Palo Alto) → (mark_schultz_(wrestler).Q3487332, place_of_birth.P19, palo_alto,_california.Q47265)
Michael_Jones_(Footballer)	Michael Jones Michael Jones is an English footballer, who plays for Cefn Druids.	(Michael Jones, is, An English Footballer) → (michael_jones_(footballer).Q6831707, sport.P641, association_football.Q2736)
Olga_Semenova_Tyan-Shanskaya	She learned to play chess at the age of 18; four years later she won the Leningrad city women's chess championship.	(She, learned, Chess) → (olga_semenova_tyan-shanskaya.Q2392240, sport.P641, chess.Q718)
Owen_Paterson	Owen Paterson Owen William Paterson is a British Conservative Party politician who was the Secretary of State for Environment, Food and Rural Affairs from 2012 to 2014.	(Owen William Paterson, is, A British Conservative Party Politician) → (owen_paterson.Q197894, occupation.P106, politician.Q82955)
Pan_Gongsheng	Pan Gongsheng Pan Gongsheng is a Chinese economist, banker, reformist and bureaucrat.	(Pan Gongsheng, is, A Chinese Economist) → (pan_gongsheng.Q9309080, country_of_citizenship.P27, china.Q148)
Paul Carrington (American Football)	Paul Carrington Paul Carrington is a former American football defensive end.	(Paul Carrington, is, A Former American Football Defensive End) → (paul_carrington_(american_football).Q9057011, sport.P641, american_football.Q41323)
Peter_Jebsen	Peter Jebsen Peter Jebsen was a Norwegian businessperson and politician.	(Peter Jebsen, was, A Norwegian Businessperson) → (peter_jebsen.Q1774966, country_of_citizenship.P27, norway.Q20); (Peter Jebsen, was, A Norwegian Businessperson) → (peter_jebsen.Q1774966, occupation.P106, businessperson.Q43845)
Pina_Piovani	Pina Piovani Pina Piovani was an Italian stage and film actress.	(Pina Piovani, was, An Italian Stage And Film Actress) → (pina_piovani.Q3388583, country_of_citizenship.P27, italy.Q38)
Robert_G._Cole	Rapidly advancing through the ranks at Fort Benning as the parachute infantry battalions were expanded to regiments, he was a lieutenant colonel commanding the 3rd Battalion of the 502nd Parachute Infantry Regiment on June 6, 1944, the date of his unit's first combat jump.	(He, was, A Lieutenant Colonel) → (robert_g._cole.Q969944, military_rank.P410, lieutenant_colonel.Q493898)
Robert_Holt_Carpenter	Robert Holt Carpenter Robert Holt Carpenter was a New Zealand bookbinder, local politician, bookseller and character.	(Robert Holt Carpenter, was, A New Zealand Bookbinder) → (robert_holt_carpenter.Q7345593, country_of_citizenship.P27, new_zealand.Q664)
Robert_Van_Boxel	He joined amateur side FC Lisse in summer 2015 from Sparta.	(He, joined, Amateur Side Fc Lisse) → (robert_van_boxel.Q2017870, member_of_sports_team.P54, fc_lisse.Q978325)
Rolf_Maurer	Rolf Maurer Rolf Maurer was a Swiss road racing cyclist who competed professionally between 1960 and 1969.	(Rolf Maurer, was, A Swiss Road Racing Cyclist) → (rolf_maurer.Q870132, country_of_citizenship.P27, switzerland.Q39); (Rolf Maurer, was, A Swiss Road Racing Cyclist) → (rolf_maurer.Q870132, sport.P641, road_bicycle_racing.Q3609)
Ron_Maudsley	Maudsley was born in Lostock Gralam, Cheshire, and educated at Malvern College and Birmingham University, then served in the Royal Army Service Corps during World War II.	(Maudsley, was born in, Lostock Gralam) → (ron_maudsley.Q16009653, place_of_birth.P19, lostock_gralam.Q4640202)
Safa_Al-Safi	Safa al-Safi Safa al-Din Mohammed al-Safi is an Iraqi politician and former Justice Minister who is currently Minister of State for the Council of Representatives.	(Mohammed Al-Safi, is, An Iraqi Politician) → (safa_al-safi.Q7398343, occupation.P106, politician.Q82955)
Salvador_Luria	In 1959, he became chair of Microbiology at the Massachusetts Institute of Technology .	(He, became chair of, Microbiology) → (salvador_luria.Q205667, field_of_work.P101, microbiology.Q7193)
Sanele_Vavae_Tuilagi	He now plays for US Carcassonne.	(He, now plays for, Us Carcassonne) → (sanele_vavae_tuilagi.Q3472139, member_of_sports_team.P54, us_carcassonne.Q3550526)
Seth_Plum	Seth Plum Seth Lewis Plum was an English international footballer who played as a wing half.	(Seth Lewis Plum, was, An English International Footballer) → (seth_plum.Q7456581, sport.P641, association_football.Q2736); (Seth Lewis Plum, was, An English International Footballer) → (seth_plum.Q7456581, member_of_sports_team.P54, england_national_football_team.Q47762)
Sofija_Korkutytė	Sofija Korkutytė Sofija Korkutytė was a Lithuanian rower who won three European titles in the eights event in 1963, 1965 and 1967; she finished second in 1964 and 1966.	(Sofija Korkutytė, was, A Lithuanian Rower) → (sofija_korkutytė.Q16014909, sport.P641, rowing_(sport).Q159354)
Thomas_Rogne	He made his first start for Celtic at home against Dundee United in a 1–0 victory on 20 February 2010, he played very well and impressed the coach Tony Mowbray.	(He, made his first start for, Celtic) → (thomas_rogne.Q609801, member_of_sports_team.P54, celtic_f.c..Q19593)
Uri_Avnery	Avnery was born in Beckum, near Münster in Westphalia, as Helmut Ostermann, the youngest of four children, to a well-established German Jewish family, his father being a private banker in the town.	(Avnery, born in, Beckum) → (uri_avnery.Q325679, place_of_birth.P19, beckum,_germany.Q2707)
Ursula_Lehr	She was a member of the Christian Democratic Union .	(She, was a member of, The Christian Democratic Union) → (ursula_lehr.Q290708, member_of_political_party.P102, christian_democratic_union_of_germany.Q49762)
Valdívia_(Brazilian_Footballer)	Valdívia Wanderson Ferreira de Oliveira , commonly known as Valdívia, is a Brazilian footballer who plays as an attacking midfielder for Avaí, on loan from Internacional.	(A Brazilian Footballer, plays as, An Attacking Midfielder) → (valdívia_(brazilian_footballer).Q18176477, position_played_on_team_/_speciality.P413, midfielder.Q193592)
Victor-François,_2Nd_Duc_De_Broglie	His victory over Prince Ferdinand at Bergen won him the rank of marshal of France from the French King Louis XV and the title of Prince of the Holy Roman Empire "Reichsfürst" from Holy Roman Emperor Francis I. In 1759, he won the Battle of Bergen and followed that with the capture the city of Minden, later fighting at the Battle of Minden under the command of Contades, whom he would succeed in command.	(His victory, won him the, Marshal) → (victor-françois,_2nd_duc_de_broglie.Q695929, award_received.P166, marshals_of_france.Q474485)

Figure 8: GPT-2_{XL} on Wikidata-OIE.

ID	Sentence	Triples to gold triples	
		BERT$_{\text{LARGE}}$	GPT-2$_{\text{XL}}$
Andy_Hilbert	In the same season he was claimed off waivers by the Pittsburgh Penguins on March 9, 2006.	-	(He, by, The Pittsburgh Penguins) → (andy_hilbert.Q526158, member_of_sports_team.P54, pittsburgh_penguins.Q193643)
Bob_Dylan.Q392	Dylan married his backup singer Carolyn Dennis on June 4, 1986.	-	(Dylan, married, Carolyn Dennis) → (Bob_Dylan.Q392, spouse.P26, Carolyn_Dennis.Q5045345)
Bones_Hillman	During the hiatus of Midnight Oil , Hillman returned to New Zealand , working as a studio and live musician with Dave Dobbyn and recorded the album " Available Light " .	(Hillman, returned, New Zealand) → (bones_hillman.Q2910026, country_of_citizenship.P27, new_zealand.Q664)	-
Brent_Hinds	Hinds continues to concentrate on Mastodon, with the majority of his time spent touring or in the studio.	-	(His Time, with, Mastodon) → (brent_hinds.Q909746, member_of.P463, mastodon_(band).Q548844)
Butch_Vig	Their debut album, "Garbage", was an unexpected smash, selling over 4 million copies and certified double platinum in the UK, United States and Australia.	-	(Over 4 Million Copies, in, United States) → (butch_vig.Q451084, country_of_citizenship.P27, united_states.Q30)
C-Kan	In addition to being known as a rapper, C-Kan is also known for supporting the free use of marijuana in México.	-	(Marijuana, in, México) → (c-kan.Q27734073, country_of_citizenship.P27, mexico.Q96)
Charles_Gratiot_Sr.	Gratiot died of a stroke in St . Louis .	(Gratiot, died, St. Louis) → (charles_gratiot_sr..Q2959257, place_of_death.P20, st._louis.Q38022)	-
Cheryl_Chan	Cheryl Chan Cheryl Chan Wei Ling , PBM , is a Singaporean politician .	(Cheryl Chan Wei Ling, is, A Singaporean Politician) → (cheryl_chan.Q22003605, occupation.P106, politician.Q82955)	-
Cyba_Audi	Audi is an experienced moderator and presenter at local, regional, and international business events.	-	(Audi, is, Presenter) → (cyba_audi.Q18921358, occupation.P106, television_presenter.Q947873)
Douglas_Lenat	He is a Fellow of the AAAS , AAAI , and Cognitive Science Society , and an editor of the J . Automated Reasoning , J . Learning Sciences , and J . Applied Ontology .	(He, is a Fellow, Aaai) → (douglas_lenat.Q559334, member_of.P463, association_for_the_advancement_of_artificial_intelligence.Q2739680)	-
Douglas_Lenat	While attending the University of Pennsylvania, Lenat supported himself through programming, notably designing and developing a natural language interface to a U.S. Navy data base question–answering system serving as an early online shipboard operations manual used on US aircraft carriers.	-	(Lenat, „ The University of Pennsylvania) → (douglas_lenat.Q559334, educated_at.P69, university_of_pennsylvania.Q49117)
Fred_Stewart_(Football_Manager)	Despite managing for 39 years he only ever took charge of two clubs, Stockport County and Cardiff City, and he holds the record for longest serving manager in the history of both clubs.	-	(He, charge of, Stockport County) → (fred_stewart_(football_manager).Q5496345, member_of_sports_team.P54, stockport_county_f.c..Q18526)
Félix_Tanco	Born Félix Manuel de Jesús Tanco y Bosmeniel , in Bogotá , Colombia , he arrived in Cuba at a very young age .	(Bogotá, „ Cuba) → (félix_tanco.Q5511649, country_of_citizenship.P27, cuba.Q241)	-
Gary_Ablett_Jr.	He also won his second Brownlow Medal , becoming the first Gold Coast player to win the award and the 14th player in VFL / AFL history to win it twice .	(He, also won, His Second Brownlow Medal) → (gary_ablett_jr..Q3098509, award_received.P166, leigh_matthews_trophy.Q6519632); (He, also won, His Second Brownlow Medal) → (gary_ablett_jr..Q3098509, award_received.P166, brownlow_medal.Q3853498)	-
Geeto_Mongol	After meeting his would - be tag team partner , Bepo Mongol , Tattrie returned to the WWWF as Geto Mongol and The Mongols were brought to the United States in 1968 .	(The Mongols, were, The United States) → (geeto_mongol.Q11324907, country_of_citizenship.P27, united_states.Q30)	-
Georgi_Plekhanov	Georgi Plekhanov Georgi Valentinovich Plekhanov was a Russian revolutionary , philosopher and a Marxist theoretician .	(Georgi Valentinovich Plekhanov, was, A Marxist Theoretician) → (georgi_plekhanov.Q182905, occupation.P106, philosopher.Q4964182)	-
Győző_Drozdy	Győző Drozdy Győző Drozdy was a Hungarian teacher , journalist , and politician .	(Győző Drozdy, was, Politician) → (győző_drozdy.Q910001, occupation.P106, politician.Q82955)	-
Horace_Rawlins	Rawlins died on 22 January 1935 in a nursing home in Reading , Berkshire .	(Rawlins, died, Reading) → (horace_rawlins.Q4436711, place_of_death.P20, reading,_berkshire.Q161491)	-
Horace_Rawlins	He played in a professional event at Stanmore Golf Club in June 1894 but finished well out of the prizes.	-	(He, played in, Stanmore Golf Club) → (horace_rawlins.Q4436711, sport.P641, golf.Q5377)
Hugo_Winterhalter	Winterhalter remained with RCA Victor until 1963, at which time he moved to Kapp; that same year, he also penned the main title theme for the film, "Diamond Head".	-	(Winterhalter, to, Kapp) → (hugo_winterhalter.Q1635872, record_label.P264, kapp_records.Q1728690)
Ioannis_Talianos	He competed in the men ' s 4 × 100 metres relay at the 1924 Summer Olympics .	(He, competed in the, The 1924 Summer Olympics) → (ioannis_talianos.Q35692847, participant_of.P1344, 1924_summer_olympics.Q8132)	-
Joan_Of_France,_Duchess_Of_Berry	The nuns of the Order of the Annunciation of the Blessed Virgin Mary still maintain their way of life in monasteries in France, Belgium, Costa Rica and Poland.	-	(Their Way, in, France) → (joan_of_france,_duchess_of_berry.Q236220, country_of_citizenship.P27, france.Q142)
John_Maynard_Keynes	John Maynard Keynes John Maynard Keynes , 1st Baron Keynes , was a British economist , whose ideas fundamentally changed the theory and practice of macroeconomics and the economic policies of governments .	(John Maynard Keynes, was, A British Economist) → (john_maynard_keynes.Q9317, occupation.P106, economist.Q188094)	-
Jonathan_Culler	Currently , he is Class of 1916 Professor of English and Comparative Literature at Cornell University .	(He, is, English) → (jonathan_culler.Q933332, languages_spoken,_written_or_signed.P1412, english_language.Q1860)	-
Jordi_Cañas_Pérez	In the regional elections of Catalonia in November 2010 , he was chosen in the primary as number three for Ciutadans ' candidacy in the province of Barcelona .	(The Province, of, Barcelona) → (jordi_cañas_pérez.Q557693, member_of_political_party.P102, citizens_(spanish_political_party).Q1393123)	-
Jules_Deligny	He competed in the freestyle lightweight event at the 1920 Summer Olympics .	(He, competed in the, The 1920 Summer Olympics) → (jules_deligny.Q16207961, participant_of.P1344, 1920_summer_olympics.Q8128)	-
Liz_Mcgregor	Liz McGregor Liz McGregor is a South African author and a journalist who worked for leading South African newspapers such as the " Sunday Times " and the " Rand Daily Mail " .	(Liz Mcgregor, is, A Journalist) → (liz_mcgregor.Q23541069, occupation.P106, journalist.Q1930187)	-
Louise_Shropshire	The granddaughter of slaves, Louise Shropshire was born Louise Jarrett on February 15, 1913 in Coffee County, Alabama.	-	(Louise Shropshire, was, Coffee County, Alabama) → (louise_shropshire.Q15430857, place_of_birth.P19, coffee_county,_alabama.Q485660)
Macit_Gürdal	He competed in the men ' s tournament at the 1952 Summer Olympics .	(He, competed in the, The 1952 Summer Olympics) → (macit_gürdal.Q57313835, participant_of.P1344, 1952_summer_olympics.Q8407)	-
Markus_Pröll	At the beginning of his career , Pröll became a target of some ridicule because his name " Pröll " literally means something like " lout " in German .	("Pröll, " in, German) → (markus_pröll.Q704630, country_of_citizenship.P27, germany.Q183)	-
Monde_Hadebe	Their Round Four match against the saw Hadebe score the first senior try of his career in a 27 – 10 victory , and the Sharks XV again finished top of the Southern Section log .	(His Career, „ The Sharks Xv) → (monde_hadebe.Q6898677, member_of_sports_team.P54, sharks_(currie_cup).Q744636)	-
Neville_Southall	He moved on to Everton for £150,000 in 1981 and established himself as the club's first-choice goalkeeper by the 1983–84 season.	-	(He, moved on to, Everton) → (neville_southall.Q436650, member_of_sports_team.P54, everton_f.c..Q5794)
Olga_Semenova_Tyan-Shanskaya	She died in Leningrad .	(She, died, Leningrad) → (olga_semenova_tyan-shanskaya.Q2392240, place_of_death.P20, saint_petersburg.Q656)	-
Parvathy_Ratheesh	Parvathy Ratheesh made her film acting debut in 2015 similar to that of her younger brother , Padmaraj Ratheesh who also made his acting debut in the same year through " Fireman " .	(Her Film, that of her younger brother, Padmaraj Ratheesh) → (parvathy_ratheesh.Q19895785, sibling.P3373, padmaraj_ratheesh.Q19895782)	-
Peter_Dalla_Riva	In 1993, he was inducted into the Canadian Football Hall of Fame.	-	(He, was inducted into the, Fame) → (peter_dalla_riva.Q7173528, award_received.P166, canadian_football_hall_of_fame.Q3517653)
Pierre_Duhem	Unlike many former historians , who denigrated the Middle Ages, he endeavored to show that the Roman Catholic Church had helped foster Western science in one of its most fruitful periods.	-	(The Roman Catholic Church, had, Foster Western Science) → (pierre_duhem.Q314172, field_of_work.P101, philosophy_of_science.Q59115)
Rafael_Arcadio_Bernal_Supelano	He then served as bishop of the Roman Catholic Diocese of Arauca, Colombia, from 1990 to 2003 and as bishop of the Roman Catholic Diocese of Libano–Honda, Colombia, from 2003 to 2004.	-	(The Roman Catholic Diocese, „ Colombia) → (rafael_arcadio_bernal_supelano.Q2126904, country_of_citizenship.P27, colombia.Q739)
Rajeshwar_Dayal	The Government of India awarded him the second highest civilian award of the Padma Vibhushan, in 1969.	-	(Him, award, The Padma Vibhushan) → (rajeshwar_dayal.Q15448539, award_received.P166, padma_vibhushan.Q672392)
Ray_Johnson_(American_Football)	Ray Johnson Raymond Robert Johnson was an American football defensive back who played three seasons in the National Football League with the Cleveland Rams and Chicago Cardinals .	(The National Football League, back who played, Raymond Robert Johnson) → (ray_johnson_(american_football).Q21077889, sport.P641, american_football.Q41323)	-
Sean_Hanish	Sean Hanish Sean Hanish is an American film writer , producer and director best known for " Saint Judy " , " Return to Zero " and " Sister Cities " .	(Sean Hanish, is, An American Film Writer) → (sean_hanish.Q19867622, country_of_citizenship.P27, united_states.Q30); (Sean Hanish, is, An American Film Writer) → (sean_hanish.Q19867622, occupation.P106, screenwriter.Q28389)	-
Seth_Plum	Born in Edmonton, Plum played professionally for Charlton Athletic, Chelsea and Southend United.	-	(Plum, played professionally for, Charlton Athletic) → (seth_plum.Q7456581, member_of_sports_team.P54, charlton_athletic_f.c..Q19462)
Seth_Plum	Plum received his only cap for England at age 23 while playing for Charlton Athletic , starting and playing the full 90 minutes in a 4 – 1 win over France on 10 May 1923 .	(His Only Cap, playing for, Charlton Athletic) → (seth_plum.Q7456581, member_of_sports_team.P54, charlton_athletic_f.c..Q19462)	-
Stanislav_Nedkov	A two - time national champion in freestyle wrestling , he has also had first - place finishes in international tournaments in Russia , Moldova , Turkey , and Bulgaria .	(Turkey, , and, Bulgaria) → (stanislav_nedkov.Q4844375, country_of_citizenship.P27, bulgaria.Q219)	-
Stiliyan_Petrov	In addition he is Bulgaria's all-time most-capped player with 105 appearances for the side.	-	(He, is, Bulgaria'S All-Time Most-Capped Player) → (stiliyan_petrov.Q166263, country_of_citizenship.P27, bulgaria.Q219)
Stiliyan_Petrov	On 14 January 2010 , it was announced that Petrov had come second in Bulgaria ' s Player of the Year .	(Petrov, in, Bulgaria'S Player) → (stiliyan_petrov.Q166263, country_of_citizenship.P27, bulgaria.Q219)	-
Thomas_A._Dunn	He graduated from DePaul University College of Law in 1971.	-	(He, graduated from, Depaul University College) → (thomas_a._dunn.Q33082990, educated_at.P69, depaul_university_college_of_law.Q5244034)
Thomas_Scott_(Commentator)	In 1803 , Scott left the Lock Hospital to become Rector of Aston Sandford in Buckinghamshire where he remained until his death in 1821 .	(His Death, he remained until, Aston Sandford) → (thomas_scott_(commentator).Q7793831, place_of_death.P20, aston_sandford.Q3088093)	-
Thor_Heyerdahl	Thor Heyerdahl Thor Heyerdahl was a Norwegian adventurer and ethnographer with a background in zoology, botany and geography.	-	(Thor Heyerdahl, was, A Norwegian Adventurer) → (thor_heyerdahl.Q133622, country_of_citizenship.P27, norway.Q20)
Venkatesh_Kulkarni	His first novel , " Naked in Deccan " , won the 1984 American Book Award of the Before Columbus Foundation and was listed among the top ten novels of the decade by the " Chicago Tribune " .	(His First Novel, , won, The 1984 American Book Award) → (venkatesh_kulkarni.Q7920091, award_received.P166, american_book_awards.Q463606)	-
Vitor_Castro_De_Souza	Vitor Castro de Souza Vítor Castro de Souza , or simply Vitor Castro, is a Brazilian striker.	-	(Vítor Castro De Souza, is, A Brazilian Striker) → (vitor_castro_de_souza.Q5604876, position_played_on_team_/_speciality.P413, forward_(association_football).Q280658)
Vyacheslav_Tsaryov	He made his debut in the Soviet Top League in 1990 for FC Dynamo Moscow .	(He, for, Fc Dynamo Moscow) → (vyacheslav_tsaryov.Q4503307, member_of_sports_team.P54, fc_dynamo_moscow.Q17497)	-
Zare_Markovski	Zare Markovski as player played in KK Rabotnički and MZT Skopje .	(Zare Markovski, played in, Kk Rabotnički) → (zare_markovski.Q4023877, member_of_sports_team.P54, kk_rabotnički.Q2603345)	-

Figure 9: BERT$_{\text{LARGE}}$ vs. GPT-2$_{\text{XL}}$ on Wikidata-OIE.