


Retrieving Complex Tables with Multi-Granular Graph Representation Learning

Fei Wang, Kexuan Sun, Muhao Chen, Jay Pujara, Pedro Szekely

Department of Computer Science & Information Sciences Institute, University of Southern California
Los Angeles, California, USA

{fwang598,kexuansu,muhaoche,jpujara,szekely}@usc.edu

ABSTRACT

The task of natural language table retrieval (NLTR) seeks to retrieve semantically relevant tables based on natural language queries. Existing learning systems for this task often treat tables as plain text based on the assumption that tables are structured as dataframes. However, tables can have complex layouts which indicate diverse dependencies between subtable structures, such as nested headers. As a result, queries may refer to different spans of relevant content that is distributed across these structures. Moreover, such systems fail to generalize to novel scenarios beyond those seen in the training set. Prior methods are still distant from a generalizable solution to the NLTR problem, as they fall short in handling complex table layouts or queries over multiple granularities. To address these issues, we propose Graph-based Table Retrieval (GTR ) , a generalizable NLTR framework with multi-granular graph representation learning. In our framework, a table is first converted into a *tabular graph*, with cell nodes, row nodes and column nodes to capture content at different granularities. Then the tabular graph is input to a Graph Transformer model that can capture both table cell content and the layout structures. To enhance the robustness and generalizability of the model, we further incorporate a self-supervised pre-training task based on graph-context matching. Experimental results on two benchmarks show that our method leads to significant improvements over the current state-of-the-art systems. Further experiments demonstrate promising performance of our method on cross-dataset generalization, and enhanced capability of handling complex tables and fulfilling diverse query intents.¹

CCS CONCEPTS

• **Information systems** → **Retrieval models and ranking.**

KEYWORDS

Table retrieval; Semantic retrieval; Graph Transformer; Pre-training

ACM Reference Format:

Fei Wang, Kexuan Sun, Muhao Chen, Jay Pujara, Pedro Szekely. 2021. Retrieving Complex Tables with Multi-Granular Graph Representation Learning. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21)*, July

¹Code and data are available at <https://github.com/FeiWang96/GTR>



This work is licensed under a Creative Commons Attribution NonCommercial-ShareAlike International 4.0 License.

SIGIR '21, July 11–15, 2021, Virtual Event, Canada

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8037-9/21/07.

<https://doi.org/10.1145/3404835.3462909>

11–15, 2021, Virtual Event, Canada. ACM, New York, NY, USA, 11 pages.
<https://doi.org/10.1145/3404835.3462909>

1 INTRODUCTION

Web tables are rich sources of semi-structured knowledge that benefit a wide range of applications. For example, Wikipedia contains millions of high-quality tables that support various knowledge-driven tasks, such as table-based question answering [41], fact verification [12] and table-to-text generation [35]. Additionally, tables are ubiquitous in scientific literature and financial reports and have inspired research efforts on table profiling [37], table-text grounding [30], tabular semantic parsing [19], etc.

As an important component of Web information, tables are presented as direct results to Web queries in search engines [3]. Traditional formal methods for information retrieval from databases (e.g. SQL, QBE [70] and QUEL [51]) and formal query generation methods (e.g. text-to-SQL [61, 63]) do not provide a flexible way to support information retrieval from Web tables with varied structures. Therefore, the task of *natural language table retrieval* (NLTR) [66] has been proposed, offering more flexibility for directly searching semantically relevant tables based on search queries described in natural language. Moreover, NLTR is a key building block for tasks that require synthesizing knowledge from tables, such as table-based reading comprehension [59], open fact verification [46], and open domain question answering [10, 13]. While Web tables are beneficial to many downstream tasks, a key issue in those tasks in real-world scenarios lies in the difficulty of efficiently collecting relevant tables from a large table corpus. NLTR can then be used to identify candidate tables for those tasks.

A common challenge of NLTR lies in the fact that tables consist of both structured table cells and unstructured contextual information (e.g. captions). A simple method to deal with both types of information is treating tables as plain text [6, 7, 43]. In this way, the problem is reduced to document retrieval, but the characteristics of different types of information, especially the semantic dependency among table cells is ignored. Text in each cell of a table contains limited knowledge and are sometimes meaningless without dependencies. For example, in Fig. 1(c), numerical cells do not provide meaningful knowledge unless aligned with attributes in row and column headers. As an attempt to capture the information in different substructures of a table, Chen et al. [15] designed an embedding-based feature selection technique to select from each table the rows, columns and cells that are relevant to a query. Then they applied the pre-trained language model BERT [17] to encode selected table content. Shrager et al. [49] treated different types of information as multimodal objects and used recurrent neural networks (RNN) or convolutional neural networks (CNN) to encode

<table><tr><th>Lake</th><th>Area</th></tr><tr><td>Windermere</td><td>5.69 sq mi</td></tr><tr><td>Ullswater</td><td>3.86 sq mi</td></tr><tr><td>Derwent Water</td><td>2.06 sq mi</td></tr></table>	Lake	Area	Windermere	5.69 sq mi	Ullswater	3.86 sq mi	Derwent Water	2.06 sq mi	<table><tr><th>Country</th><td>United States</td></tr><tr><th>State</th><td>California</td></tr><tr><th>County</th><td>Los Angeles</td></tr><tr><th>Region</th><td>South California</td></tr></table>	Country	United States	State	California	County	Los Angeles	Region	South California	<table><tr><th></th><th>Right-handed</th><th>Left-handed</th></tr><tr><td>Males</td><td>43</td><td>9</td></tr><tr><td>Females</td><td>44</td><td>4</td></tr><tr><td>Totals</td><td>87</td><td>12</td></tr></table>		Right-handed	Left-handed	Males	43	9	Females	44	4	Totals	87	12	<table><tr><th colspan="2"></th><th colspan="3">To</th></tr><tr><th rowspan="4">From</th><th>Solid</th><td>Solid</td><td>Liquid</td><td>Gas</td></tr><tr><th>Solid trans</th><td>Solid trans</td><td>Melting</td><td>Sublimation</td></tr><tr><th>Liquid</th><td>Freezing</td><td>-</td><td>Boiling</td></tr><tr><th>Gas</th><td>Deposition</td><td>Condensation</td><td>-</td></tr></table>			To			From	Solid	Solid	Liquid	Gas	Solid trans	Solid trans	Melting	Sublimation	Liquid	Freezing	-	Boiling	Gas	Deposition	Condensation	-
Lake	Area																																																				
Windermere	5.69 sq mi																																																				
Ullswater	3.86 sq mi																																																				
Derwent Water	2.06 sq mi																																																				
Country	United States																																																				
State	California																																																				
County	Los Angeles																																																				
Region	South California																																																				
	Right-handed	Left-handed																																																			
Males	43	9																																																			
Females	44	4																																																			
Totals	87	12																																																			
		To																																																			
From	Solid	Solid	Liquid	Gas																																																	
	Solid trans	Solid trans	Melting	Sublimation																																																	
	Liquid	Freezing	-	Boiling																																																	
	Gas	Deposition	Condensation	-																																																	
(a) Relational table	(b) Entity table	(c) Matrix table	(d) Nested table																																																		

(a) Relational table

(b) Entity table

(c) Matrix table

(d) Nested table

Figure 1: Example snippets of tables with diverse layout structures.

each of them. Sun et al. [53] utilized attention mechanism to select cell embeddings over each row and each column. The approaches in prior studies have offered decent performance in intrinsic evaluation settings [15, 53, 66]. However, they are still quite distant from a generalizable solution to the NLTR problem.

To effectively address the NLTR problem, a learning-based system needs to tackle three aspects of generalization issues, which are however overlooked by prior approaches. First, as shown in Fig. 1, table cells are organized in diverse layouts to express the complex dependencies between cells. For example, we find that around 63.8% of WikiTables [66] come with nested structures of merged cells. Failing to consider these complex layout structures prevents the extraction and synthesizing of semantic knowledge from these tables. Second, natural language queries may have varied intents, referring to various granularities of content stored in different subunits of a table, such as cells, rows and columns. For example, the table in Figure 2(a) is relevant to queries for “taxable wages” at table-level, “dependent allowance” at row-level, and “yearly aggregates” at column-level. Sun et al. [53] analyzed a subset of the WebQuery-Tables dataset and found that about 24.5% queries are asking for information in specific subtable units while about 69.5% are asking for a whole table. Third, tables and queries from different datasets can possess dissimilar content, therefore requiring a generally applicable retrieval model to be adaptive to different datasets. Some work [15, 53] collects datasets from different sources and perform intrinsic evaluation on each of them. Nonetheless, prior methods fall short under cross-dataset evaluation [14], i.e. training on one dataset and testing on another dataset.

To this end, this paper proposes a novel table retrieval framework, namely **Graph-based Table Retrieval (GTR)**, to tackle the generalization issues (Sect. 3). GTR leverages state-of-the-art graph representation learning techniques to capture both content and layout structures of complex tables. Specifically, GTR incorporates a process to convert layout structures to *tabular graphs*, with cell nodes, row nodes, and column nodes to capture tabular information in different subunits. A variant of Graph Transformer [34] is then applied for automatic feature extraction from tabular content (Sect. 3.2), providing a structure-aware and multi-granular semantic representation for each unit. Then given a query, the framework uses two matching modules to measure the relevance scores based on both cells and contextual information of tables (Sect. 3.3). To further improve the adaptivity of GTR, we design a pre-training process to enhance the robustness of the query-graph matching module (Sect. 3.4). Experimental results on two benchmark datasets show that our method achieves significant improvements over prior state-of-the-art methods, even without pre-training (Sect. 4.2). In particular, our method outperforms the best-performing baseline on both datasets by 8.27% and 4.97% in terms of MAP. Further experimentation (Sect. 4.3) also demonstrates that the GTR framework

exhibits promising cross-dataset generalization performance, and shows stronger ability to handle complex tables and diverse query intents than existing methods.

In summary, this paper makes three major contributions: (1) a novel graph representation strategy captures complex layouts of tables and preserves structural dependencies between table units; (2) a query-graph matching process with a pre-trained Graph Transformer provides robust characterization of tables and supports multi-granular feature extraction for varied query intents; (3) a comprehensive set of experimentation shows the superior performance of the proposed method based on NLTR benchmarks and verifies the effectiveness in terms of cross-dataset generalization, complex table representation and query intent adaptation.

2 RELATED WORK

We review two relevant research topics. Both topics have a large body of work, for which we provide a selected summary.

2.1 Natural Language Table Retrieval

Earlier approaches for NLTR [6, 38, 43] treated tabular data as plain text and used BM25-style [45] methods to retrieve tables in the same way as document retrieval. Following this line, later work attempted to improve via better feature engineering, which involved handcrafted statistical and semantic features [5, 7], or utilized lexical embeddings [65, 66]. These studies provided feasible solutions to NLTR. However, the extracted features used in these approaches have limited coverage on queries and tabular content, as they focused on specific facets. Moreover, some strong features (e.g. entities and categorical features [66]) were only available in specific scenarios, limiting their generalizability.

Recent work focused on neural network approaches [15, 49, 53, 67]. Sun et al. [53] proposed query-specific attention mechanisms to aggregate table cell embeddings, which provided a flexible way to induce the relevance between a query and different parts of a table with softmax classifiers. Through this direction, Chen et al. [15] designed an embedding-based feature selection technique to select most relevant content from cells, rows and columns of each table, where BERT [17] was used to encode the concatenated text sequence of selected table content. Chen et al. [15] also observed that combining neural network methods and feature-based methods achieved further improvements. Shraga et al. [49] treated different facets of a table, including descriptions, schemas, rows and columns as different data modalities, and incorporated a multi-channel neural network to capture all modalities to be retrieved. The network was trained with both query-independent and query-dependent objectives. Zhang et al. [67] constructed a graph for the query, and headers, captions and cells of a table, and incorporated a graph convolutional network (GCN) [32] based classifier to predict the

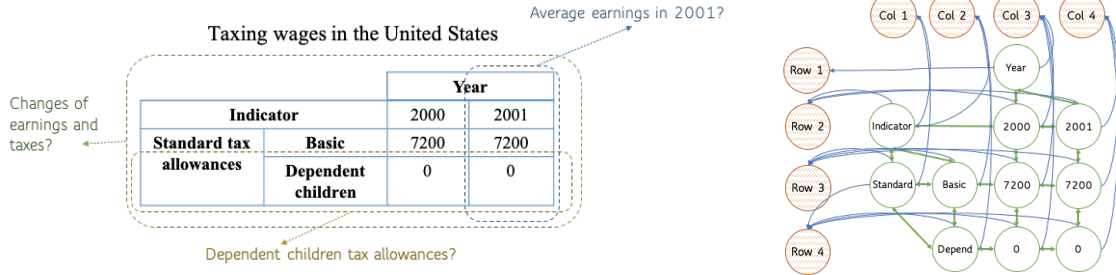


Figure 2: Multi-granular query intents and the tabular graph construction process for a table with complex layout.

query-content relevance scores. In addition, there have been research efforts on cascade re-ranking based on the results of single [50] or multiple [48] table retrieval methods.

The proposed table retrieval approach in this paper is connected to neural network approaches. The main difference lies in two perspectives: (1) our approach offers flexible representations of tables suitable for characterizing diverse and complex table layouts; (2) it better captures semantic information of a table in various subunits, hence can handle situations where queries are intended for different granularities of content.

2.2 Pre-training on Semi-structured Data

Some recent efforts have been made to pre-train graph neural networks for modeling structured or semi-structured data. Previous studies proposed node-level and graph-level pre-training tasks. Node-level pre-training methods, including Variational Graph Auto-Encoders [31], GraphSAGE [23], Graph Infomax [56], and GPT-GNN [27], aimed to support downstream tasks that relied on node representations. Graph-level pre-training methods, such as InfoGraph [52], sought to support inductive representation learning for global prediction tasks on graphs or subgraphs. In addition, Hu et al. [26] combined both node-level and graph-level pre-training. Our pre-training process is connected to the supervised graph-level property prediction by Hu et al. [26], but we match the graph representations to textual representations.

Fewer works have attempted pre-training on tabular data. Early work [20, 22] pre-trained embeddings for words or cells in tabular data according to co-occurrence. Inspired by the recent success of BERT [17] in language modeling, researchers extended BERT for encoding sequences extracted from tables and achieved state-of-the-art performance on semantic parsing over relational tables. Yin et al. [64] proposed TaBERT, which was pre-trained by recovering masked words, masked cells, and names and data types of masked columns. Herzig et al. [24] proposed TAPAS, which used a masked language model objective as BERT, but applied whole word and whole cell masking. Those aforementioned techniques however, do not capture the diverse and complex layout structures of tables, nor do they support a way of multi-granular aggregation of table information, which are both essential to tackling the NLTR task.

3 METHOD

In this section, we first provide a formal definition of the NLTR task (Sect. 3.1), and then introduce the architectures of the main

components in our framework (Sections 3.2 and 3.3). This is followed by the technical details of training, pre-training (Sect. 3.4) and inference (Sect. 3.5) processes.

3.1 Preliminaries

Task Definition. Given a natural language search query $q \in Q$ and a set of tables $\mathcal{T}_q = \{T_1, T_2, \dots, T_p\}$, the goal of the NLTR task is to rank tables from \mathcal{T}_q according to how likely q can be satisfied by the information in each table. The main body of a table contains two types of content, i.e. *table cells* and *contextual information*. Table cells can contain both header cells that describe attributes or names, and basic data cells² [67]. Different from previous works, we do not make the assumption that these table cells necessarily form a matrix-like layout [15, 49]. Meanwhile, tables are associated with *contextual information* (e.g. captions and footnotes), which have also been used as side information to characterize a table in previous works [15, 49, 67]. Following the aforementioned works, our work also leverages these two components, i.e. table cells and contextual information, to characterize a table. Fig. 2(a) shows an example. The caption “Taxing wages in the United States” is the contextual information and the main content is presented in several table cells.

Method Overview. The overall architecture of GTR is given in Fig. 3. Specifically, the framework consists of three model components: (1) a query-graph matching module captures the cells of a table with a multi-granular graph representation (Sect. 3.2), and estimates how relevant the cell content is based on such a representation (Sect. 3.3); (2) a query-context matching module assesses the relevance of a table to a query based on the contextual information associated with the table (Sect. 3.3); and (3) a ranking module that combines the outputs of two aforementioned matching modules and calculates the relevance score for each candidate query-table pair (Sect. 3.3). To further explore the potential of our framework, especially the generalizability of the query-graph matching module, we also incorporate a novel pre-training process (Sect. 3.4).

3.2 Graph Representation of Tables

Tabular Graph Construction. To effectively characterize a table with arbitrary layout, the first step is to transform the table into a multi-granular graph representation. In detail, a table T can be split into a set of units. Each unit refers to an individual cell, a row or a column. Units are interconnected in different ways depending

²Sometimes table titles and footnotes also appear as (merged) cells.

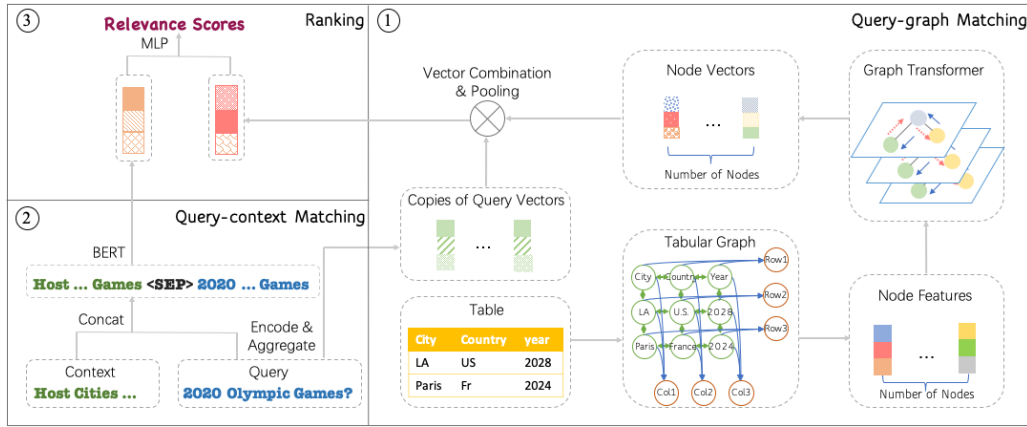


Figure 3: The overall architecture of the GTR framework. *Vector Combination* thereof refers to operations in Sect. 3.3. Numbered markers correspond to the components described in Method Overview (Sect. 3.1).

on the table layout, and each unit can contain different types of information. This characteristic of tables is similar to that of graphs. In particular, each table unit can be treated as a node $v_i \in V$ in a graph $G = (V, E)$, and the connection between two table units becomes an edge $e = (v_i, v_j) \in E$ between their corresponding nodes. In this paper, we consider two kinds of connections between table units: between adjacent cells; and between a unit and its subunit, such as a row and a cell in this row.

Fig. 2 shows an example of a table and its tabular graph. Suppose we have a table about taxing wages. The table has 11 cells in total where four cells are merged cells. In the corresponding tabular graph, each cell has an associated node. In addition, we have global nodes that capture the information of each row and column. Between each pair of adjacent cell nodes, we add a bidirectional edge. Every global node has a unidirectional edge sourced from each constituent cell node. These edges allow global nodes to aggregate information from the respective cells in the row or column.

Tabular Graph Transformer. Once a table is converted to a tabular graph, we use a variant of Graph Transformer [34] to characterize both cell content and layout structures. In detail, this process starts with initial representations $\mathbf{V}^0 = \{\mathbf{v}_i^0\}$ of node features, which are obtained using a pre-trained text encoder to encode the table unit content referring to each node³. We compare different text encoders in Sect. 4.4. These encodings along with the tabular graph introduced in the previous section are used as inputs to a Graph Transformer. Every l -th layer of the Graph Transformer thereof incorporates a multi-head self-attention layer with residual connection, a feedforward neural network layer (*FFNN*) and layer normalization (*LayerNorm*) [1]:

$$\mathbf{v}_i^{l+1} = \text{LayerNorm}(\text{FFNN}(\sigma(\mathbf{W}_g^l \mathbf{v}_i^l + \sum_{j \in \mathcal{N}_i} \alpha_{ij}^{lh} \mathbf{W}_g^l \mathbf{v}_j^l))),$$

where σ denotes the Leaky Rectified Linear Unit (LeakyReLU) [40], \mathbf{W}_g^l is a trainable weight matrix that transforms \mathbf{v}_i^l to the same size of the output of multi-head self-attention layer, \parallel denotes the concatenation operation over H attention heads, \mathcal{N}_i denotes the

neighborhood of node v_i in G , and α_{ij}^{lh} is the attention⁴ score of node v_j to node v_i in the h -th head of the l -th layer:

$$\alpha_{ij}^{lh} = \frac{\exp(a_{ij}^{lh})}{\sum_{j' \in \mathcal{N}_i} \exp(a_{ij'}^{lh})},$$

$$a_{ij}^{lh} = \sigma(\mathbf{w}_{lh}^T [\mathbf{W}_a^{lh} \mathbf{v}_i^l \parallel \mathbf{W}_a^{lh} \mathbf{v}_j^l]).$$

We stack L layers of Graph Transformer to allow tabular features to pass through the graph structure. Note that both local neighborhood information of cell nodes and global information of row and column nodes are captured through message passing.

3.3 Query-Table Matching

Query-Graph Matching. After obtaining the embedding representation of tabular content, we then need to match the tabular graph with the query. This process is completed by the query-graph matching module. Given representations of individual nodes in the graph, we first apply a linear transformation and layer normalization to these representations:

$$\mathbf{v}_i = \text{LayerNorm}(\mathbf{W}_1 \mathbf{v}_i^l + b_1).$$

Meanwhile, a sentence encoder is used to encode the query (implementation details see Sect. 4.1). Each encoded node representation $\mathbf{v}_i \in \mathbf{V}$ and the query representation \mathbf{q} are concatenated together with their element-wise subtraction and Hadamard product:

$$\hat{\mathbf{h}}_i = [\mathbf{v}_i \parallel \mathbf{q} \parallel \mathbf{v}_i - \mathbf{q} \parallel \mathbf{v}_i \circ \mathbf{q}]$$

Note that this is shown to be a comprehensive way to model embedding interactions in previous works [58, 69]. Then another non-linear transformation with \tanh ⁵ activation function is applied to produce hidden representations:

$$\mathbf{h}_i = \text{Tanh}(\mathbf{W}_2 \hat{\mathbf{h}}_i + b_2).$$

⁴We use additive attention mechanism [2]. Other mechanisms, such as dot-product attention [39], can also be applied, though we observe similar performance.

⁵The sequence output of BERT is activated by tanh function, which keeps the outputs of query-context matching and query-graph matching modules at the same scale.

³For global nodes, we average text embeddings of constituent cells as node features.

Finally, we aggregate the hidden representations of all nodes with a max-pooling operation, where $|V|$ is the number of nodes in G :

$$\mathbf{h}_{qd} = \text{MaxPooling}(\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{|V|}).$$

Note that the design of this module is motivated by the need for fulfilling varied queried intents in NLTR. The representations of cell, row and column nodes encoded by Graph Transformer naturally summarize on various subunits of the table as potential references for different queries. The pooling operator over nodes is for selecting summarized content that is most relevant to a query.

Query-Context Matching. Contextual information associated with tables can provide side information indicating the content of tables. To capture such side information \mathbf{h}_{qc} , we use a query-context matching module. Specifically, query-context matching can be seen as short-short or long-short text matching regarding to the length of context. Successful text matching models, such as BERT, can be used as the backbone of this module.

Learning Objective. The final query-table matching representation $\mathbf{h}_{qt} = [\mathbf{h}_{qd} \parallel \mathbf{h}_{qc}]$ is then fed to a multi-layer perceptron (MLP) to calculate the relevance score s_k . Recall that the goal of the NLTR task is to rank a collection of tables $\mathcal{T}_q = \{T_1, T_2, \dots, T_p\}$ according to their relevance scores to query $q \in Q$. Since we have the relevance score s_k for each table T_k , any ranking objectives can be used here. Following Chen et al. [15], the default setting approximates point-wise ranking with a mean square error (MSE) loss:

$$\text{MSE} = \frac{1}{|Q|} \sum_{q \in Q} \frac{1}{|\mathcal{T}_q|} \sum_{k=1}^{|\mathcal{T}_q|} (s_k - y_k)^2,$$

where y_k is the gold label (relevance score) of table t_k . Specifically, in the scenario where each query has only one relevant table, using the negative loglikelihood objective can achieve better performance [53]. Following Sun et al. [53], we use negative loglikelihood (NLL) as the loss function for this specific situation:

$$\text{NLL} = -\frac{1}{|Q|} \sum_{q \in Q} \log\left(\frac{\exp(s_{\hat{k}})}{\sum_{k=1}^{|\mathcal{T}_q|} \exp(s_k)}\right),$$

where \hat{k} is the index of the only relevant table to each query.

3.4 Pre-training

In order to support robust characterization of tables, we design a self-supervised pre-training task, following the principles suggested by Chang et al. [8]: (1) the pre-training task should capture self-supervision signals that are relevant to the downstream task, so that the pre-trained model can acquire essential characteristics for solving the downstream task; and (2) the pre-training task should be cost-efficient in terms of pre-training data, ideally relying on free or self-generated labels. Considering that a query and the contextual information of a relevant table contain semantically related information, we use the contextual information as free-labels during pre-training.

Graph-Context Matching. Specifically, we reuse the architecture of the query-graph matching module. Different from the original query-graph matching process, the pre-training process treats the contextual information of tables as queries and performs *graph-context matching*. During pre-training, for each table T with the

contextual information c , we randomly select the contextual information c' of another table. We refer c as a positive context, and c' as a negative context. The “context” here provides the same functionality as the “query” in Sect. 3.3. The objective of pre-training is to make T more relevant to c than c' . Suppose the relevance scores from the query-graph matching module for c and c' , are s and s' , we also apply MSE as the loss function such that the ground-truth scores for c and c' are 1 and 0, respectively.

3.5 Inference

During inference, GTR retrieves tables based on both the table cells and contextual information. Fig. 3 depicts the whole pipeline. Given a query, for each candidate table that has been converted to its tabular graph, both query-graph matching and query-context matching modules yield the combined representation of the query and corresponding information of the table. As described above, the query-graph matching module has performed a pooling operation to filter the relevant information in the tabular graph. Then, representations from both modules are further combined for an estimation of relevance score $s_k = \text{MLP}(\mathbf{h}_{qt})$. For each query, we sort all candidate tables directly by the estimated relevance scores.

4 EXPERIMENT

In this section, we conduct experiments based on two benchmark datasets (Sect. 4.1) and compare the performance of GTR against a series of recent baselines (Sect. 4.2). We also provide quantitative analysis on the generalizability of our method (Sect. 4.3), and conduct detailed ablation studies (Sect. 4.4) and case studies (Sect. 4.5) to help understand the contribution of different model components.

4.1 Experimental Setup

Datasets. We conduct experiments on two benchmark datasets, i.e. WikiTables [66] and WebQueryTable [53]. The relevant query-table pairs of these two datasets are collected from different sources. More details and statistics of the datasets are described as follows:

- **WikiTables:** WikiTables is a widely-used dataset for the NLTR task. It contains 60 queries that are contributed by two previous studies [6, 57]. 3,120 candidate tables were extracted from Wikipedia. All candidate tables were labeled by annotators with one of three relevance scores: 0 (irrelevant), 1 (relevant), and 2 (highly relevant). Each table is associated with contextual information including a caption, Wikipedia’s page title and section title. We analyzed the cell adjacency of these tables and discovered that 1,886 of them came with nested layout structures involving merged cells. Following previous works [15, 66], we run 5-fold cross-validation on this dataset.
- **WebQueryTable:** The WebQueryTable dataset contains 21,113 queries collected from search logs of a commercial search engine and 273,816 tables. For each query, one relevant table was obtained from the top ranked Web page of the same search engine after manual evaluation. Captions of tables are also given in this dataset as contextual information. We use the originally released training, validation and test set splits [53] for evaluation.

Table 1: Retrieval performance on WikiTables. The best performing method in each column is boldfaced, and the second best method is underscored. Baselines are organized into (1) unsupervised, (2) feature engineering and (3) end-to-end groups.

Method	NDCG@5	NDCG@10	NDCG@15	NDCG@20	MAP
BM25	0.3196	0.3377	0.3732	0.4045	0.4260
WebTable	0.2980	0.3150	0.3486	0.3922	-
SDR	0.4573	0.4841	0.5195	0.5534	-
MDR	0.5021	0.5116	0.5451	0.5761	-
Tab-Lasso	0.5161	0.5018	0.5330	0.5481	-
LTR	0.5910	0.5712	0.5858	0.6041	0.5615
TaBERT	0.5926	0.6108	0.6451	0.6668	0.6326
BERT4TR	0.6052	0.6171	0.6386	0.6689	0.6191
GTR (w/o pre-training)	0.6554	0.6747	0.6978	0.7211	0.6665
GTR	0.6671	0.6856	0.7065	0.7272	0.6859

Table 2: Retrieval performance on WebQueryTable. P@1 is not reported by the BERT4TR paper.

Method	P@1	MAP
BM25	0.4712	0.5823
MDF	0.4779	0.6102
MNN	0.4902	0.6194
TaBERT	0.5067	0.6338
BERT4TR	-	0.7104
GTR (w/o pre-training)	0.6257	0.7369
GTR	0.6358	0.7457

Baselines. We compare our framework GTR with the following 10 strong baseline methods⁶.

- **BM25 [45]:** Okapi BM25 is an unsupervised method using token-matching with TF-IDF [44] weights as the scoring function.
- **WebTable [7]:** WebTable is a method based on linear regression using hand-crafted features.
- **SDR [6]:** Single-field Document Ranking (SDR) treats a table as a regular document, and uses a symmetric conditional probability model with Dirichlet smoothing to capture query-table relevance.
- **MDR [43]:** Multi-field Document Ranking (MDR) extends SDR by treating each table as multiple separated fields of text. Each field corresponds to page titles, table section titles, table captions, table body, or table headings, respectively. MDR also uses coordinate ascent algorithm [62] to learn the aggregation weights.
- **Tab-Lasso [5]:** Tab-Lasso is a Lasso [54] model with coordinate ascent, taking well-designed hand-crafted features as input.
- **LTR [66]:** Lexical Table Retrieval (LTR) is a strong non-neural baseline, which employs point-wise regression using Random Forest [25] with features from WebTable [7] and Tab-Lasso [5].
- **MDF [53]:** Matching with Designed Features (MDF) matches queries and tables based on lexical similarity using IDF scores, phrasal similarity using phrase dictionary tables [33], and sentential similarity using the CDSSM [47] model, respectively.

⁶We did not compare with STR [66] which used additional entity and categorical information. Also, MTR [49] was reported in a different experimental setting, but its implementation had not been released by the time this paper was written.

- **MNN [53]:** Matching with Neural Networks (MNN) is a method that uses bi-directional gated recurrent unit (GRU) [16] to encode queries and captions, and applies query-specific attention mechanisms to aggregate cell embeddings to represent table units.
- **BERT4TR [15]:** BERT for Table Retrieval (BERT4TR) is the previous state-of-the-art method that applies the pre-trained language model BERT [17] to encode flattened tables. An embedding-based selection process is first utilized to select the most relevant rows from tables with respect to queries. Then the query, contextual information of a table and selected tabular content is flattened as a sequence and encoded by BERT, on top of which an MLP is stacked to compute the relevance score.
- **TaBERT [64]:** TaBERT is a more recently released language model that is pre-trained on a large corpus of 26 million tables and their English contexts. It has been previously applied to semantic parsing on tables and offered state-of-the-art performance. We apply this strong table representation learning method for NLTR. Similar to BERT4TR, we also stack an MLP scorer to calculate the query-table relevance scores.

For WebTable, SDR, MDR, Tab-Lasso and LTR, we use the implementation from Zhang and Balog [66]⁷. The results for BM25, MDF, MNN and BERT4TR on WebQueryTable dataset are obtained from their original papers, where experiments are conducted using the same data split. For BERT4TR on WikiTables dataset and TaBERT on both datasets, we use the original implementation, configuration and preprocessing steps released by the authors.

Evaluation Metrics. Considering the different annotation strategies of the two datasets, we adopt different groups of metrics to evaluate the retrieval performance on each dataset, as to be consistent with prior studies. On WikiTables, following previous works [15, 66], we report Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (NDCG@ k) with cut-off points $k = \{5, 10, 15, 20\}$. On WebQueryTable, we report MAP and Precision at 1 (P@1) following Sun et al. [53]. Specifically, MAP and NDCG metrics are calculated using the TREC evaluation tool⁸.

⁷Some baseline results in Tab. 1 may be different from those reported in previous works due to different setups of cross-validation.

⁸https://github.com/usnistgov/trec_eval

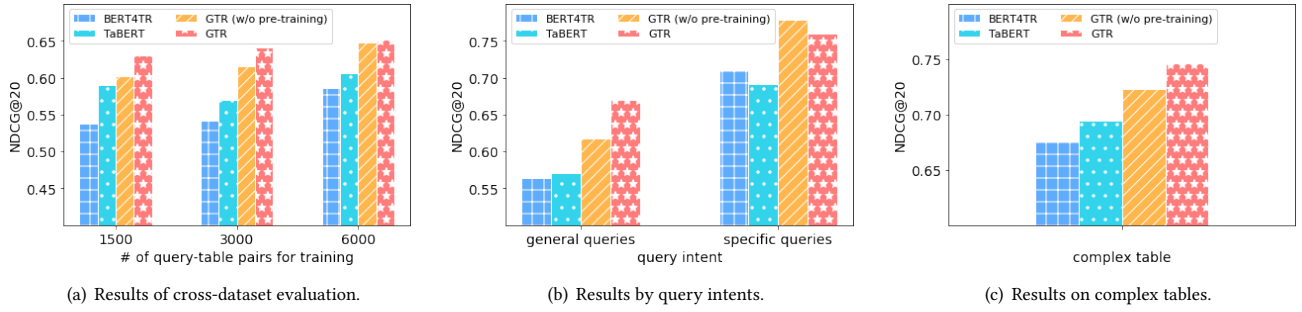


Figure 4: Generalizability analysis results.

Implementation Details. For the default version of GTR, we use BERT as the query-context matching module and FastText [28] as the text encoder for query-graph matching module (see Sect. 4.4 for ablation on text encoders). In query-graph matching module, we use four layers of Graph Transformer ($L = 4$) with four self-attention heads ($H = 4$). The dimensionality of hidden states is set to 300. In query-context matching module, multiple text sequences from the query and context are concatenated and fed to BERT. Following Devlin et al. [17], we add a [CLS] token at the beginning of the input sequence, and separate query and context with a [SEP] token. Different segment embeddings are assigned to distinguish query from context. We use the final output of the first token [CLS] as the hidden state of query-context matching. As described in Sect. 3.3, the learning objective is set to be MSE on WikiTables where multiple relevant tables coexist for each query, and that is set to a NLL loss on WebQueryTable where each query has only one relevant table.

Both pre-training and the main training process use an Adam optimizer with learning rate set as 0.0001. Pre-training of the query-graph matching module uses tables from both WikiTables and WebQueryTable datasets. It is conducted for 20 epochs with a batch size of 16, so as to fit into one RTX 2080 GPU. The main training process on both datasets takes 5 epochs with a linear learning rate scheduler with warmup steps. The training configurations on both datasets are slightly different, as being limited by the GPU memory. On the WikiTables dataset, we use batch size of 16 and warmup steps of 100. On the WebQueryTable dataset, we use batch size of 4 and warmup steps of 1000. Trainable parameters other than text encoders are initialized using the Xavier initializer [21]. Dropout rate of 0.1 is applied to each Graph Transformer layer and before the final MLP. The negative slope of LeakyReLU is set as 0.2. We use Pytorch [42] and DGL [60] to implement our framework.

4.2 Main Results

The main results presented here are under the intrinsic evaluation protocol following the design of the two NLTR benchmarks.

As reported in Tab. 1 and Tab. 2, among the baseline methods, the pre-trained Transformer language model based BERT4TR demonstrates state-of-the-art performance over other baselines. The reason is that that pre-trained language models more comprehensively capture the semantic information of table cell content and context information, in comparison to a number of other baselines based on explicit features or static embeddings. In particular, though TaBERT

follows is similar to BERT4TR, it offers a less performance. We hypothesize that since TaBERT is designed for semantic parsing tasks where the focus is to capture column relations, it does not necessarily support well summarizing table content and inferring the query-table affinity.

We observe that our method, even without pre-training, outperforms BERT4TR and TaBERT with at least relative improvements of 8.29% in terms of NDCG@5, 9.34% in terms of NDCG@10, 8.17% in terms of NDCG@15, 7.80% in terms of NDCG@20 and 5.36% in terms of MAP, on the WikiTables dataset. It is noteworthy that, both BERT4TR and TaBERT are Transformer-based architectures that flatten table cells to sequences. This strategy of representation necessarily discards the dependencies between subtable content that are modeled in the table layouts. The graph representation and coupled tabular Graph Transformer in our framework preserve the original structures of table content, and encapsulate features of table cells in different granularities. The experimental results verify our hypothesis that structure-aware representations and multi-granular information of tables are conducive to general purpose NLTR.

Pre-training the query-graph matching module further leads to at least a relative improvement of 1.78% in terms of NDCG@5, and that of 2.91% in terms of MAP. This is attributed to that the query-graph matching module acquires more robust characteristics of tables during the pre-training process, hence particularly benefits the training that does not involve lots of data.

The evaluation on the WebQueryTable dataset supports the same conclusion. GTR outperforms BERT4TR with a relative increase of MAP by 3.73% without pre-training and 4.97% with pre-training, with more improvement in comparison to other baselines. The two datasets have different annotation strategies and sources of relevant query-table pairs. These results indicate that GTR adapts well to different scenarios of NLTR.

4.3 Generalizability Analysis

We further present several aspects of generalizability experiments, with detailed analysis on cross-dataset generalization, reactions to query intents and performance on complex tables. In these experiments, we compare GTR with the two best-performing baselines BERT4TR and TaBERT.

Cross-dataset Evaluation. In the first experiment, we compare NLTR methods in an inductive evaluation setting, seeking to examine how well they can transfer knowledge to retrieve tables across datasets. Specifically, we train GTR and the two baselines on

Table 3: Ablation study on framework components. *Removing Tabular Graph* removes the entire query-graph matching module. *Removing Edges* keeps nodes but removes all edges in the tabular graph. *Multi-head GAT* uses a multi-head Graph Attention Network as the encoder. *Removing Row and Col Nodes* removes row and column nodes. *Node Initialization with BERT* replaces FastText with BERT as the cell text encoder. ↓ marks a significant drop of a metric by at least 4% relatively.

Setting	NDCG@5	NDCG@10	NDCG@15	NDCG@20	MAP
Default	0.6554	0.6747	0.6978	0.7211	0.6665
- Removing Tabular Graph	0.5979 ↓	0.6118 ↓	0.6395 ↓	0.6606 ↓	0.6231 ↓
- Removing Edges	0.6190 ↓	0.6438 ↓	0.6669 ↓	0.6956	0.6513
- Multi-head GAT	0.6458	0.6553	0.6728	0.6977	0.6546
- Removing Row and Col Nodes	0.6403	0.6566	0.6704	0.6922 ↓	0.6494
- Node Initialization with BERT	0.6472	0.6417 ↓	0.6652 ↓	0.6967	0.6472

a subset of query-table pairs from WebQueryTables, and evaluate on WikiTables. Note that the relevant tables in the two datasets are collected from different sources. In addition, to study the data efficiency of training the models, we vary the size of the training set to be 1,500, 3,000, and 6,000, which are approximately half, equal and double of the size of the test set, respectively.

The results are accordingly presented in Fig. 4(a), which indicate that our method yields better performance than BERT4TR and TaBERT on each setting of the training data. In particular, GTR exhibits better generalization performance even with half of the training data (by offering 0.6157 in terms of NDCG@20), in comparison to BERT4TR and TaBERT that are trained with full data (which achieve 0.5859 and 0.6060 in terms of NDCG@20, respectively). We also observe that when the number of training samples decreases, the performance of the previous state-of-the-art system BERT4TR drops more drastically, while both TaBERT and GTR are relatively more stable. Moreover, when the training set is small (1,500), the performance of TaBERT is close to GTR and is much better than BERT4TR. We believe this is because TaBERT has learned more adaptive table encoding than BERT by pre-training on large table corpus, hence offering better cross-dataset generalization than BERT4TR when without sufficient fine-tuning data. However, it still drastically fall behind GTR.

Meanwhile, we observe noticeable performance drop by our method when pre-training is disabled, especially in cases with less training data. This indicates the effectiveness of pre-training to improve cross-dataset generalization. Though even without pre-training, GTR still consistently outperforms the two strong baselines.

Performance by Query Intents. In the second experiment, we show how well GTR and both baselines react to query intents on different granularities of content. Following Sun et al. [53], we split queries from the WikiTables dataset into two main groups, i.e. *general queries* and *specific queries*, based on their intents. A general query usually refers to a whole table involving several aspects of objects while a specific query usually asks about a specific local aspect of the table in a row, a column or individual cells. For example, “*world interest rates table*” refers to a general query and “*2008 Olympics gold medal winners*” is a more specific query.

Fig. 4(b) presents the results. For both general and specific query intents, GTR significantly outperforms both BERT4TR and TaBERT. The reason is mainly attributed to that our graph representation

strategy, especially the incorporation of multi-granular node encoding, naturally provides a multi-granular content summarization to fulfill query intents of different specificities. Meanwhile, the two strong baseline methods perform differently on reacting to the query intents. Specifically, TaBERT performs slightly better than BERT4TR on general queries, but being worse on specific queries. This is most likely due to the difference in the table unit selection processes of these two methods. Both methods select highly relevant table units according to a given query as model inputs, but TaBERT creates synthetic rows by regrouping cell content from each column. Although this process remains the most relevant table content to queries, it may hinder the model to capture the original semantics of table units. We also observe that our method benefits much from pre-training to deal with general queries. We believe this benefits from contextual information (e.g. captions) in the pre-training task of graph-context matching (Sect. 3.4), where the contextual information serves as general descriptions of tables.

Performance on Complex Tables. In the third experiment, we evaluate how different methods are capable of handling complex tables. To do so, we preserve only 1,886 tables with nested structures in cross-validation. Results in Fig. 4(c) show that GTR, with or without pre-training, notably outperforms both TaBERT and BERT4TR under this setting. It verifies that our graph representation strategy is better at capturing complex table layouts than Transformer language model based methods, as our method captures the essential structural layout information rather than flattening table cells into a sequence. Interestingly, we observe that TaBERT performs better than BERT4TR on retrieving complex tables, which is just the opposite when retrieving from the whole table corpus (Tab. 1). Meanwhile, GTR also performs much better with pre-training, indicating pre-training task to be beneficial to complex table representation.

According to the experiments, our method is capable of effectively transferring knowledge cross datasets. In addition, the graph representation strategy allows for capturing multi-granular information to fulfill both general and specific intents.

4.4 Ablation Study

To further help understand the contribution of each incorporated model component, we hereby conduct several aspects of ablation studies based on WikiTables. The discussed results are in Tab. 3.

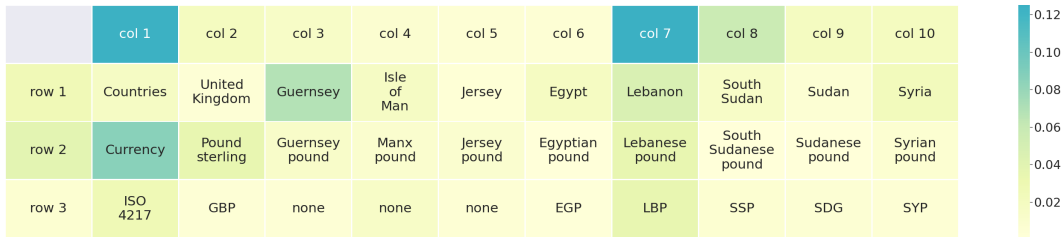


Figure 5: The heat-map of pooling operation in query-graph matching module by index selection frequency on a retrieved table for the query “asian countries currency”. Row nodes and column nodes are shown as the first column and the first row in the figure, respectively. Nodes with higher index selection frequency are displayed in darker colors.

Table Graph Representation. We first examine the effectiveness of the graph representation as well as the tabular Graph Transformer. As expected, the performance drastically drops when ignoring the information in table cells and ignoring the table layouts, leading to a relative drop of NDCG@5 by 8.77%. Besides, NDCG@5 decreases by relatively 5.56% when each table unit is captured independently (i.e., removing edges). The results indicate that both information inside each cell and dependencies among cells are important for comprehensive table understanding. Lastly, using multi-head Graph Attention Network (GAT) [55] instead of Graph Transformer leads to relatively 1.46% of drop in NDCG@5.

Row and Column Nodes. When removing row and column nodes from tabular graphs, the performance is lessened by 2.30% in terms of NDCG@5, and by 4.00% in terms of NDCG@20 relatively. Presumably this is because Graph Transformer with cell nodes alone cannot effectively capture row-wise and column-wise information. Thus, it is essential to use row nodes and column nodes for that level of coarse-grained information aggregation.

Text Encoders. We test if node initialization can benefit from a deep contextualized embedding. Interestingly, we observe that initializing node features by encoding cell text with BERT [17] performs worse than with FastText [28]. The NDCG@5 drops by 1.25% and the NDCG@20 drops by 3.38%, relatively. This is understandable, since each cell content is a standalone short piece of text that does not necessarily benefit from contextualized text embedding by BERT. On the contrary, the static embedding by FastText support with more stable semantic representation to jump-start the node features based on the short cell content. This is in line with the observation where static embeddings outperform contextualized embeddings on lexical and phrasal tasks [9, 18, 36].

4.5 Case Study

We present a case study with a representative example (Fig. 5) to illustrate how the graph representation supports with multi-granular information aggregation to fulfill the query intent. The importance of tabular graph nodes in the heat-map reflects the index frequency of node representations in max-pooling operation.

We observe when answering the query “asian countries currency”, the first column and the seventh column of the retrieved table contribute the most. This is reasonable as both columns cover some aspects of the query. The first column is the header of rows, indicating that this table is about countries and their currencies. The seventh column is the currency of an Asian country Lebanon. This

phenomenon shows the effectiveness of global nodes in tabular graphs. However, the tenth column, which is also about the currency of an Asian country Syria, does not attract as much attention as the seventh column does. One possible reason is that when more than one node covers similar information, the max-pooling operator may take the most informative one to leave capacity for other kinds of information. In this case, one of Lebanon and Syria is sufficient to fulfill the query intent about Asian countries. Moreover, we observe that some cell nodes, such as the node of “Currency” in the first column and nodes of cells in the seventh column, also have a high frequency to be selected by max-pooling. This shows that both global nodes and local nodes can contribute to query-graph matching. Some neighbors of the highly influential nodes mentioned above may also be frequently selected by pooling. This is likely attributed to that relevant information is propagated through the tabular graph from highly influential nodes to their neighbors.

5 CONCLUSION

In this paper, we proposed a novel framework for complex table retrieval. The GTR framework includes a tabular graph representation strategy that captures the cell structure dependencies, with row nodes and column nodes for high-level feature aggregation. GTR applies a tabular Graph Transformer to effectively support multi-granular feature extraction with tabular graphs as inputs. In addition, we introduced a self-supervised pre-training task which leverages the contextual information as free-labels, so as to enhance the robustness of the tabular Graph Transformer. At last, a comprehensive set of experiments and analysis show GTR’s state-of-the-art performance based on NLTR benchmarks, and demonstrate the capability of this framework in terms of cross-dataset generalization, handling complex table structures, and fulfilling diverse query intents. For future work, we plan to extend the use of GTR to other table-related tasks, such as table summarization [4, 11] and table-text grounding [30]. Applying the graph-based table representation for perceptual tasks, such as cell structure recognition [68] and functional block detection [29], is another meaningful direction.

ACKNOWLEDGEMENT

We appreciate the anonymous reviewers for their insightful comments and suggestions. This material is based upon work sponsored by the DARPA MCS program under Contract No. N660011924033 with the United States Office Of Naval Research, and by Air Force Research Laboratory under agreement number FA8750-20-2-10002.

REFERENCES

- [1] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. 2016. Layer normalization. *arXiv preprint arXiv:1607.06450* (2016).
- [2] Dzmitry Bahdanau, Kyung Hyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015*.
- [3] Sreeram Balakrishnan, Alon Halevy, Boulos Harb, Hongrae Lee, Jayant Madhavan, Afshin Rostamizadeh, Warren Shen, Kenneth Wilder, Fei Wu, and Cong Yu. 2015. Applying webtables in practice. In *7th Biennial Conference on Innovative Data Systems Research (CIDR '15)*.
- [4] Junwei Bao, Duyu Tang, Nan Duan, Zhao Yan, Yuanhua Lv, Ming Zhou, and Tiejun Zhao. 2018. Table-to-text: Describing table region with natural language. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.
- [5] Chandra Sekhar Bhagavatula, Thanapon Noraset, and Doug Downey. 2013. Methods for exploring and mining tables on wikipedia. In *Proceedings of the ACM SIGKDD Workshop on Interactive Data Exploration and Analytics*. 18–26.
- [6] Michael J Cafarella, Alon Halevy, and Nodira Khousainova. 2009. Data integration for the relational web. *Proceedings of the VLDB Endowment* 2, 1 (2009), 1090–1101.
- [7] Michael J Cafarella, Alon Halevy, Daisy Zhe Wang, Eugene Wu, and Yang Zhang. 2008. Webtables: exploring the power of tables on the web. *Proceedings of the VLDB Endowment* 1, 1 (2008), 538–549.
- [8] Wei-Cheng Chang, X Yu Felix, Yin-Wen Chang, Yiming Yang, and Sanjiv Kumar. 2019. Pre-training Tasks for Embedding-based Large-scale Retrieval. In *International Conference on Learning Representations*.
- [9] Muhao Chen, Weijia Shi, Pei Zhou, and Kai-Wei Chang. 2019. Retrofitting Contextualized Word Embeddings with Paraphrases. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.
- [10] Wenhui Chen, Ming-Wei Chang, Eva Schlinger, William Yang Wang, and William W. Cohen. 2021. Open Question Answering over Tables and Text. In *International Conference on Learning Representations*. <https://openreview.net/forum?id=MmCRsw1UYI>
- [11] Wenhui Chen, Jianshu Chen, Yu Su, Zhiyu Chen, and William Yang Wang. 2020. Logical Natural Language Generation from Open-Domain Tables. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 7929–7942.
- [12] Wenhui Chen, Hongmin Wang, Jianshu Chen, Yunkai Zhang, Hong Wang, Shiyang Li, Xiyu Zhou, and William Yang Wang. 2019. TabFact: A Large-scale Dataset for Table-based Fact Verification. In *International Conference on Learning Representations*.
- [13] Wenhui Chen, Hanwen Zha, Zhiyu Chen, Wenhan Xiong, Hong Wang, and William Yang Wang. 2020. HybridQA: A Dataset of Multi-Hop Question Answering over Tabular and Textual Data. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. 1026–1036.
- [14] Yiran Chen, Pengfei Liu, Ming Zhong, Zi-Yi Dou, Danqing Wang, Xipeng Qiu, and Xuan-Jing Huang. 2020. An Empirical Study of Cross-Dataset Evaluation for Neural Summarization Systems. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. 3679–3691.
- [15] Zhiyu Chen, Mohamed Trabelsi, Jeff Heflin, Yanan Xu, and Brian D. Davison. 2020. Table Search Using a Deep Contextualized Language Model. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (2020).
- [16] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1724–1734.
- [17] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 4171–4186.
- [18] Kavin Ethayarajh. 2019. How Contextual are Contextualized Word Representations? Comparing the Geometry of BERT, ELMo, and GPT-2 Embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 55–65.
- [19] Jing Fang, Prasenjit Mitra, Zhi Tang, and C Lee Giles. 2012. Table header detection and classification. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.
- [20] Majid Ghasemi-Gol and Pedro Szekely. 2018. Tabvec: Table vectors for classification of web tables. *arXiv preprint arXiv:1802.06290* (2018).
- [21] Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. 249–256.
- [22] Majid Ghasemi Gol, Jay Pujara, and Pedro Szekely. 2019. Tabular Cell Classification Using Pre-Trained Cell Embeddings. In *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 230–239.
- [23] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Advances in neural information processing systems*. 1024–1034.
- [24] Jonathan Herzig, Paweł Krzysztof Nowak, Thomas Müller, Francesco Piccinno, and Julian Eisenschlos. 2020. TaPas: Weakly Supervised Table Parsing via Pre-training. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 4320–4333.
- [25] Tin Kam Ho. 1995. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, Vol. 1. IEEE, 278–282.
- [26] Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. 2019. Strategies for Pre-training Graph Neural Networks. In *International Conference on Learning Representations*.
- [27] Ziniu Hu, Yuxiao Dong, Kuansan Wang, Kai-Wei Chang, and Yizhou Sun. 2020. GPT-GNN: Generative Pre-Training of Graph Neural Networks (*KDD '20*). 1857–1867.
- [28] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of Tricks for Efficient Text Classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. Association for Computational Linguistics, 427–431.
- [29] Sun Kexuan, Rayudu Harsha, and Jay Pujara. 2021. A Hybrid Probabilistic Approach for Table Understanding. In *Thirty-Fifth AAAI Conference on Artificial Intelligence*.
- [30] Dae Hyun Kim, Enamul Hoque, Juho Kim, and Maneesh Agrawala. 2018. Facilitating document reading by linking text and tables. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 423–434.
- [31] Thomas N Kipf and Max Welling. 2016. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308* (2016).
- [32] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations (ICLR)*.
- [33] Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*. 48–54.
- [34] Rik Koncel-Kedziorski, Dhanush Bekal, Yi Luan, Mirella Lapata, and Hannaneh Hajishirzi. 2019. Text Generation from Knowledge Graphs with Graph Transformers. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 2284–2293.
- [35] Rémi Lebret, David Grangier, and Michael Auli. 2016. Neural Text Generation from Structured Data with Application to the Biography Domain. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 1203–1213.
- [36] Qianchu Liu, Diana McCarthy, and Anna Korhonen. 2020. Towards Better Context-aware Lexical Semantics: Adjusting Contextualized Representations through Static Anchors. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 4066–4075.
- [37] Ying Liu, Kun Bai, Prasenjit Mitra, and C Lee Giles. 2007. Tableseer: automatic table metadata extraction and searching in digital libraries. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*. 91–100.
- [38] Ying Liu, Kun Bai, Prasenjit Mitra, C Lee Giles, et al. 2007. Tablerank: A ranking algorithm for table search and retrieval. In *Proceedings of the National Conference on Artificial Intelligence*, Vol. 22. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 317.
- [39] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective Approaches to Attention-based Neural Machine Translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 1412–1421.
- [40] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. 2013. Rectifier nonlinearities improve neural network acoustic models. In *International Conference on Machine Learning (ICML)*.
- [41] Panupong Pasupat and Percy Liang. 2015. Compositional Semantic Parsing on Semi-Structured Tables. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 1470–1480.
- [42] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library.. In *NeurIPS*.
- [43] Rakesh Pimplikar and Sunita Sarawagi. 2012. Answering table queries on the web using column keywords. *Proceedings of the VLDB Endowment* 5, 10 (2012), 908–919.
- [44] Juan Ramos et al. 2003. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, Vol. 242. Citeseer, 29–48.
- [45] SE ROBERTSON, S WALKER, S JONES, MM HANCOCK-BEAULIEU, and M GATFORD. 1995. Okapi at TREC-3. *NIST special publication* 500225 (1995), 109–123.

- [46] Michael Schlichtkrull, Vladimir Karpukhin, Barlas Öguz, Mike Lewis, Wen-tau Yih, and Sebastian Riedel. 2020. Joint Verification and Reranking for Open Fact Checking Over Tables. *arXiv preprint arXiv:2012.15115* (2020).
- [47] Yelong Shen, Xiaodong He, Jianfeng Gao, Li Deng, and Grégoire Mesnil. 2014. A latent semantic model with convolutional-pooling structure for information retrieval. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*. 101–110.
- [48] Roei Shraga, Haggai Roitman, Guy Feigenblat, and Mustafa Canim. 2020. Ad hoc table retrieval using intrinsic and extrinsic similarities. In *Proceedings of The Web Conference 2020*. 2479–2485.
- [49] Roei Shraga, Haggai Roitman, Guy Feigenblat, and Mustafa Cannim. 2020. Web Table Retrieval using Multimodal Deep Learning. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1399–1408.
- [50] Roei Shraga, Haggai Roitman, Guy Feigenblat, and Bar Weiner. 2020. Projection-based Relevance Model for Table Retrieval. In *Companion Proceedings of the Web Conference 2020*. 28–29.
- [51] MICHAEL STONEBRAKER, EUGENE WONG, PETER KREPS, and GERALD HELD. 1976. The Design and Implementation of INGRES. *ACM Transactions on Database Systems* 1, 3 (1976), 189–222.
- [52] Fan-Yun Sun, Jordan Hoffman, Vikas Verma, and Jian Tang. 2019. InfoGraph: Unsupervised and Semi-supervised Graph-Level Representation Learning via Mutual Information Maximization. In *International Conference on Learning Representations*.
- [53] Yibo Sun, Zhao Yan, Duyu Tang, Nan Duan, and Bing Qin. 2019. Content-based table retrieval for web queries. *Neurocomputing* 349 (2019), 183–189.
- [54] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58, 1 (1996), 267–288.
- [55] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. *International Conference on Learning Representations* (2018). <https://openreview.net/forum?id=rjXmpikCZ> accepted as poster.
- [56] Petar Veličković, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. 2018. Deep Graph Infomax. In *International Conference on Learning Representations*.
- [57] Petros Venetis, Alon Halevy, Jayant Madhavan, Marius Pasca, Warren Shen, Fei Wu, Gengxin Miao, and Chung Wu. 2011. Recovering semantics of tables on the web. *Proceedings of the VLDB Endowment* 4, 9 (2011), 528–538.
- [58] Haoyu Wang, Muhao Chen, Hongming Zhang, and Dan Roth. 2020. Joint Constrained Learning for Event-Event Relation Extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 696–706.
- [59] Hao Wang, Xiaodong Zhang, Shuming Ma, Xu Sun, Houfeng Wang, and Mengxiang Wang. 2018. A neural question answering model based on semi-structured tables. In *Proceedings of the 27th International Conference on Computational Linguistics*. 1941–1951.
- [60] Minjie Wang, Da Zheng, Zihao Ye, Quan Gan, Mufei Li, Xiang Song, Jinjing Zhou, Chao Ma, Lingfan Yu, Yu Gai, Tianjun Xiao, Tong He, George Karypis, Jinyang Li, and Zheng Zhang. 2019. Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks. *arXiv preprint arXiv:1909.01315* (2019).
- [61] Yushi Wang, Jonathan Berant, and Percy Liang. 2015. Building a semantic parser overnight. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 1332–1342.
- [62] Stephen J Wright. 2015. Coordinate descent algorithms. *Mathematical Programming* 151, 1 (2015), 3–34.
- [63] Xiaojun Xu, Chang Liu, and Dawn Song. 2017. SQLNet: Generating Structured Queries From Natural Language Without Reinforcement Learning. *arXiv preprint arXiv:1711.04436* (2017).
- [64] Pengcheng Yin, Graham Neubig, Wen-tau Yih, and Sebastian Riedel. 2020. TaBERT: Pretraining for Joint Understanding of Textual and Tabular Data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 8413–8426.
- [65] Li Zhang, Shuo Zhang, and Krisztian Balog. 2019. Table2Vec: neural word and entity embeddings for table population and retrieval. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1029–1032.
- [66] Shuo Zhang and Krisztian Balog. 2018. Ad hoc table retrieval using semantic similarity. In *Proceedings of the 2018 World Wide Web Conference*. 1553–1562.
- [67] Xingyao Zhang, Linjun Shou, Jian Pei, Ming Gong, Lijie Wen, and Daxin Jiang. 2020. A Graph Representation of Semi-structured Data for Web Question Answering. In *Proceedings of the 28th International Conference on Computational Linguistics*. 51–61.
- [68] Xinyi Zheng, Douglas Burdick, Lucian Popa, Xu Zhong, and Nancy Xin Ru Wang. 2021. Global table extractor (gte): A framework for joint table identification and cell structure recognition using visual context. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 697–706.
- [69] Guangyu Zhou, Muhao Chen, Chelsea JT Ju, Zheng Wang, Jyun-Yu Jiang, and Wei Wang. 2020. Mutation effect estimation on protein–protein interactions using deep contextualized representation learning. *NAR Genomics and Bioinformatics* 2, 2 (2020), lqaa015.
- [70] Moshé M Zloof. 1975. Query-by-example: the invocation and definition of tables and forms. In *Proceedings of the 1st International Conference on Very Large Data Bases*. 1–24.