InfoSync: Information Synchronization across Multilingual Semi-structured Tables

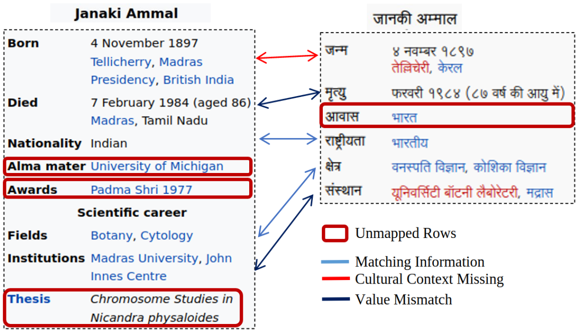
Siddharth Khincha[1], **Chelsi Jain**[2], **Vivek Gupta**[3†*], **Tushar Kataria**[3†], **Shuo Zhang**[4]
[1]IIT Guwahati, [2]CTAE, Udaipur, [3]University of Utah[‡], [4]Bloomberg,
s.khincha@iitg.ac.in, chelsiworld@gmail.com
{vgupta, tkataria}@cs.utah.edu, {szhang611}@bloomberg.net

Abstract

Information Synchronization of semi-structured data across languages is challenging. For instance, Wikipedia tables in one language should be synchronized across languages. To address this problem, we introduce a new dataset InfoSync and a two-step method for tabular synchronization. InfoSync contains 100K entity-centric tables (Wikipedia Infoboxes) across 14 languages, of which a subset (∼3.5K pairs) are manually annotated. The proposed method includes 1) *Information Alignment* to map rows and 2) *Information Update* for updating missing/outdated information for aligned tables across multilingual tables. When evaluated on InfoSync, information alignment achieves an F1 score of 87.91 (en ↔ non-en). To evaluate information updation, we perform human-assisted Wikipedia edits on Infoboxes for 603 table pairs. Our approach obtains an acceptance rate of 77.28% on Wikipedia, showing the effectiveness of the proposed method.

1 Introduction

English articles across the web are more timely updated than other languages on particular subjects. Meanwhile, culture differences, topic preferences, and editing inconsistency lead to information mismatch across multilingual data, e.g., outdated information or missing information (Jang et al., 2016; Nguyen et al., 2018). Online encyclopedia, e.g., Wikipedia, contains millions of articles that need to be updated constantly, involving expanding existing articles, modifying content such as correcting facts in sentences (Shah et al., 2019) and altering Wikipedia categories (Zhang et al., 2020b). However, more than 40% of Wikipedia's active editors are in English. At the same time, only 15% of the world population speak English as their first language. Therefore, information in languages other than English may not be as updated (Bao et al., 2012). See Figure 1 for an example of an information mismatch for the same entity across different languages. In this work, we look at synchronizing information across multilingual content.



Janaki Ammal	
Born	4 November 1897 Tellicherry, Madras Presidency, British India
Died	7 February 1984 (aged 86) Madras, Tamil Nadu
Nationality	Indian
Alma mater	University of Michigan
Awards	Padma Shri 1977
Scientific career	
Fields	Botany, Cytology
Institutions	Madras University, John Innes Centre
Thesis	*Chromosome Studies in Nicandra physaloides*

जानकी अम्माल	
जन्म	४ नवम्बर १८९७ तेल्लिचेरी, केरल
मृत्यु	फरवरी १९८४ (८७ वर्ष की आयु में)
आवास	भारत
राष्ट्रीयता	भारतीय
क्षेत्र	वनस्पति विज्ञान, कोशिका विज्ञान
संस्थान	यूनिवर्सिटी बॉटनी लैबोरेटरी, मद्रास

Figure 1: Janaki Ammal Infoboxes in English (right) and Hindi (left). Hindi Table lacks the "British Rule of India" as a cultural context. Two value mismatches (a) The Hindi table doesn't list *Died* key's state (b) Institution values differ. The Hindi table mentions "residence" while the English table doesn't. Hindi Table is missing Thesis, Awards, and Alma Mater keys. Both don't mention parents, early education, or honors.

To overcome the above-mentioned problem, we formally introduce the task of Information Synchronization for multilingual articles, which includes paragraphs, tables, lists, categories, and images. But due to its magnitude and complexity, synchronizing all of the information across different modalities on a webpage is daunting. Therefore, this work focuses on semi-structured data, a.k.a. table synchronization in a few languages, as the first step toward our mission.

We consider Infobox, a particular type of semi-structured Wikipedia tables (Zhang and Balog, 2020a), which contain entity-centric information, where we observe various information mismatches, e.g., missing rows (cf. Figure 1). One intuitive idea to address them is translation-based. However, the Infoboxes contain rows with implicit context; translating these short phrases is prone to errors and leads to ineffective synchronization (Minhas

[*] Corresponding Author
[†] Equal Contribution

Findings of the Association for Computational Linguistics: ACL 2023, pages 2536–2559
July 9-14, 2023 ©2023 Association for Computational Linguistics

et al., 2022). To systematically assess the challenge, we curate a dataset, namely InfoSync, consisting of 100K multilingual Infobox tables across 14 languages and covering 21 Wikipedia categories. ∼3.5K table pairs of English to non-English or non-English to non-English are sampled and manually synchronized.

We propose a table synchronization approach that comprises two steps: (1.) **Information Alignment:** align table rows, and (2.) **Information Update:** update missing or outdated rows across language pairs to circumvent the inconsistency. The *information alignment* component aims to align the rows in multilingual tables. The proposed method uses corpus statistics across Wikipedia, such as key and value-based similarities. The *information update* step relies on an effective rule-based approach. We manually curate nine rules: row transfer, time-based, value trends, multi-key matching, append value, high to low resource, number of row differences, and rare keys. Both tasks are evaluated on InfoSync to demonstrate their effectiveness. Apart from the automatic evaluation, we deploy an online experiment that submits the detected mismatches by our method to Wikipedia after strictly following Wikipedia editing guidelines. We monitor the number of accepted and rejected edits by Wikipedia editors to demonstrate its efficacy. All proposed edits are performed manually, in accordance with Wikipedia's editing policies and guidelines[1], rule set[2], and policies[3]. These changes were subsequently accepted by Wikipedia editors, demonstrating the efficacy of our methodology.

The contributions in this work are as follows: 1) We investigate the problem of Information Synchronization across multilingual semi-structured data, i.e., tables, and construct a large-scale dataset InfoSync; 2) We propose a two-step approach (alignment and updation) and demonstrate superiority over exiting baselines; 3) The rule-based updation system achieves excellent acceptance when utilized for human-assisted Wikipedia editing. Our InfoSync dataset and method source code are available at https://info-sync.github.io/info-sync/.

2 Motivation

2.1 Challenges in Table Synchronization

We observe the following challenges when taking Wikipedia Infoboxes as a running example. Note this is not an exhaustive list.

MI: Missing Information represents the problem where information appears in one language and is missing in others. This may be due to the fact that the table is out-of-date or to cultural, social, or demographic preferences for modification (cf. Figure 1).

OI: Outdated Information denotes that information is updated in one language but not others.

IR: Information Representation varies across languages. For example, one attribute about "parents" can be put in a single row or separate rows ("Father" and "Mother").

UI: Unnormalized Information presents cases where table attributes can be expressed differently. For example, "known for" and "major achievements" of a person represent the same attribute (i.e., paraphrase).

LV: Language Variation means that information is expressed in different variants across languages. This problem is further exaggerated by the implicit context in tables when translating. E.g., "Died" in English might be translated to "Overleden" (Pass Away) or "overlijdensplaats" (Place of Death) in Dutch due to missing context.

SV: Schema Variation denotes that the schema (template structure) varies. For example, extraction of "awards" in Musician tables can be harrowing due to dynamic on-click lists (*Full Award Lists*).

EEL: Erroneous Entity Linking is caused by mismatched linkages between table entities among multiple languages, e.g., "ABV" and "Alcohol by Volume".

2.2 Wikipedian "Biases"

Wikipedia is a global resource across over 300 languages. However, the information is skewed toward English-speaking countries (Roy et al., 2020) as English has the most significant Wikipedia covering 23% (11%) of total pages (articles). Most users' edits (76%) are also done in English Wikipedia. English Wikipedia also has the highest number of page reads (49%) and page edits (34%), followed by German (20% and 12%) and Spanish (12% and 6%), respectively. Except for the top 25 languages, the total number of active editors, pages, and edits is less than 1% (Warncke-Wang et al., 2012;

[1] https://en.wikipedia.org/wiki/Wikipedia:List_of_policies_and_guidelines

[2] https://en.wikipedia.org/wiki/Wikipedia:Simplified_ruleset

[3] https://en.m.wikipedia.org/wiki/Wikipedia:Editing_policy

Alonso and Robinson, 2016).

Multilingual Wikipedia articles evolve separately due to cultural and geographical bias (Callahan and Herring, 2011; REAGLE and RHUE, 2011; Tinati et al., 2014), which prevents information synchronization. For example, information on "Narendra Modi" (India's Prime Minister) is more likely to be better reflected in Hindi Wikipedia than in other Wikipedias. This means that in addition to the obvious fact that smaller Wikipedias can be expanded by incorporating content from larger Wikipedias, larger Wikipedias can also be augmented by incorporating information from smaller Wikipedias. Thus, information synchronization could assist Wikipedia communities by ensuring that information is consistent and of good quality across all language versions.

3 The InfoSync Dataset

To systematically assess the challenge of information synchronization and evaluate the methodologies, we aim to build a large-scale table synchronization dataset InfoSync based on entity-centric Wikipedia Infoboxes.

3.1 Table Extraction

We extract Wikipedia Infoboxes from pages appearing in multiple languages on the same date to simultaneously preserve Wikipedia's original information and potential discrepancies. These extracted tables are across 14 languages and cover 21 Wikipedia categories.

Languages Selection. We consider the following languages English(en), French(fr), German(de), Korean(ko), Russian(ru), Arabic(ar), Chinese(zh), Hindi(hi), Cebuano(ceb), Spanish(es), Swedish(sv), Dutch(nl), Turkish(tr), and Afrikaans(ak). We extracted tables across 14 languages and covered 21 diverse Wikipedia categories. In these 14 languages, four are low resource (af, ceb, hi, tr) < 6000, seven of them medium resource (ar, ko,nl, sv, zh,ru, de,es) (6000–10000), and the remaining one are high resource (en, en, fr), w.r.t. to the number of infobox total tables (see Table 1 in paper). Our choices were motivated by the following factors:- a) Cover all the continents, thus covering the majority and diverse population. Out of chosen languages, 7 (English, French, German, Spanish, Swedish, Dutch, and Turkish) are European. b). They have sufficient pages with info boxes; each entity info box is present in at least five languages, and c) an adequate number of rows (5 and above) facilitates better data extraction.

Categories. Extracted tables cover twenty-one simple, diverse, and popular topics: *Airport, Album, Animal, Athlete, Book, City, College, Company, Country, Food, Monument, Movie, Musician, Nobel, Painting, Person, Planet, Shows, and Stadiums*. We observe that *Airport* has the most number of entity tables followed by *Movie* and *Shows*, as shown in Table 10. Other extraction details are provided in Appendix A.1.

3.2 Tabular Information Mismatched

Ln	**Average Table Transfer %**		**Language Statistics**	
C1	**C1** → $\sum_L ln$	$\sum_L ln$ → **C1**	# **Tables**	**AR**
af	17.46	400.5	1575	9.91
ar	34.02	27.38	7648	13.01
ceb	42.87	134.88	3870	7.82
de	40.73	27.12	8215	7.88
en	45.85	0.32	12431	12.60
es	38.78	9.00	9920	12.59
fr	41.25	4.73	10858	10.30
hi	18.39	358.97	1724	10.91
ko	31.13	40.51	6601	9.35
nl	33.69	24.6	7837	10.46
ru	36.98	14.54	9066	11.41
sv	35.53	24.62	7985	9.89
tr	28.99	59.33	5599	10.14
zh	32.16	32.71	7140	12.43

Table 1: **Average Table Transfer**:- Column 2 shows the average number of tables missing in other languages which can be transferred from C1. Column 3 shows the average number of tables missing in C1, which we can transfer from all languages to C1. Here L is the set of all languages (ln) except source or transfer language. **Language Statistics**:- The number of tables and average rows (AR) per table across different categories for each language.

We analyze the extracted tables in the context of the synchronization problem and identify the information gap. The number of tables is biased across languages, as shown in Table 1. We observe Afrikaans, Hindi, and Cebuano have a significantly less number of tables. Similarly, the table size is biased across several languages. Dutch and Cebuano have the last rows. In addition, the number of tables across categories is uneven; refer to Table 2. Airport and Movie have the highest number of tables. Table 2 also reports the average number of rows for a category. Planet, Company, and Movie have the highest average number of rows.

When synchronizing a table from one language to another, we observe that the maximum number of tables can be transferred from English, French, and Spanish from Column 1 in Table 1. Afrikaans,

Topic	# Tables	AR	Topic	# Tables	AR
Airport	18512	9.66	Diseases	3973	6.03
Food	6184	7.93	Monument	1550	9.71
Album	5833	7.58	Medicine	2516	15.20
Animal	3304	8.27	Movie	12082	13.29
Athlete	3209	9.09	Musician	2729	9.53
Book	1550	9.99	Nobel	9522	9.84
Painting	3542	7.05	Country	3338	22.85
City	3088	14.45	Person	2252	11.87
College	1857	11.01	Planet	1233	16.80
Company	2225	13.85	Shows	5644	13.86
Stadium	6326	10.94			

Table 2: **Category Statistics** :- Number of tables in each category and average number of rows (AR) across different languages.

Hindi, and Cebuano have the least overlapping information (Column 3) with all other languages. The number of rows (Column 5) varies substantially between languages, with Spanish and Arabic having the highest number.

3.3 InfoSync Evaluation Benchmark

We construct the evaluation benchmark by manually mapping the table's pairs in two languages. The table pairs we consider can be broadly split into English ↔ Non-English and Non-English ↔ Non-English. The annotations are conducted as follows.

English ↔ Non-English: We sample 1964 table pairs, where a minimum of 50 pairs for each category and language are guaranteed. We divide the annotated dataset, ratio of $1:2$, into validation and test sets. The non-English tables are translated into English first and then compared against the English version. Furthermore, native speakers annotated 200 table pairs for English $\overset{*}{\leftrightarrow}$ Hindi and English $\overset{*}{\leftrightarrow}$ Chinese to avoid minor machine translation errors.

Non-English ↔ Non-English: We consider six non-English languages: two from each High resource (French, Russian), Medium Resource (German, Korean), and Low Resource (Hindi, Arabic), w.r.t. the number of tables in InfoSync. We sample and annotate 1589 table pairs distributed equally among these languages, where we choose an average of $\sim$ 50 tables for all pairs of languages. Both are translated into English before manually mapping them.

In addition, for more detailed analysis, we also annotate metadata around table synchronization challenges such as MI, IR, LV, OI, UI, SV, and EEL, as discussed in §2.1.

4 Table Synchronization Method

This section will explain our proposed table synchronization method for addressing missing or outdated information. This method includes two steps: information alignment and update. The former approach aims to align rows across a pair of tables, and the latter helps to update missing or outdated information. We further deploy our update process in a human-assisted Wikipedia edit framework to test the efficacy in the real world.

4.1 Information Alignment

An Infobox consists of multiple rows where each row has a key and value pair. Given a pair of tables $T_x = [..., (k_x^i, v_x^i), ...]$ and $T_y = [..., (k_y^j, v_y^j), ...]$ in two languages, table alignment aims to align all the possible pairs of rows, e.g., (k_x^i, v_x^i) and (k_y^j, v_y^j) refer to the same information and should be aligned. We propose a method that consists of five modules, each of which relaxes matching requirements in order to create additional alignments.

M1. Corpus-based. The pair of rows (k_x, v_y) in T_x and (k_y, v_y) in T_y are supposed to be aligned if $cosine(em(tr_x^{en}(k_x)), em(tr_y^{en}(k_y))) > \theta_1$, where em is the embedding, θ_1 is the threshold, and $tr_y^{en}()$ denotes the English translation of k if k is not in English. In order to achieve accurate key translations, we adopt a majority voting approach, considering multiple translations of the same key from different category tables. We consider the key's values and categories as additional context for better translation during the voting process. To simplify the voting procedure, we pre-compute mappings by selecting only the most frequent keys for each category across all languages.

M2. Key-only. This module attempts to align the unaligned pairs in module 1. Using their English translation, it first computes cosine similarity for all possible key pairs. k^x will be aligned to k^y only if they are mutually most similar key and the similarity is above a certain threshold θ_2. This is similar to maximum bipartite matching, treating similarity scores as edge weights followed by threshold-based pruning. And it ensures we are capturing the highest similarity mapping from both language directions. Note that here we use only keys as the text for similarity computation.

M3. Key value bidirectional. This module is similar to step 2, except it uses the entire table row for computing similarities, i.e., key + value, using threshold θ_3.

M4. Key value unidirectional. This module further relaxes the bidirectional mapping constraint in step 3, i.e., thus removing the requirement of the highest similarity score matching from both sides. We shift to unidirectional matching between row pairs, i.e., consider the highest similarity in either direction. However, this may result in adding spurious alignments. To avoid this, we have a higher threshold (θ_4) than the prior step.

M5. Multi-key. Previous modules only take the most similar key for alignment if exceeding the threshold. In this module, we further relax the constraint to select multiple keys (maximum two), given exceeding a threshold (θ_5). Multi-key mapping is sparse, but the above procedure will lead to dense mapping. To avoid this, we introduce a *soft constraint* for value-combination alignment, where multi-key values are merged. We consider valid multi-key alignment when the merge value-combination similarity score exceeds that of the most similar key.

The thresholds of five modules are tuned in the sequence as stated above.

4.2 Information Updation

Information modification includes *Row Append* (adding missing rows), *Row Update* (replacing or adding values), and *Merge* Rows. We propose a rule-based heuristic approach for information updates. The rules are in form of logical expression ($\forall_{(\mathrm{R}_{T_x},\mathrm{R}_{T_y})}$ L $\mapsto$ R) applied on infobox tables, where, R_{T_x} and R_{T_y} represent table rows for language x and y respectively. These rules are applied sequentially according to their priority rank (P.R.). Rules explanations are described below.

R1. Row Transfer. Following the logistic rule of

$$\forall_{(\mathrm{R}_{T_x},\mathrm{R}_{T_y})} \mathrm{Al}_{T_x}^{T_y}(\mathrm{R}_{T_x};\mathrm{R}_{T_y}) = 0$$

$$\mapsto T_y \cup tr_x^y(\mathrm{R}_{T_x}) \bigwedge \mathrm{Al}_{T_x}^{T_y}(\mathrm{R}_{T_x}; tr_x^y(\mathrm{R}_{T_x})) = 1$$

, where $\mathrm{Al}_{T_x}^{T_y}(.;.)$ represents the alignment mapping between two tables T_y and T_x. Unaligned rows are transferred from one table to another.

R2. Multi-Match. We update the table by removing multi-alignments and replacing them with merged information to handle multikey alignments.

R3. Time-based. We update aligned values using the latest timestamp.

R4. Trends (positive/negative). This update applies to cases where the value is highly likely to follow a monotonic pattern (increasing or decreasing) w.r.t. time, e.g., athlete career statistics. The authors curated the positive/negative trend lists.

R5. Append Values. Additional value information from an up-to-date row is appended to the outdated row.

R6. HR to LR. This rule transfers information from high to low resource language to update outdated information.

R7. #Rows. This rule transfers information from bigger (more rows) to smaller (fewer rows) tables.

R8. Rare Keys (Non Popular). We update information from the table where non-popular keys are likely to be added recently to the outdated table. The authors also curate non-popular keys.

Detailed formulation of logical rules and their priority ranking are listed in Table 3. Figure 3 in Appendix shows an example of table update.

Human-assisted Wikipedia Infobox Edits: We apply the above rules to assist humans in updating Wikipedia infoboxes. Following Wikipedia edit guidelines[4], rule set[5], and policies[6], we append our update request with a description to provide evidence, which contains (a) up-to-date entity page URL in the source language, (b) exact table rows information, the source language, and the details of the changes, (c) and one additional citation discovered by the editor for extra validation. [7] We further update beyond our heuristic-based rules but are aligned through our information alignment method.

5 Experiments

Our experiments assess the efficacy of our proposed two-stage approach by investigating the following questions.

- What is the efficacy of the unsupervised multilingual method for table alignment? (§5.2)

- How significant are the different modules of the alignment algorithm? (§5.2 and §A.6)

- Does the rule-based updating approach effective for information synchronization? (§5.3)

- Can the two-step approach assist humans in updating Wikipedia Infoboxes? (§5.3)

[4] https://en.wikipedia.org/wiki/Wikipedia:List_of_policies_and_guidelines

[5] https://en.wikipedia.org/wiki/Wikipedia:Simplified_ruleset

[6] https://en.m.wikipedia.org/wiki/Wikipedia:Editing_policy

[7] We use a search engine such as Google, Bing, you.com, perplexity.ai, find, etc. for additional citation.

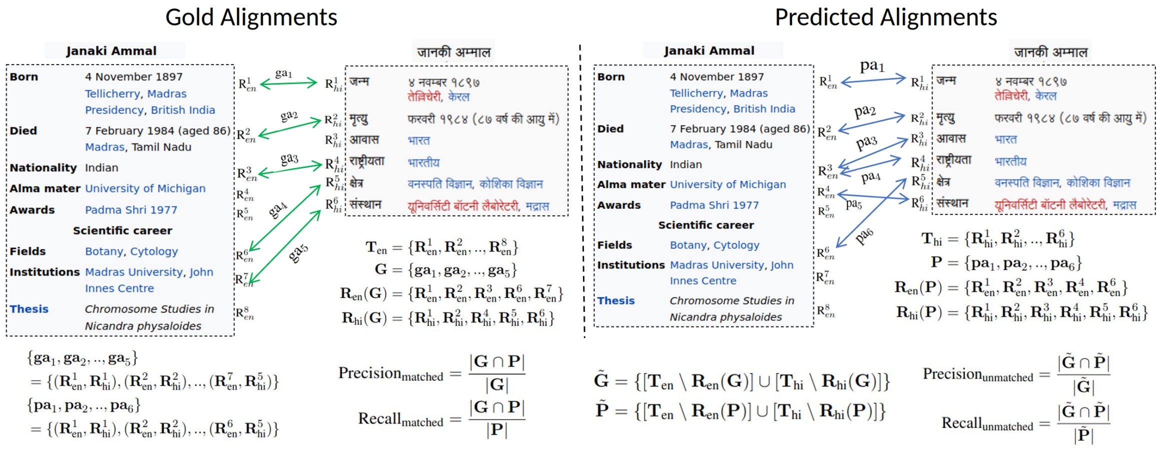

Gold Alignments
Predicted Alignments
Janaki Ammal
जानकी अम्माल
Born
4 November 1897
Tellicherry, Madras Presidency, British India
Died
7 February 1984 (aged 86)
Madras, Tamil Nadu
Nationality
Indian
Alma mater
University of Michigan
Awards
Padma Shri 1977
Scientific career
Fields
Botany, Cytology
Institutions
Madras University, John Innes Centre
Thesis
Chromosome Studies in Nicandra physaloides
जन्म
४ नवम्बर १८९७
तेलिचेरी, केरल
मृत्यु
फरवरी १९८४ (८७ वर्ष की आयु में)
आवास
भारत
राष्ट्रीयता
भारतीय
क्षेत्र
वनस्पति विज्ञान, कोशिका विज्ञान
संस्थान
यूनिवर्सिटी बॉटनी लैबोरेटरी, मद्रास
ga_1 ga_2 ga_3 ga_4 ga_5
pa_1 pa_2 pa_3 pa_4 pa_5 pa_6
$\mathbf{T}_{\mathrm{en}} = \{\mathbf{R}^1_{\mathrm{en}}, \mathbf{R}^2_{\mathrm{en}}, .., \mathbf{R}^8_{\mathrm{en}}\}$
$\mathbf{G} = \{\mathrm{ga}_1, \mathrm{ga}_2, .., \mathrm{ga}_5\}$
$\mathbf{R}_{\mathrm{en}}(\mathbf{G}) = \{\mathbf{R}^1_{\mathrm{en}}, \mathbf{R}^2_{\mathrm{en}}, \mathbf{R}^3_{\mathrm{en}}, \mathbf{R}^6_{\mathrm{en}}, \mathbf{R}^7_{\mathrm{en}}\}$
$\mathbf{R}_{\mathrm{hi}}(\mathbf{G}) = \{\mathbf{R}^1_{\mathrm{hi}}, \mathbf{R}^2_{\mathrm{hi}}, \mathbf{R}^4_{\mathrm{hi}}, \mathbf{R}^5_{\mathrm{hi}}, \mathbf{R}^6_{\mathrm{hi}}\}$
$\mathbf{T}_{\mathrm{hi}} = \{\mathbf{R}^1_{\mathrm{hi}}, \mathbf{R}^2_{\mathrm{hi}}, .., \mathbf{R}^6_{\mathrm{hi}}\}$
$\mathbf{P} = \{\mathrm{pa}_1, \mathrm{pa}_2, .., \mathrm{pa}_6\}$
$\mathbf{R}_{\mathrm{en}}(\mathbf{P}) = \{\mathbf{R}^1_{\mathrm{en}}, \mathbf{R}^2_{\mathrm{en}}, \mathbf{R}^3_{\mathrm{en}}, \mathbf{R}^4_{\mathrm{en}}, \mathbf{R}^6_{\mathrm{en}}\}$
$\mathbf{R}_{\mathrm{hi}}(\mathbf{P}) = \{\mathbf{R}^1_{\mathrm{hi}}, \mathbf{R}^2_{\mathrm{hi}}, \mathbf{R}^3_{\mathrm{hi}}, \mathbf{R}^4_{\mathrm{hi}}, \mathbf{R}^5_{\mathrm{hi}}, \mathbf{R}^6_{\mathrm{hi}}\}$
$\{\mathrm{ga}_1, \mathrm{ga}_2, .., \mathrm{ga}_5\} = \{(\mathbf{R}^1_{\mathrm{en}}, \mathbf{R}^1_{\mathrm{hi}}), (\mathbf{R}^2_{\mathrm{en}}, \mathbf{R}^2_{\mathrm{hi}}), .., (\mathbf{R}^7_{\mathrm{en}}, \mathbf{R}^5_{\mathrm{hi}})\}$
$\{\mathrm{pa}_1, \mathrm{pa}_2, .., \mathrm{pa}_6\} = \{(\mathbf{R}^1_{\mathrm{en}}, \mathbf{R}^1_{\mathrm{hi}}), (\mathbf{R}^2_{\mathrm{en}}, \mathbf{R}^2_{\mathrm{hi}}), .., (\mathbf{R}^6_{\mathrm{en}}, \mathbf{R}^5_{\mathrm{hi}})\}$
$\mathrm{Precision}_{\mathrm{matched}} = \frac{|\mathbf{G} \cap \mathbf{P}|}{|\mathbf{G}|}$
$\mathrm{Recall}_{\mathrm{matched}} = \frac{|\mathbf{G} \cap \mathbf{P}|}{|\mathbf{P}|}$
$\tilde{\mathbf{G}} = \{[\mathbf{T}_{\mathrm{en}} \setminus \mathbf{R}_{\mathrm{en}}(\mathbf{G})] \cup [\mathbf{T}_{\mathrm{hi}} \setminus \mathbf{R}_{\mathrm{hi}}(\mathbf{G})]\}$
$\tilde{\mathbf{P}} = \{[\mathbf{T}_{\mathrm{en}} \setminus \mathbf{R}_{\mathrm{en}}(\mathbf{P})] \cup [\mathbf{T}_{\mathrm{hi}} \setminus \mathbf{R}_{\mathrm{hi}}(\mathbf{P})]\}$
$\mathrm{Precision}_{\mathrm{unmatched}} = \frac{|\tilde{\mathbf{G}} \cap \tilde{\mathbf{P}}|}{|\tilde{\mathbf{G}}|}$
$\mathrm{Recall}_{\mathrm{unmatched}} = \frac{|\tilde{\mathbf{G}} \cap \tilde{\mathbf{P}}|}{|\tilde{\mathbf{P}}|}$

Figure 2: Explanation of Alignment Performance Metrics: $\mathbf{T_{en}}$ and $\mathbf{T_{hi}}$ are a collection of all rows in the English and Hindi tables, respectively. $\mathbf{R}^n_x$ represents the n^{th} row in the language table. $\mathbf{R}_x(\mathbf{X})$ retrieves all rows in the language x using mapping $\mathbf{X}$. $|.|$ represents the set's cardinality. Every alignment is saved as a tuple in form $(\mathbf{R}^m_x, \mathbf{R}^n_y)$. $\mathbf{G}$ is a collection of all gold (human) alignments. $\mathbf{P}$ is a collection of predicted alignments (can see there are mistakes in the alignment.

5.1 Experimental Setup

Baselines Models. We compare our approach with LaBSE (Feng et al., 2022), and SimCSE (Gao et al., 2021), multilingual sentence transformers embeddings (Reimers and Gurevych, 2020a) in which we include mBERT (case2) with mean pooling (mp) (Reimers and Gurevych, 2020b), and its distill versions (distill mBERT) (Sanh et al., 2019) all in base form. We also compared with XLM-RoBERTa (XLM-R) (Conneau et al., 2019) with mean pooling, and its distill version (Reimers and Gurevych, 2019a) trained via MPNet-based teacher model (MPNet) (Song et al., 2020). For all baseline implementation, we use the Hugging Face transformers (Wolf et al., 2020) and sentence transformers (Reimers and Gurevych, 2019a) library for the multilingual models' implementation.

Hyper-parameter Tuning. For our method, we embed translated English keys and values using MPNet model (Reimers and Gurevych, 2019b). We tune the threshold hyper-parameters using the validation set, $\frac{1}{3}$ of the total annotated set. We sequentially tune the hyper-parameters thresholds (θ_1 to θ_5) in modules training order. Optimal threshold after tuning are θ_1 = (0.8,0.8); θ_2 = (0.64, 0.6); θ_3= θ_3 = (0.54, 0.54); θ_4 =(0.9, 0.54); θ_5 = (0.88, 0.96) for $\mathrm{T}_{en} \leftarrow \mathrm{T}_x$ and $\mathrm{T}_x \leftarrow \mathrm{T}_y$ respectively. We retain the default setting for other models' specific hyperparameters.

Information Alignment. We consider English as our reference language for alignment. Specifically, we translate all multilingual tables to English using an effective table translation approach of XInfoTabS (Minhas et al., 2022). Then, we apply incremental modules as discussed in §4.1. We tune independently on the validation set for Non-English ↔ Non-English and English ↔ Non-English.

The method is assessed on two sets of metric (a.) matched score: measure the F1-score between ground truth matched row and predicted alignment, and (b.) unmatched score: measure the F1-score between independent (unmatched) rows in ground truth with predicted unaligned rows. See Figure 2 for the explanations of these metrics.

Information Updation. We apply the heuristic-based approach and deploy the predicted updates for human-assisted edits on Wikipedia Infoboxes. 532 table pairs are edited distributed among $T_{en} \rightarrow T_x$, $T_x \rightarrow T_y$, and $T_x \rightarrow T_{en}$, where x and y are non-English languages.

5.2 Information Alignment

Algorithm Efficacy. Table 4 reports the matched and unmatched scores. For match scores, we observe that the corpus-based module achieves an F1 score exceeding 50 for all language pairs. Using a key-only module boosts the performance by about 5-15 points. Taking the whole row context (key-value pair) with strict constraints on bidirectional mapping, i.e., two-way similarity, improves performance substantially (more than 16 points). Further relaxing the bi-direction constraint to unidirectional matching (one-way similarity), we improve our results marginally with less than 0.5 performance points. Thus relaxation of the bi-

P.R.	Rule Name	Logical Rule $\forall_{(\mathbf{R}_{T_x}, \mathbf{R}_{T_y})}$ **L** $\mapsto$ **R**	Update Type
1	Row Transfer	$\forall_{(\mathrm{R}_{T_x}, \mathrm{R}_{T_y})} \mathrm{Al}_{T_x}^{T_y}(\mathrm{R}_{T_x}; \mathrm{R}_{T_y}) = 0$ $\mapsto T_y \cup tr_x^y(\mathrm{R}_{T_x}) \bigwedge \mathrm{Al}_{T_x}^{T_y}(\mathrm{R}_{T_x}; tr_x^y(\mathrm{R}_{T_x})) = 1$	Row Addition
2	Multi-Match	$\forall_{(\mathrm{R}_{T_x}, \mathrm{R}_{T_y})} (\sum_{\mathrm{R}_{T_y}} \mathrm{Al}_{T_x}^{T_y}(\mathrm{R}_{T_x}; \mathrm{R}_{T_y})) > 1$ $\mapsto \{T_y \setminus \cup_{(\forall_{\mathrm{R}_{T_y}} \mathrm{Al}_{T_x}^{T_y}(\mathrm{R}_{T_x}; \mathrm{R}_{T_y}) = 1)} \mathrm{R}_{T_y}\} \bigcup tr_x^y(\mathrm{R}_{T_x}) \bigwedge \mathrm{Al}_{T_x}^{T_y}(\mathrm{R}_{T_x}; tr_x^y(\mathrm{R}_{T_x})) = 1$	Row Delete
3	Time-based	$\forall_{(\mathrm{R}_{T_x}, \mathrm{R}_{T_y})} \mathrm{Al}_{T_x}^{T_y}(\mathrm{R}_{T_x}; \mathrm{R}_{T_y}) = 1 \bigwedge (\mathrm{isTime}(\mathrm{R}_{T_x}, \mathrm{R}_{T_y}) = 1)$ $\bigwedge (\mathrm{exTime}(\mathrm{R}_{T_x}) > \mathrm{exTime}(\mathrm{R}_{T_y})) \mapsto \mathrm{R}_{T_y} \leftarrow tr_x^y(\mathrm{R}_{T_x})$	Value Substitute
4	Positive Trend or Negative Trend	$\forall_{(\mathrm{R}_{T_x}, \mathrm{R}_{T_y}, \mathrm{PosTrend})} \mathrm{Al}_{T_x}^{T_y}(\mathrm{R}_{T_x}; \mathrm{R}_{T_y}) = 1 \bigwedge \mathrm{exKey}(\mathrm{R}_{T_x}) \in \mathrm{PosTrend}$ $\bigwedge \mathrm{R}_{T_x} > \mathrm{R}_{T_y} \mapsto \mathrm{R}_{T_y} \leftarrow \mathrm{R}_{T_x}$ $\forall_{(\mathrm{R}_{T_x}, \mathrm{R}_{T_y}, \mathrm{NegTrend})} \mathrm{Al}_{T_x}^{T_y}(\mathrm{R}_{T_x}; \mathrm{R}_{T_y}) = 1 \bigwedge \mathrm{exKey}(\mathrm{R}_{T_x}) \in \mathrm{NegTrend}$ $\bigwedge \mathrm{R}_{T_x} < \mathrm{R}_{T_y} \mapsto \mathrm{R}_{T_y} \leftarrow \mathrm{R}_{T_x}$	Value Substitute Value Substitute
5	Append Value	$\mathrm{R}_{T_x} = \mathrm{V} \bigwedge \forall_{(\mathrm{R}_{T_x}, \mathrm{R}_{T_y})} \mathrm{Al}_{T_x}^{T_y}(\mathrm{R}_{T_x}; \mathrm{R}_{T_y}) = 1 \bigwedge \lvert \mathrm{R}_{T_x}[k] \rvert > \lvert \mathrm{R}_{T_y}[k] \rvert$ $\mapsto \forall_{(v \in \mathrm{R}_{T_x}[k] \wedge \notin tr_x^y(\mathrm{R}_{T_x}[k]))} \mathrm{R}_{T_y} \leftarrow \mathrm{R}_{T_y} \cup tr_x^y(v)$	Value Addition
6	HR to LR	$(T_x, T_y) \in (HR, LR) \bigwedge \forall_{(\mathrm{R}_{T_x}, \mathrm{R}_{T_y})} \mathrm{Al}_{T_x}^{T_y}(\mathrm{R}_{T_x}; \mathrm{R}_{T_y}) = 1$ $\bigwedge tr_x^{en}(\mathrm{R}_{T_x}) \neq tr_y^{en}(\mathrm{R}_{T_y}) \mapsto \mathrm{R}_{T_y} \leftarrow tr_x^y(\mathrm{R}_{T_x})$	Value Substitute
7	# Rows	$\lvert T_x \rvert >> \lvert T_y \rvert \bigwedge \forall_{(\mathrm{R}_{T_x}, \mathrm{R}_{T_y})} \mathrm{Al}_{T_x}^{T_y}(\mathrm{R}_{T_x}; \mathrm{R}_{T_y}) = 1 \bigwedge tr_x^{en}(\mathrm{R}_{T_x}) \neq tr_y^{en}(\mathrm{R}_{T_y})$ $\mapsto \mathrm{R}_{T_y} \leftarrow tr_x^y(\mathrm{R}_{T_x})$	Value Substitute
8	Rare Keys	$\forall_{(\mathrm{R}_{T_x}, \mathrm{R}_{T_y}, \mathrm{RarKeys})} \mathrm{Al}_{T_x}^{T_y}(\mathrm{R}_{T_x}; \mathrm{R}_{T_y}) = 1 \bigwedge \mathrm{tr}_x^{en}(\mathrm{R}_{t_x}) \neq \mathrm{tr}_y^{en}(\mathrm{R}_{t_y})$ $\bigwedge \forall_{(\mathrm{R}_{T_x}, \mathrm{R}_{T_y}} \lvert \mathrm{exKey}(\mathrm{R}_{T_x}) \in \mathrm{RarKey} \rvert > \lvert \mathrm{exKey}(\mathrm{R}_{T_y}) \in \mathrm{RarKey} \rvert \mapsto \mathrm{R}_{T_y} \leftarrow \mathrm{R}_{T_x}$	Value Substitute

Table 3: **Logical Rules for Information Updation**. Notation:- T_z represents a table in language z, R_{T_z} represents a row of the table. In $\mathrm{R}_{T_z}[k] = v$, k,v represent key and value pair. For $\mathrm{R}_{T_z}[k] = V$, V denotes value list mapped to a key k. $\mathrm{Al}_{T_x}^{T_y}(.;.)$ represents the alignment mapping between two tables T_y and T_x. Translation between two languages(p and q) is represented by $tr_q^p(.)$. *exKey* extract key from a table row. *isTime* is true if the row has time entry. *exTime* extract time from table row. *PosTrend/NegTrend* represent list of keys whose value always increase or decrease with time. *RarKey* represent set of keys are least frequent in the corpora.

Method	Match				UnMatch			
	$T_{en} \leftrightarrow T_x$	$T_x \leftrightarrow T_y$	$T_{en} \overset{*}{\leftrightarrow} T_{hi}$	$T_{en} \overset{*}{\leftrightarrow} T_{zh}$	$T_{en} \leftrightarrow T_x$	$T_x \leftrightarrow T_y$	$T_{en} \overset{*}{\leftrightarrow} T_{hi}$	$T_{en} \overset{*}{\leftrightarrow} T_{zh}$
SimCSE	75.78	68.46	77.93	80.47	79.11	76.3	73.31	74.91
LaBSE	85.25	78.44	88.98	89.1	87.03	81.7	88.98	85.06
mBERT-mp	80.98	73.74	82.9	86.73	82.68	80.22	76.73	81.85
XLM-R	83.38	75.02	86.85	88.08	85.42	80.65	83.14	83.1
MPNet	82.85	78.63	86.08	87.58	84.2	83.45	83.14	83.76
distill mBERT	84.55	77.45	87.64	88.7	86.3	82.28	83.14	84.3
Our Approach								
Corpus-based	61.86	56.74	57.34	69.33	70.51	71.73	54.01	63.11
+ Key Only	70.41	62.14	73.4	74.67	73.85	73.52	62.49	66.23
+ Key-Val-Bi	87.71	84.2	90.07	93.04	89.51	85.52	85.06	89.2
+ Key-Val-Uni	87.89	84.33	90.34	93.12	89.52	85.42	85.16	88.62
+ Multi-Key	**87.91**	**84.36**	**90.14**	**92.8**	**89.3**	**85.46**	**84.98**	**88.15**

Table 4: **Matched and UnMatch Score :** F1-Score for all test sets of INFOSYNC.

direction mapping constraint doesn't lead to significantly better alignments. The multi-key module, which considers one-to-many alignments, further improves the accuracy marginally. The reason for the marginal improvements is very few instances of one-to-many mappings.

For unmatch scores, we see similar results to match scores. The only significant difference is in key-only performance, where we observe a 0.5x performance improvement compared to match scores. We also analyze the precision-recall in Tables 17, 18, 19 and 20 of Appendix §A.3. We observe that the precision reduces and recall increases for match scores with module addition, whereas the reverse is true for unmatch scores. The number of alignments increases as we add more modules with relaxed constraints. This increases the number of incorrect alignments reducing the precision but increasing the recall. [8] Similarly, we can note the accuracy of unaligned rows increases because more incorrect alignments are added with relaxed constraints. We also report each module coverage in Appendix A.4. The performance of our proposed approach grouped by languages, category, and rows keys are detailed in Appendix A.5.

Error Analysis. Error analysis (cf §2.1) for

[8] There are more incorrect alignments ${}^{N}C_2$ compared to correct alignments which is $O(n)$.

Method	$T_{en} \leftrightarrow T_x$						$T_x \leftrightarrow T_y$					
	OI	**IR**	**SV**	**LV**	**UI**	**EL**	**OI**	**IR**	**SV**	**LV**	**UI**	**EL**
w/o Align	298	286	22	158	388	118	245	226	33	146	486	148
Corpus-based	81	284	15	141	337	74	108	218	26	102	366	109
+Key Only	110	281	7	120	262	48	77	212	19	94	284	97
+Key-Val-Bi	75	232.33	6	35	108	8	44	197	15	28	60	18
+Key-Val-Uni	74	206.67	6	30	99	8	43	188	15	28	59	17
+Multi-Key	**74**	**179.67**	**6**	**30**	**99**	**8**	**43**	**180.33**	**15**	**28**	**59**	**17**

Table 5: **Error Analysis for Matched Score :** $T_{en} \leftrightarrow T_x$ and $T_x \leftrightarrow T_y$.

Method	$T_{en} \leftrightarrow T_x$	$T_x \leftrightarrow T_y$
Corpus-based	**157**	**245**
+Key Only	422	343
+Key-Value-Bi	526	399
+Key-Value-Uni	572	415
+Multi-Key	619	437

Table 6: **Error Analysis for UnMatch Score :** Total Unaligned mistakes for $T_{en} \leftrightarrow T_x$ and $T_x \leftrightarrow T_y$.

matched and unmatched are reported in Table 5 and 6, respectively. Our proposed method works sequentially, relaxing constraints, and the number of falsely aligned rows increases with module addition (cf. Table 6). Different modules contribute unequally to unaligned mistakes, (25%, 56%) of the mistakes come from corpus-based module, (39%, 22%) from Key Only Module, (17%, 35%) from Key-Value-Bidirectional module, (7%, 4%) from Key-Value-uni-directional module, and (7.6%, 5%) from multi-key alignment module, for $T_{en} \leftrightarrow T_x$ and $T_x \leftrightarrow T_y$ respectively. The corpus-based module is worst performing in $T_x \leftrightarrow T_y$ because of difficulty in multilingual mapping. The key-only module is the worst performing in $T_{en} \leftrightarrow T_x$ because it's the first relaxation in the algorithm. Further analysis of the error cases is in Appendix (§A.7).

5.3 Information Updation

Rules	**Gold**			**Predicted**	
	$T_{en} \rightarrow T_x$	$T_x \rightarrow T_y$	**Live Set**	$T_{en} \rightarrow T_x$	$T_x \rightarrow T_y$
R1	20320	18055	4213	21246	17675
R2	648	502	207	1395	1852
R3	546	399	75	443	347
R4	142	151	4	120	147
R5	3507	2116	784	3193	1960
R6	5237	3047	332	5062	2891
R7	2748	1899	990	2732	1855
R8	25	77	5	29	82
Al	14967	9715	2851	14864	10657

Table 7: **Updates on Test Corpora:** Count of the number of updates done by different rules listed in §4.2.*Al* is the number of Alignments. R1-R8 are the rules listed in the same sequential manner as listed in §4.2.

Table 7 reports the results of different updation types of rules explained in §4.2. We observe that the row addition rule accounts for the most updated, ∼64% of total updates for gold and predicted aligned table pairs. The flow of information from high resource to low resource accounts for ∼13% of the remaining updates, whereas a high number of rows too low adds another 8% of the updates. ∼9% of the updates are done by the value updates rule. All the other rules combined give 8% of the remaining suggested updates. From the above results, most information gaps can be resolved by row transfer. The magnitude of rules like value updates and multi-key shows that table information needs to be synchronized regularly. Examples of edited infoboxes using the proposed algorithm are shown in Appendix Figures 4 and 5.

Type	**Total**	**Accept**	**Reject**
Row Transfer	461	368(79.82%)	93(20.17%)
Value Substitution	70	52(74.28%)	18(25.72%)
Append Value	72	46(63.88%)	26(36.12%)
Total	603	466 (77.28%)	136(22.72%)

Table 8: **Analysis of Human-Assisted Updates:** Accept/Reject rate of different types of edits for human-assisted Wikipedia infobox updates.

Ln Pairs	**Total**	**Accept**	**Reject**
$T_{en} \rightarrow T_x$	204	161(78.92%)	43(21.07%)
$T_x \rightarrow T_y$	216	169(78.25%)	47(21.75%)
$T_x \rightarrow T_{en}$	183	136(74.31%)	47(25.68%)
Total	603	466(77.28%)	137(22.71%)

Table 9: **Human-Assisted Wikipedia infobox updates:** Accept/Reject rate for different flows of information.

Table 8 reports a similar analysis for human-assisted Wikipedia infobox edits. We also report Wikipedia editors' accept/reject rate for the above-deployed system in Table 9. We obtained an acceptance rate of 77.28% (as of May 2023), with the highest performance obtained when information flows across non-English languages. The lowest performance is obtained when the information flows from non-English to an English info box. This highlights that our two-step procedure is effective in a real-world scenario. Examples of live updates are shown in Appendix Figures 6 and 7.

6 Related Works

Information Alignment. Multilingual Table attribute alignment has been previously addressed via supervised (Adar et al., 2009; Zhang et al., 2017; Ta and Anutariya, 2015) and unsupervised methods (Bouma et al., 2009; Nguyen et al., 2011). Supervised methods trained classifiers on features extracted from multilingual tables. These features include cross-language links, text similarity, and schema features. Unsupervised methods made use of corpus statistics and template/schema matching for alignments. Other techniques by Jang et al. (2016); Nguyen et al. (2018) focus on using external knowledge graphs such as DBpedia for the updation of Infoboxes or vice versa. In their experiments, most of these methods use less than three languages, and machine translation is rarely used. Additionally, we don't require manual feature curation for strong supervision. We study the problem more thoroughly with grouped analysis along languages, categories, and keys direction. The works closest to our approach are Nguyen et al. (2011); Rinser et al. (2013), both of which use cross-language hyperlinks for feature or entity matching. Nguyen et al. (2011) uses translations before calculating text similarity. Utilizing cross-language links can provide a robust alignment supervision signal. In contrast to our approach, we do not use external knowledge or cross-language links for alignments. This additional information is rarely available for languages other than English.

Information Updation. Prior work for information updates (Iv et al., 2022; Spangher et al., 2022; Panthaplackel et al., 2022; Zhang et al., 2020b,d) covers Wikipedia or news articles than semi-structured data like tables. Spangher et al. (2022) studies the problem of updating multilingual news articles across different languages over 15 years. They classify the edits as addition, deletion, updates, and retraction. These were the primary intuitions behind our challenge classified in §2.1. Iv et al. (2022) focused on automating article updates with new facts using large language models. Panthaplackel et al. (2022) focused on generating updated headlines when presented with new information. Some prior works also focus on the automatic classification of edits on Wikipedia for content moderation and review (Sarkar et al., 2019; Daxenberger and Gurevych, 2013). Evening modeling editor's behavior for gauging collaborative editing and development of Wikipedia pages has been studied (Jaidka et al., 2021; Yang et al., 2017). Other related works include automated sentence updation based on information arrival (Shah et al., 2020; Dwivedi-Yu et al., 2022). None of these works focus on tables, especially Wikipedia Infoboxes. Also, they fail to address multilingual aspects of information updation.

7 Conclusion and Future Work

Information synchronization is a common issue for semi-structured data across languages. Taking Wikipedia Infoboxes as our case study, we created INFOSYNC and proposed a two-step procedure that consists of alignment and updation. The alignment method outperforms baseline approaches with an F1-score greater than 85; the rule-based method received a 77.28 percent approval rate when suggesting updates to Wikipedia.

We identify the following future directions. (a) *Beyond Infobox Synchronization.* While our technique is relatively broad, it is optimized for Wikipedia Infoboxes. We want to test whether the strategy applies to technical, scientific, legal, and medical domain tables (Wang et al., 2013; Gottschalk and Demidova, 2017). It will also be intriguing to widen the updating rules to include social, economic, and cultural aspects. (b) *Beyond Pairwise Alignment.* Currently, independent language pairs are considered for (bi) alignment. However, multiple languages can be utilized jointly for (multi) alignment. (c) *Beyond Pairwise Updates.* Similar to (multi) alignment, one can jointly update all language variants simultaneously. This can be done in two ways: (1.) *With English as pivot language :* To update across all languages. Here, English act as a central server with message passing. (2.) *Round-Robin Fashion:* where pairwise language updates between language pairs are transferred in a round-robin ring across all language pairs. In every update, we selected a leader similar to a leader election in distributed systems. (d) *Joint Alignment and Updation.* Even while our current approach is accurate, it employs a two-step process for synchronization, namely alignment followed by updating. We want to create rapid approaches aligning and updating in a single step. (e) *Text for Updation*: Our method doesn't consider Wikipedia articles for updating tables (Lange et al., 2010; Sáez and Hogan, 2018; Sultana et al., 2012).

Limitations

We only consider 14 languages and 21 categories, whereas Wikipedia has pages in more than 300 languages and 200 broad categories. Increasing the scale and diversity will further improve method generalization. Our proposed method relies on the good multilingual translation of key and value from table pairs. Although we use key, value, and category together for better context, enhancement in table translation (Minhas et al., 2022) will benefit our approach. Because our rule-based system requires manual intervention, it has automation limits. Upgrading to completely automated methods based on a large language model may be advantageous. We are only considering updates for semi-structured tables. However, updating other page elements, such as images and article text, could also be considered. Although a direct expansion of our method to a multi-modal setting is complex (Suzuki et al., 2012).

Ethics Statement

We aimed to create a balanced, bias-free dataset regarding demographic and socioeconomic factors. We picked a wide range of languages, even those with limited resources, and we also ensured that the categories were diversified. Humans curate the majority of information on Wikipedia. Using unrestricted automated tools for edits might result in biased information. For this reason, we adhere to the "human in the loop" methodology (Smith et al., 2020) for editing Wikipedia. Additionally, we follow Wikipedia editing guidelines[9], rule set[10], and policies[11] for all manual edits. Therefore, we ask the community to use our method only as a recommendation tool for revising Wikipedia. As a result, we ask that the community utilize InfoSync strictly for scientific and non-commercial purposes from this point forward.

Acknowledgements

We thank members of the Utah NLP group for their valuable insights and suggestions at various stages of the project, and reviewers for their helpful comments. Additionally, we appreciate the inputs provided by Vivek Srikumar and Ellen Riloff. Vivek Gupta acknowledges support from Bloomberg's Data Science Ph.D. Fellowship.

References

Faheem Abbas, Muhammad Malik, Muhammad Rashid, and Rizwan Zafar. 2016. Wikiqa — a question answering system on wikipedia using freebase, dbpedia and infobox. pages 185–193.

Eytan Adar, Michael Skinner, and Daniel S. Weld. 2009. Information arbitrage across multi-lingual wikipedia. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, WSDM '09, page 94–103, New York, NY, USA. Association for Computing Machinery.

Elisa Alonso and Bryan J. Robinson. 2016. Exploring translators' expectations of wikipedia: A qualitative review. *Procedia - Social and Behavioral Sciences*, 231:114–121. International Conference; Meaning in Translation: Illusion of Precision, MTIP2016, 11-13 May 2016, Riga, Latvia.

Gilbert Badaro and Paolo Papotti. 2022. Transformers for tabular data representation: A tutorial on models and applications. *Proc. VLDB Endow.*, 15(12):3746–3749.

Patti Bao, Brent Hecht, Samuel Carton, Mahmood Quaderi, Michael Horn, and Darren Gergle. 2012. Omnipedia: Bridging the wikipedia language gap. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 1075–1084, New York, NY, USA. Association for Computing Machinery.

Gosse Bouma, Sergio Duarte, and Zahurul Islam. 2009. Cross-lingual alignment and completion of Wikipedia templates. In *Proceedings of the Third International Workshop on Cross Lingual Information Access: Addressing the Information Need of Multilingual Societies (CLIAWS3)*, pages 21–29, Boulder, Colorado. Association for Computational Linguistics.

Ewa S. Callahan and Susan C. Herring. 2011. Cultural bias in wikipedia content on famous persons. *J. Am. Soc. Inf. Sci. Technol.*, 62(10):1899–1915.

Qianglong Chen, Feng Ji, Xiangji Zeng, Feng-Lin Li, Ji Zhang, Haiqing Chen, and Yin Zhang. 2021a. KACE: Generating knowledge aware contrastive explanations for natural language inference. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2516–2527, Online. Association for Computational Linguistics.

Wenhu Chen, Ming-Wei Chang, Eva Schlinger, William Yang Wang, and William W. Cohen. 2021b. Open question answering over tables and text. In *International Conference on Learning Representations*.

[9] https://en.wikipedia.org/wiki/Wikipedia:List_of_policies_and_guidelines

[10] https://en.wikipedia.org/wiki/Wikipedia:Simplified_ruleset

[11] https://en.m.wikipedia.org/wiki/Wikipedia:Editing_policy

Wenhu Chen, Hongmin Wang, Jianshu Chen, Yunkai Zhang, Hong Wang, Shiyang Li, Xiyou Zhou, and William Yang Wang. 2020a. Tabfact: A large-scale dataset for table-based fact verification. In *International Conference on Learning Representations*.

Wenhu Chen, Hanwen Zha, Zhiyu Chen, Wenhan Xiong, Hong Wang, and William Yang Wang. 2020b. HybridQA: A dataset of multi-hop question answering over tabular and textual data. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1026–1036, Online. Association for Computational Linguistics.

Zhiyu Chen, Wenhu Chen, Charese Smiley, Sameena Shah, Iana Borova, Dylan Langdon, Reema Moussa, Matt Beane, Ting-Hao Huang, Bryan Routledge, and William Yang Wang. 2021c. FinQA: A dataset of numerical reasoning over financial data. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3697–3711, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Zhiyu Chen, Wenhu Chen, Hanwen Zha, Xiyou Zhou, Yunkai Zhang, Sairam Sundaresan, and William Yang Wang. 2020c. Logic2Text: High-fidelity natural language generation from logical forms. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2096–2111, Online. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *CoRR*, abs/1911.02116.

Johannes Daxenberger and Iryna Gurevych. 2013. Automatically classifying edit categories in Wikipedia revisions. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 578–589, Seattle, Washington, USA. Association for Computational Linguistics.

Xiang Deng, Huan Sun, Alyssa Lees, You Wu, and Cong Yu. 2022. Turl: Table understanding through representation learning. *SIGMOD Rec.*, 51(1):33–40.

Haoyu Dong, Zhoujun Cheng, Xinyi He, Mengyu Zhou, Anda Zhou, Fan Zhou, Ao Liu, Shi Han, and Dongmei Zhang. 2022. Table pretraining: A survey on model architectures, pretraining objectives, and downstream tasks. *arXiv preprint arXiv:2201.09745*.

Jane Dwivedi-Yu, Timo Schick, Zhengbao Jiang, Maria Lomeli, Patrick Lewis, Gautier Izacard, Edouard Grave, Sebastian Riedel, and Fabio Petroni. 2022. Editeval: An instruction-based benchmark for text improvements. *arXiv preprint arXiv:2209.13331*.

Julian Eisenschlos, Syrine Krichene, and Thomas Müller. 2020. Understanding tables with intermediate pre-training. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 281–296, Online. Association for Computational Linguistics.

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. Language-agnostic BERT sentence embedding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 878–891, Dublin, Ireland. Association for Computational Linguistics.

Sébastien Ferré. 2012. Squall: A controlled natural language for querying and updating rdf graphs. In *International Workshop on Controlled Natural Language*, pages 11–25. Springer.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6894–6910, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Michael Glass, Mustafa Canim, Alfio Gliozzo, Saneem Chemmengath, Vishwajeet Kumar, Rishav Chakravarti, Avi Sil, Feifei Pan, Samarth Bharadwaj, and Nicolas Rodolfo Fauceglia. 2021. Capturing row and column semantics in transformer based question answering over tables. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1212–1224, Online. Association for Computational Linguistics.

Simon Gottschalk and Elena Demidova. 2017. Multiwiki: Interlingual text passage alignment in wikipedia. *ACM Trans. Web*, 11(1).

Vivek Gupta, Riyaz A. Bhat, Atreya Ghosal, Manish Shrivastava, Maneesh Singh, and Vivek Srikumar. 2022a. Is my model using the right evidence? systematic probes for examining evidence-based tabular reasoning. *Transactions of the Association for Computational Linguistics*, 10:659–679.

Vivek Gupta, Maitrey Mehta, Pegah Nokhiz, and Vivek Srikumar. 2020. INFOTABS: Inference on tables as semi-structured data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2309–2324, Online. Association for Computational Linguistics.

Vivek Gupta, Shuo Zhang, Alakananda Vempala, Yujie He, Temma Choji, and Vivek Srikumar. 2022b. Right for the right reason: Evidence extraction for trustworthy tabular reasoning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3268–3283, Dublin, Ireland. Association for Computational Linguistics.

Jonathan Herzig, Thomas Müller, Syrine Krichene, and Julian Eisenschlos. 2021. Open domain question answering over tables via dense retrieval. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 512–519, Online. Association for Computational Linguistics.

Jonathan Herzig, Pawel 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*, pages 4320–4333, Online. Association for Computational Linguistics.

Hiroshi Iida, Dung Thai, Varun Manjunatha, and Mohit Iyyer. 2021. TABBIE: Pretrained representations of tabular data. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3446–3456, Online. Association for Computational Linguistics.

Robert Iv, Alexandre Passos, Sameer Singh, and Ming-Wei Chang. 2022. FRUIT: Faithfully reflecting updated information in text. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3670–3686, Seattle, United States. Association for Computational Linguistics.

Mohit Iyyer, Wen-tau Yih, and Ming-Wei Chang. 2017. Search-based neural structured learning for sequential question answering. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1821–1831, Vancouver, Canada. Association for Computational Linguistics.

Kokil Jaidka, Andrea Ceolin, Iknoor Singh, Niyati Chhaya, and Lyle Ungar. 2021. WikiTalkEdit: A dataset for modeling editors' behaviors on Wikipedia. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2191–2200, Online. Association for Computational Linguistics.

Saemi Jang, Mun Yong Yi, et al. 2016. Utilization of dbpedia mapping in cross lingual wikipedia infobox completion. In *Australasian Joint Conference on Artificial Intelligence*, pages 303–316. Springer.

Divyansh Kaushik, Eduard Hovy, and Zachary Lipton. 2020. Learning the difference that makes a difference with counterfactually-augmented data. In *International Conference on Learning Representations*.

Aneta Koleva, Martin Ringsquandl, and Volker Tresp. Analysis of the attention in tabular language models. In *NeurIPS 2022 First Table Representation Workshop*.

Dustin Lange, Christoph Böhm, and Felix Naumann. 2010. Extracting structured information from wikipedia articles to populate infoboxes. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, page 1661–1664, New York, NY, USA. Association for Computing Machinery.

Xi Victoria Lin, Richard Socher, and Caiming Xiong. 2020. Bridging textual and tabular data for cross-domain text-to-SQL semantic parsing. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4870–4888, Online. Association for Computational Linguistics.

Bhavnick Minhas, Anant Shankhdhar, Vivek Gupta, Divyanshu Aggarwal, and Shuo Zhang. 2022. XInfoTabS: Evaluating multilingual tabular natural language inference. In *Proceedings of the Fifth Fact Extraction and VERification Workshop (FEVER)*, pages 59–77, Dublin, Ireland. Association for Computational Linguistics.

Thomas Müller, Julian Eisenschlos, and Syrine Krichene. 2021. TAPAS at SemEval-2021 task 9: Reasoning over tables with intermediate pre-training. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 423–430, Online. Association for Computational Linguistics.

Linyong Nan, Chiachun Hsieh, Ziming Mao, Xi Victoria Lin, Neha Verma, Rui Zhang, Wojciech Kryściński, Hailey Schoelkopf, Riley Kong, Xiangru Tang, Mutethia Mutuma, Ben Rosand, Isabel Trindade, Renusree Bandaru, Jacob Cunningham, Caiming Xiong, Dragomir Radev, and Dragomir Radev. 2022. FeTaQA: Free-form table question answering. *Transactions of the Association for Computational Linguistics*, 10:35–49.

Linyong Nan, Dragomir Radev, Rui Zhang, Amrit Rau, Abhinand Sivaprasad, Chiachun Hsieh, Xiangru Tang, Aadit Vyas, Neha Verma, Pranav Krishna, Yangxiaokang Liu, Nadia Irwanto, Jessica Pan, Faiaz Rahman, Ahmad Zaidi, Mutethia Mutuma, Yasin Tarabar, Ankit Gupta, Tao Yu, Yi Chern Tan, Xi Victoria Lin, Caiming Xiong, Richard Socher, and Nazneen Fatema Rajani. 2021. DART: Open-domain structured data record to text generation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 432–447, Online. Association for Computational Linguistics.

J. Neeraja, Vivek Gupta, and Vivek Srikumar. 2021. Incorporating external knowledge to enhance tabular reasoning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2799–2809, Online. Association for Computational Linguistics.

Nhu Nguyen, Dung Cao, and Anh Nguyen. 2018. Automatically mapping wikipedia infobox attributes to dbpedia properties for fast deployment of vietnamese dbpedia chapter. In *Asian Conference on Intelligent Information and Database Systems*, pages 127–136. Springer.

Thanh Nguyen, Viviane Moreira, Huong Nguyen, Hoa Nguyen, and Juliana Freire. 2011. Multilingual schema matching for wikipedia infoboxes. *Proceedings of the VLDB Endowment*, 5(2).

Barlas Oguz, Xilun Chen, Vladimir Karpukhin, Stan Peshterliev, Dmytro Okhonko, Michael Schlichtkrull, Sonal Gupta, Yashar Mehdad, and Scott Yih. 2022. UniK-QA: Unified representations of structured and unstructured knowledge for open-domain question answering. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1535–1546, Seattle, United States. Association for Computational Linguistics.

Sheena Panthaplackel, Adrian Benton, and Mark Dredze. 2022. Updated headline generation: Creating updated summaries for evolving news stories. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6438–6461, Dublin, Ireland. Association for Computational Linguistics.

Ankur Parikh, Xuezhi Wang, Sebastian Gehrmann, Manaal Faruqui, Bhuwan Dhingra, Diyi Yang, and Dipanjan Das. 2020. ToTTo: A controlled table-to-text generation dataset. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1173–1186, Online. Association for Computational Linguistics.

Panupong Pasupat and Percy Liang. 2015a. 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)*, pages 1470–1480, Beijing, China. Association for Computational Linguistics.

Panupong Pasupat and Percy Liang. 2015b. 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)*, pages 1470–1480, Beijing, China. Association for Computational Linguistics.

Aniket Pramanick and Indrajit Bhattacharya. 2021. Joint learning of representations for web-tables, entities and types using graph convolutional network. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1197–1206, Online. Association for Computational Linguistics.

JOSEPH REAGLE and LAUREN RHUE. 2011. Gender bias in wikipedia and britannica. *International Journal of Communication*, 5:1138–1158.

Nils Reimers and Iryna Gurevych. 2019a. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2019b. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. 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)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2020a. Making monolingual sentence embeddings multilingual using knowledge distillation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4512–4525, Online. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2020b. Making monolingual sentence embeddings multilingual using knowledge distillation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4512–4525, Online. Association for Computational Linguistics.

Daniel Rinser, Dustin Lange, and Felix Naumann. 2013. Cross-lingual entity matching and infobox alignment in wikipedia. *Information Systems*, 38(6):887–907.

Dwaipayan Roy, Sumit Bhatia, and Prateek Jain. 2020. A topic-aligned multilingual corpus of Wikipedia articles for studying information asymmetry in low resource languages. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 2373–2380, Marseille, France. European Language Resources Association.

Ohad Rozen, Vered Shwartz, Roee Aharoni, and Ido Dagan. 2019. Diversify your datasets: Analyzing generalization via controlled variance in adversarial datasets. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 196–205, Hong Kong, China. Association for Computational Linguistics.

Tomás Sáez and Aidan Hogan. 2018. Automatically generating wikipedia info-boxes from wikidata. In *Companion Proceedings of the The Web Conference 2018*, WWW '18, page 1823–1830, Republic and Canton of Geneva, CHE. International World Wide Web Conferences Steering Committee.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Soumya Sarkar, Bhanu Prakash Reddy, Sandipan Sikdar, and Animesh Mukherjee. 2019. StRE: Self attentive edit quality prediction in Wikipedia. In *Proceedings of the 57th Annual Meeting of the Association for*

Computational Linguistics, pages 3962–3972, Florence, Italy. Association for Computational Linguistics.

Darsh Shah, Tal Schuster, and Regina Barzilay. 2020. Automatic fact-guided sentence modification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):8791–8798.

Darsh J. Shah, Tal Schuster, and Regina Barzilay. 2019. Automatic fact-guided sentence modification. *CoRR*, abs/1909.13838.

Abhilash Shankarampeta, Vivek Gupta, and Shuo Zhang. 2022. Enhancing tabular reasoning with pattern exploiting training. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 706–726, Online only. Association for Computational Linguistics.

Tianze Shi, Chen Zhao, Jordan Boyd-Graber, Hal Daumé III, and Lillian Lee. 2020. On the potential of lexico-logical alignments for semantic parsing to SQL queries. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1849–1864, Online. Association for Computational Linguistics.

C Estelle Smith, Bowen Yu, Anjali Srivastava, Aaron Halfaker, Loren Terveen, and Haiyi Zhu. 2020. Keeping community in the loop: Understanding wikipedia stakeholder values for machine learning-based systems. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14.

Gowthami Somepalli, Micah Goldblum, Avi Schwarzschild, C Bayan Bruss, and Tom Goldstein. 2021. Saint: Improved neural networks for tabular data via row attention and contrastive pre-training. *arXiv preprint arXiv:2106.01342*.

Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2020. MPNet: Masked and permuted pre-training for language understanding. In *Advances in Neural Information Processing Systems*, volume 33, pages 16857–16867. Curran Associates, Inc.

Alexander Spangher, Xiang Ren, Jonathan May, and Nanyun Peng. 2022. NewsEdits: A news article revision dataset and a novel document-level reasoning challenge. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 127–157, Seattle, United States. Association for Computational Linguistics.

Afroza Sultana, Quazi Mainul Hasan, Ashis Kumer Biswas, Soumyava Das, Habibur Rahman, Chris Ding, and Chengkai Li. 2012. Infobox suggestion for wikipedia entities. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12, page 2307–2310, New York, NY, USA. Association for Computing Machinery.

Huan Sun, Hao Ma, Xiaodong He, Wen-tau Yih, Yu Su, and Xifeng Yan. 2016. Table cell search for question answering. In *Proceedings of the 25th International Conference on World Wide Web*, pages 771–782.

Yu Suzuki, Yuya Fujiwara, Yukio Konishi, and Akiyo Nadamoto. 2012. Good quality complementary information for multilingual wikipedia. In *International Conference on Web Information Systems Engineering*, pages 185–198. Springer.

Thang Hoang Ta and Chutiporn Anutariya. 2015. A model for enriching multilingual wikipedias using infobox and wikidata property alignment. In *Joint International Semantic Technology Conference*, pages 335–350. Springer.

Ramine Tinati, Paul Gaskell, Thanassis Tiropanis, Olivier Phillipe, and Wendy Hall. 2014. Examining wikipedia across linguistic and temporal borders. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14 Companion, page 445–450, New York, NY, USA. Association for Computing Machinery.

Mohamed Trabelsi, Zhiyu Chen, Shuo Zhang, Brian D. Davison, and Jeff Heflin. 2022. Strubert: Structure-aware bert for table search and matching. In *Proceedings of the ACM Web Conference 2022*, WWW '22, page 442–451, New York, NY, USA. Association for Computing Machinery.

Zhigang Wang, Zhixing Li, Juanzi Li, Jie Tang, and Jeff Z. Pan. 2013. Transfer learning based cross-lingual knowledge extraction for Wikipedia. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 641–650, Sofia, Bulgaria. Association for Computational Linguistics.

Morten Warncke-Wang, Anuradha Uduwage, Zhenhua Dong, and John Riedl. 2012. In search of the ur-wikipedia: Universality, similarity, and translation in the wikipedia inter-language link network. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, WikiSym '12, New York, NY, USA. Association for Computing Machinery.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Wenhan Xiong, Jingfei Du, William Yang Wang, and Veselin Stoyanov. 2020. Pretrained encyclopedia: Weakly supervised knowledge-pretrained language model. In *International Conference on Learning Representations*.

Diyi Yang, Aaron Halfaker, Robert Kraut, and Eduard Hovy. 2017. Identifying semantic edit intentions from revisions in Wikipedia. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2000–2010, Copenhagen, Denmark. Association for Computational Linguistics.

Jingfeng Yang, Aditya Gupta, Shyam Upadhyay, Luheng He, Rahul Goel, and Shachi Paul. 2022. TableFormer: Robust transformer modeling for table-text encoding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 528–537, Dublin, Ireland. Association for Computational Linguistics.

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*, pages 8413–8426, Online. Association for Computational Linguistics.

Ori Yoran, Alon Talmor, and Jonathan Berant. 2022. Turning tables: Generating examples from semi-structured tables for endowing language models with reasoning skills. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6016–6031, Dublin, Ireland. Association for Computational Linguistics.

Tao Yu, Chien-Sheng Wu, Xi Victoria Lin, Bailin Wang, Yi Chern Tan, Xinyi Yang, Dragomir R. Radev, Richard Socher, and Caiming Xiong. 2021. Grappa: Grammar-augmented pre-training for table semantic parsing. In *International Conference of Learning Representation*.

Tao Yu, Rui Zhang, Kai Yang, Michihiro Yasunaga, Dongxu Wang, Zifan Li, James Ma, Irene Li, Qingning Yao, Shanelle Roman, Zilin Zhang, and Dragomir Radev. 2018. Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and text-to-SQL task. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3911–3921, Brussels, Belgium. Association for Computational Linguistics.

Vicky Zayats, Kristina Toutanova, and Mari Ostendorf. 2021. Representations for question answering from documents with tables and text. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2895–2906, Online. Association for Computational Linguistics.

Hongzhi Zhang, Yingyao Wang, Sirui Wang, Xuezhi Cao, Fuzheng Zhang, and Zhongyuan Wang. 2020a. Table fact verification with structure-aware transformer. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1624–1629, Online. Association for Computational Linguistics.

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*, SIGIR'19, page 1029–1032, New York, NY, USA. Association for Computing Machinery.

Shuo Zhang and Krisztian Balog. 2020a. Web table extraction, retrieval, and augmentation: A survey. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(2):1–35.

Shuo Zhang and Krisztian Balog. 2020b. Web table extraction, retrieval, and augmentation: A survey. *ACM Trans. Intell. Syst. Technol.*, 11(2).

Shuo Zhang, Krisztian Balog, and Jamie Callan. 2020b. Generating categories for sets of entities. CIKM '20, page 1833–1842, New York, NY, USA. Association for Computing Machinery.

Shuo Zhang, Zhuyun Dai, Krisztian Balog, and Jamie Callan. 2020c. Summarizing and exploring tabular data in conversational search. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, page 1537–1540, New York, NY, USA. Association for Computing Machinery.

Shuo Zhang, Edgar Meij, Krisztian Balog, and Ridho Reinanda. 2020d. Novel entity discovery from web tables. In *Proceedings of The Web Conference 2020*, WWW '20, pages 1298–1308.

Yan Zhang, Thomas Paradis, Lei Hou, Juanzi Li, Jing Zhang, and Haitao Zheng. 2017. Cross-lingual infobox alignment in wikipedia using entity-attribute factor graph. In *International Semantic Web Conference*, pages 745–760. Springer.

Fengbin Zhu, Wenqiang Lei, Youcheng Huang, Chao Wang, Shuo Zhang, Jiancheng Lv, Fuli Feng, and Tat-Seng Chua. 2021. TAT-QA: A question answering benchmark on a hybrid of tabular and textual content in finance. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3277–3287, Online. Association for Computational Linguistics.

A Appendix

A.1 Table Extraction Details

Table formats and HTML code styles differ from one language to another and even across categories in the same language. Extraction is modified to

Category	Entities	*std_dev*	Category	Entities	*std_dev*
Airport	2563	5.03	Country	259	10.28
Album	840	3.81	Diseases	462	4.20
Animal	368	3.37	Food	692	4.34
Athlete	369	5.80	Medicine	334	9.58
Book	218	5.24	Monument	203	5.23
City	262	7.95	Movie	1524	6.75
College	202	5.83	Musician	284	5.09
Company	267	6.87	Nobel	967	5.29
Painting	743	3.51	Stadium	742	5.86
Person	198	6.32	Shows	1044	6.83
Planet	188	8.46			

Table 10: **Missing information Analysis in Categories**:- For each category unique number of entities and their average standard deviation across languages.

C1	Row Diff	C1	Row Diff
af	5.28	hi	5.06
ar	5.84	ko	4.30
ceb	3.33	nl	3.86
de	5.96	ru	4.1
en	4.80	sv	3.92
es	5.17	tr	4.23
fr	4.42	zh	4.76

Table 11: **Row Difference Across Paired Languages**:- Column 2 shows average row count difference between languages for all entities.

handle these variations, which requires the following steps: (a) *Detecting Infoboxes:* We locate Wikipedia infoboxes that appear in at least five languages. (b) *Extracting HTML:* After detection, we extract HTML and preprocess to remove images, links, and signatures. (c) *Table Representation:* we convert the extracted table and store them in JSON.

Row Difference Across Paired Languages: There is substantial variation in the number of rows for infobox across different languages, i.e., rows difference = $\frac{1}{|L|}\sum_{ln \in L \setminus C1} ||R_{c1}| - |R_{ln}||$, where L is set of all 14 languages under consideration. Table 11 shows that German followed by Arabic and Afrikaans, has the highest row difference. This indicates that tables in these languages are incomplete (with missing rows).

A.2 Table Updation Examples

An example of table updation is shown in the Figure 3.

A.3 Precision and Recall

We also evaluated precision-recall values in information alignment for matched and unmatched scores (§5.2). Precision recall values for $T_{en} \leftrightarrow T_x$, $T_x \leftrightarrow T_y$, $T_{en} \overset{*}{\leftrightarrow} T_{hi}$ and $T_{en} \overset{*}{\leftrightarrow} T_{zh}$ are reported in Tables 17, 18, 19, and 20, respectively.

A.4 Algorithm Coverage

We measure the coverage on the entire corpus, the rate of rows aligned w.r.t. the smaller table in a table pair. Table 12 reports ablations results of coverage for various modules. Our proposed method aligns 72.54% and 67.96% of rows for $T_{en} \leftrightarrow T_x$ and $T_x \leftrightarrow T_y$, respectively. Corpus-based is the most constrained module, focusing more on precision; hence removing corpus-based gives better coverage for both cases. Key-Only-Unidirectional is the most important module for coverage, followed by the Key-Only module for both cases.

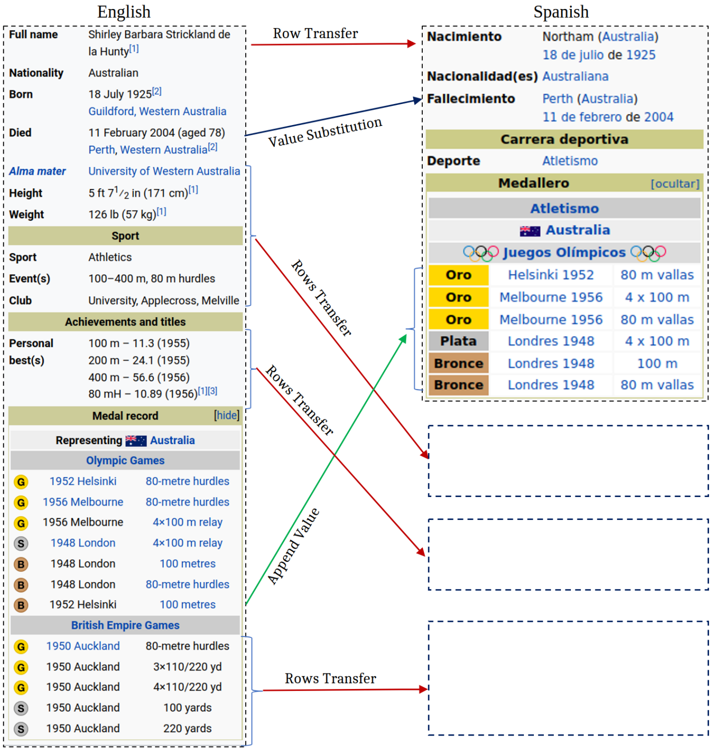
A.5 Domain and Language Wise Analysis

Table 13, 14, and 15 show the performance of our proposed method grouped by languages, domains, and keys, respectively.

Group-wise Analysis. From Table 13, for $T_{en} \leftrightarrow T_x$, Cebuano, Arabic, German, and Dutch are the worst performing languages with F1-score close to 85 for alignment. Whereas Turkish, Chinese, and Hindi have F1-score greater than 90. Korean, German, and Swedish are the lowest-performing language groups, with an F1-Score close to 86 for unaligned settings. Cebuano, Turkish, and Dutch get the highest score for unaligned metrics (greater than 90). For non-English language pairs, the lowest F1-score for match table pairs is observed for German-Arabic and Hindi-Korean pairs with an F1-score close to 78, as shown in Table 13. The highest F1-score is observed for Russian-German and Hindi-German, with F1-scores exceeding 88.8. For unmatched data, Korean-Hindi, French-Hindi, French-Korean, and Russian-Korean pairs have the lowest F1 scores, less than 85. In contrast, German-Hindi and Russion-German have exceeded the unaligned F1-Score of 90.

Category-wise Analysis. As reported in Table 14, our method performs worst in Airport and College categories for match settings when one of the languages is English. For non-English match settings, Movie and City are the worst-performing categories. For unmatch setting with English as one of the languages, Airport and Painting have the lowest F1-score, whereas Movie and Stadium have the most inferior performance for non-English languages.

Key-wise Analysis. Table 15 shows the average F1-scores across tables for frequent and non-frequent keys. We observed an F1-score degrada-



Full name	Shirley Barbara Strickland de la Hunty[1]
Nationality	Australian
Born	18 July 1925[2] Guildford, Western Australia
Died	11 February 2004 (aged 78) Perth, Western Australia[2]
Alma mater	University of Western Australia
Height	5 ft 7$^1/_2$ in (171 cm)[1]
Weight	126 lb (57 kg)[1]
Sport	
Sport	Athletics
Event(s)	100–400 m, 80 m hurdles
Club	University, Applecross, Melville
Achievements and titles	
Personal best(s)	100 m – 11.3 (1955) 200 m – 24.1 (1955) 400 m – 56.6 (1956) 80 mH – 10.89 (1956)[1][3]

Medal record [hide]

Representing Australia

Olympic Games		
G	1952 Helsinki	80-metre hurdles
G	1956 Melbourne	80-metre hurdles
G	1956 Melbourne	4×100 m relay
S	1948 London	4×100 m relay
B	1948 London	100 metres
B	1948 London	80-metre hurdles
B	1952 Helsinki	100 metres
British Empire Games		
G	1950 Auckland	80-metre hurdles
G	1950 Auckland	3×110/220 yd
G	1950 Auckland	4×110/220 yd
S	1950 Auckland	100 yards
S	1950 Auckland	220 yards

Nacimiento	Northam (Australia) 18 de julio de 1925
Nacionalidad(es)	Australiana
Fallecimiento	Perth (Australia) 11 de febrero de 2004
Carrera deportiva	
Deporte	Atletismo

Medallero [ocultar]

Atletismo

Australia

Juegos Olímpicos		
Oro	Helsinki 1952	80 m vallas
Oro	Melbourne 1956	4 x 100 m
Oro	Melbourne 1956	80 m vallas
Plata	Londres 1948	4 x 100 m
Bronce	Londres 1948	100 m
Bronce	Londres 1948	80 m vallas

Figure 3: Update Example:- "Shirley Strickland de la Hunty " Infoboxes for two languages, i.e., English and Spanish. Shows rows transfer for missing information. Value substitution because "Aged 78" is absent in Died. One medal information (Bronze,1952, 100m) is added to the medal tally.

Ablation	$T_{en} \leftrightarrow T_x$					$T_x \leftrightarrow T_y$				
	Corpus	Key	K-V-Bi	K-V-Uni	Multi	Corpus	Key	K-V-Bi	K-V-Uni	Multi
w/o	16.28	40.83	58.17	71.95	**72.54**	17.15	39.78	57.53	67.58	**67.96**
Corpus	-	33.15	50.17	74.69	**75.3**	-	25.04	49.82	64.98	**65.41**
Key	16.28	-	57.88	71.14	**71.6**	17.15	-	55.05	64.8	**65.1**
K-V-Bi	16.28	38.88	-	71.9	**72.3**	17.15	37.83	-	70.32	**70.58**
K-V-Uni	16.28	46.19	21.34	-	**67.53**	17.15	40.19	62.91	-	**63.59**
Multi	16.28	40.96	58.4	**72.23**	-	17.15	36.36	55.03	**67.13**	-

Table 12: **Coverage Ablation:** $T_{en} \leftrightarrow T_x$ and $T_x \leftrightarrow T_y$.

tion of 10 points for rare keys with a low occurrence compared to frequent keys.

A.6 Ablation Study

We report ablation performance to highlight the significance of each module in Table 16. Key-Value-Bidirectional mapping (two-way) is the most critical module, followed by Key Only corpus-based modules. We also observe Uni-directional mapping being the second most important for non-English alignments. The multi-key module was consistently was least significant for the same reason as the discussion above (very few instances). Similar observations were valid for unmatching scores.

A.7 Further Details: Error Analysis

We discussed challenges to table information synchronization across languages in §2.1. Table 5 (main paper) shows the number of instances of these challenges in evaluation for matched cases after applying various modules of the alignment algorithm.

- Corpus-based module solves approximately (40%, 56%) of outdated information,(31%, 21%) of schema variation, (10%, 30%) of language variation, (13%, 25%) of unnormalized information and (37%, 26%) of erroneous entity linking challenges in $T_{en} \leftrightarrow T_x$ and $T_x \leftrightarrow T_y$, respectively.

$T_{en} \leftrightarrow T_x$	Match	UnMatch	$T_x \leftrightarrow T_y$	Match	UnMatch
af	88.08	89.48	de ↔ hi	88.85	90.4
ar	85.24	88.77	de ↔ ko	85.27	88.7
ceb	85.17	91.07	fr ↔ ar	85.35	87.21
de	85.41	86.65	fr ↔ de	84.97	88.94
es	89.83	89.7	fr ↔ hi	83.95	84.58
fr	89.41	89.8	fr ↔ ko	83.59	84.36
hi	90.56	87.07	fr ↔ ru	87.63	88.83
ko	85.69	86.22	hi ↔ ar	84.33	89.38
nl	86.4	90.28	ko ↔ ar	82.18	89.08
ru	87.46	88.54	ko ↔ hi	78.8	83.03
sv	84.89	86.76	ru ↔ ar	82.18	86.96
tr	92.07	91.3	ru ↔ de	89.93	91.92
zh	91.61	89.31	ru ↔ hi	82.38	87.78
			ru ↔ ko	81.62	84.47
			de ↔ ar	78.05	87.23

Table 13: **Language Wise Analysis** :-Alignment F1-score reported for same language for $T_{en} \leftrightarrow T_x$ and $T_x \leftrightarrow T_y$ averaged over all entities.

- Further adding of key only similarity module resolves extra (24%, 13%) of outdated information, (36%, 21%) of schema variation, (13%, 5%) of language variation, (19%, 17%) of unnormalized information and (22%, 8%) of erroneous entity linking challenges in $T_{en} \leftrightarrow T_x$ and $T_x \leftrightarrow T_y$, respectively.

- Applying key-value-bidirectional module resolves another (12%, 13.5%) of outdated information, (18%, 6.6%) of information representation, 54%, 45% of language variation, (40%, 46%) of unnormalized information and (34%, 53%) of erroneous entity linking challenges in $T_{en} \leftrightarrow T_x$ and $T_x \leftrightarrow T_y$, respectively.

- Key-Val-Unidirectional and Multi-key together solves another (18.5%, 7.5%) of the information representation in $T_{en} \leftrightarrow T_x$ and $T_x \leftrightarrow T_y$, respectively, but are not effective against other challenges.

A.8 Other Related Work

Tabular Reasoning. Addressing NLP tasks on semi-structured tabular data has received substantial attention. There is work on tabular NLI (Gupta et al., 2020; Chen et al., 2020a; Gupta et al., 2022b), question-answering task (Zhang and Balog, 2020b; Zhu et al., 2021; Pasupat and Liang, 2015a; Abbas et al., 2016; Sun et al., 2016; Chen et al., 2021a, 2020b; Lin et al., 2020; Zayats et al., 2021; Oguz et al., 2022, and others) and table-to-text generation (Zhang et al., 2020c; Parikh et al., 2020; Nan et al., 2021; Yoran et al., 2022; Chen et al., 2021b).

Category	$T_{en} \leftrightarrow T_x$ Match	$T_{en} \leftrightarrow T_x$ UnMatch	$T_x \leftrightarrow T_y$ Match	$T_x \leftrightarrow T_y$ UnMatch
Airport	79.77	82.64	85.79	90.9
Album	93.9	91.33	88.6	85.01
Animal	93.79	94.2	90	96.24
Athlete	86.6	90.21	83.75	88.81
Book	86.48	90.96	81.29	83.13
City	86.14	93.67	77.4	86.6
College	82.47	87.53	81.05	86.24
Company	87.49	85.15	85.5	86.7
Country	86.38	92.47	86.53	92.32
Food	88.58	90.04	85.65	91.67
Monument	84.86	86.14	87.66	89.6
Movie	91.2	85.7	74.33	76.19
Musician	89.47	85.62	89.04	93.27
Nobel	88.2	91.08	88.84	87.1
Painting	90.27	82.35	86.52	89.72
Person	87.37	87.79	79.85	87.74
Planet	90.93	85.77	85.01	87.18
Shows	91.23	88.89	83.65	78.84
Stadium	88.59	87.72	83.2	77.38

Table 14: **Category Wise Analysis** :- Alignment F1-score reported for same group entities average over all languages.

Key Freq	Range	# of Keys (all)	Avg Score
High	$100 \leq x$	33	90.71
Mid	$50 \leq x \leq 100$	49	89.33
Low	$x \leq 50$	700	81.82

Table 15: **Key Wise Analysis**:- F1-Score report for grouped keys.

Tabular Representation and Learning. There are also several works representing Wikipedia tables, such papers are TAPAS (Herzig et al., 2020), StrucBERT (Trabelsi et al., 2022), Table2vec (Zhang et al., 2019), TaBERT (Yin et al., 2020), TABBIE (Iida et al., 2021), TabStruc (Zhang et al., 2020a), TabGCN (Pramanick and Bhattacharya, 2021), RCI (Glass et al., 2021), TURL (Deng et al., 2022), and TableFormer (Yang et al., 2022). Some papers such as (Yu et al., 2018, 2021; Eisenschlos et al., 2020; Neeraja et al., 2021; Müller et al., 2021; Somepalli et al., 2021; Shankarampeta et al., 2022; Dong et al., 2022, and others) study pre-training for tabular tasks. Paper related to tabular probing includes (Koleva et al.; Gupta et al., 2022a).

Tabular Datasets. There are several tabular task datasets on (a.) tabular NLI: (Gupta et al., 2020; Rozen et al., 2019; Müller et al., 2021; Kaushik et al., 2020; Xiong et al., 2020; Chen et al., 2020a; Eisenschlos et al., 2020; Chen et al., 2020c, and others); (b.) Tabular QA: WikiTableQA (Pasupat and Liang, 2015b), HybridQA (Chen et al., 2020b; Zayats et al., 2021; Oguz et al., 2022),WikiSQL

Ablation	Match				UnMatch			
	$T_{en} \leftrightarrow T_x$	$T_x \leftrightarrow T_y$	$T_{en} \overset{*}{\leftrightarrow} T_{hi}$	$T_{en} \overset{*}{\leftrightarrow} T_{zh}$	$T_{en} \leftrightarrow T_x$	$T_x \leftrightarrow T_y$	$T_{en} \overset{*}{\leftrightarrow} T_{hi}$	$T_{en} \overset{*}{\leftrightarrow} T_{zh}$
Corpus-based	86.67	82.3	89.13	92.33	87.95	87.03	83.11	87.38
Key Only	89	80.09	87.35	91.49	89.42	85.88	79.83	87.13
Key-Val-Bi	84.98	75.39	86.95	90.41	86.39	82.06	80.48	84.4
Key-Val-Uni	87.73	79.35	90	92.67	89.03	85.35	84.83	88.74
Multi-Key	87.89	84.33	-	-	89.52	85.42	-	-
w/o	87.91	84.36	90.14	92.8	89.03	85.46	84.98	88.17

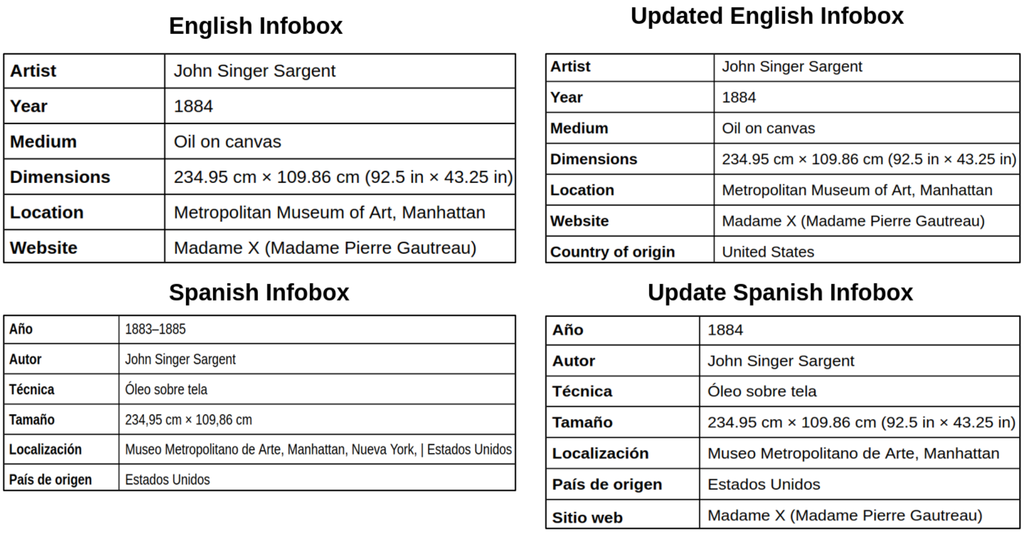
Table 16: **Ablation Study of Matched and UnMatch Score :** i.e. F1-Score for all test sets of InfoSync.

Alignment	Match			UnMatch		
	Precision	Recall	F1	Precision	Recall	F1
Corpus-based	93.51	46.22	61.86	55.66	96.17	70.51
+ Key Only	88.09	58.62	70.4	60.75	94.16	73.85
+ Key-Value-Bi	89.6	85.89	87.71	85.87	93.47	89.51
+ Key-Value-Uni	89.3	86.52	87.89	86.24	93.07	89.52
+ Multi-Key	88.85	86.99	87.91	86.51	92.27	89.3

Table 17: $T_{en} \leftrightarrow T_x$ alignment performance on Human-Annotated Test Data

(Iyyer et al., 2017), SQUALL (Ferré, 2012; Shi et al., 2020), OpenTableQA (Chen et al., 2021b), FinQA (Chen et al., 2021c), FeTaQA (Nan et al., 2022), TAT-QA (Zhu et al., 2021), SQA (Iyyer et al., 2017), NQ-Tables (Herzig et al., 2021); (c.) and Table Generation: ToTTo (Parikh et al., 2020), Turing Tables (Yoran et al., 2022), LogicNLG (Chen et al., 2020c).

Furthermore, there are also several works discussed on web table extraction, retrieval, and augmentation (Zhang and Balog, 2020a), and utilizing the transformers model for table representation (Badaro and Papotti, 2022).



English Infobox

Artist	John Singer Sargent
Year	1884
Medium	Oil on canvas
Dimensions	234.95 cm × 109.86 cm (92.5 in × 43.25 in)
Location	Metropolitan Museum of Art, Manhattan
Website	Madame X (Madame Pierre Gautreau)

Updated English Infobox

Artist	John Singer Sargent
Year	1884
Medium	Oil on canvas
Dimensions	234.95 cm × 109.86 cm (92.5 in × 43.25 in)
Location	Metropolitan Museum of Art, Manhattan
Website	Madame X (Madame Pierre Gautreau)
Country of origin	United States

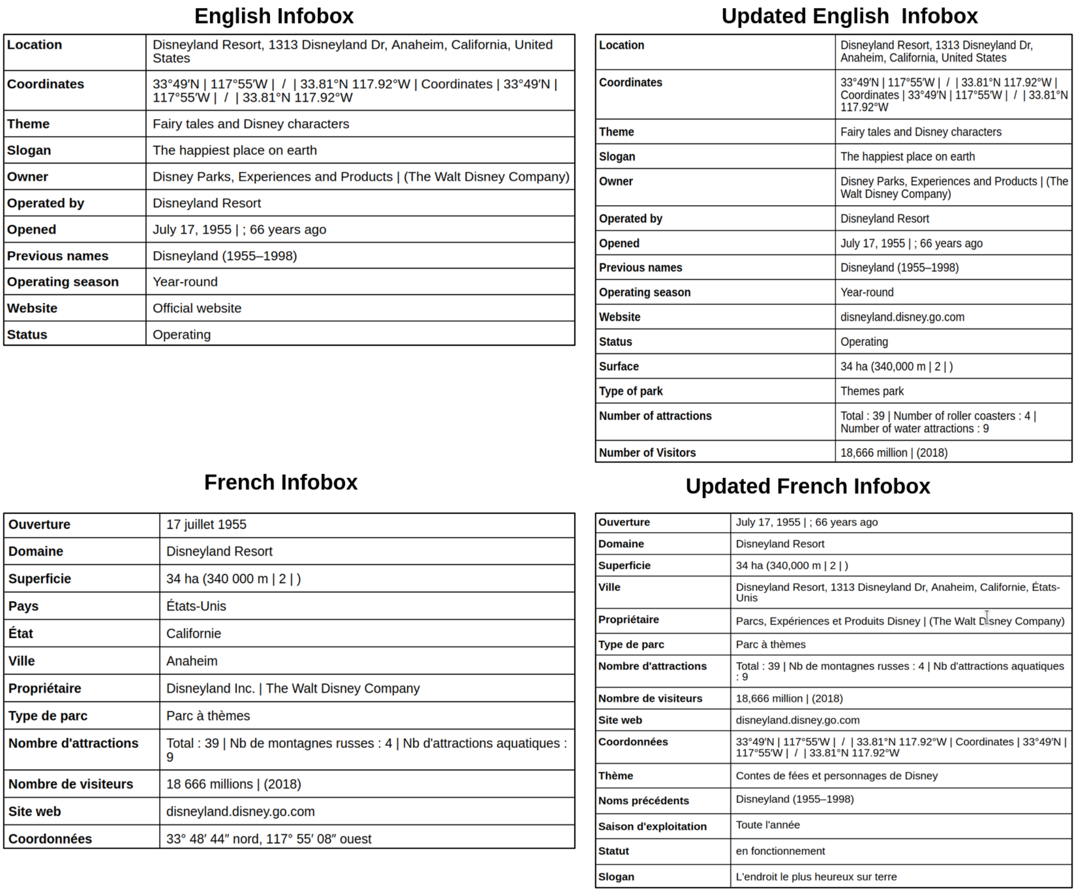
Spanish Infobox

Año	1883–1885
Autor	John Singer Sargent
Técnica	Óleo sobre tela
Tamaño	234,95 cm × 109,86 cm
Localización	Museo Metropolitano de Arte, Manhattan, Nueva York, | Estados Unidos
País de origen	Estados Unidos

Update Spanish Infobox

Año	1884
Autor	John Singer Sargent
Técnica	Óleo sobre tela
Tamaño	234.95 cm × 109.86 cm (92.5 in × 43.25 in)
Localización	Museo Metropolitano de Arte, Manhattan
País de origen	Estados Unidos
Sitio web	Madame X (Madame Pierre Gautreau)

Figure 4: **Example From Update Algorithm Proposed**: Update English Infobox is obtained by using Spanish Infobox as a reference and vice versa."Country of origin" is updated in English infobox and "website" is updated in Spanish infobox.



English Infobox

Location	Disneyland Resort, 1313 Disneyland Dr, Anaheim, California, United States
Coordinates	33°49′N | 117°55′W | / | 33.81°N 117.92°W | Coordinates | 33°49′N | 117°55′W | / | 33.81°N 117.92°W
Theme	Fairy tales and Disney characters
Slogan	The happiest place on earth
Owner	Disney Parks, Experiences and Products | (The Walt Disney Company)
Operated by	Disneyland Resort
Opened	July 17, 1955 | ; 66 years ago
Previous names	Disneyland (1955–1998)
Operating season	Year-round
Website	Official website
Status	Operating

Updated English Infobox

Location	Disneyland Resort, 1313 Disneyland Dr, Anaheim, California, United States
Coordinates	33°49′N | 117°55′W | / | 33.81°N 117.92°W | Coordinates | 33°49′N | 117°55′W | / | 33.81°N 117.92°W
Theme	Fairy tales and Disney characters
Slogan	The happiest place on earth
Owner	Disney Parks, Experiences and Products | (The Walt Disney Company)
Operated by	Disneyland Resort
Opened	July 17, 1955 | ; 66 years ago
Previous names	Disneyland (1955–1998)
Operating season	Year-round
Website	disneyland.disney.go.com
Status	Operating
Surface	34 ha (340,000 m | 2 |)
Type of park	Themes park
Number of attractions	Total : 39 | Number of roller coasters : 4 | Number of water attractions : 9
Number of Visitors	18,666 million | (2018)

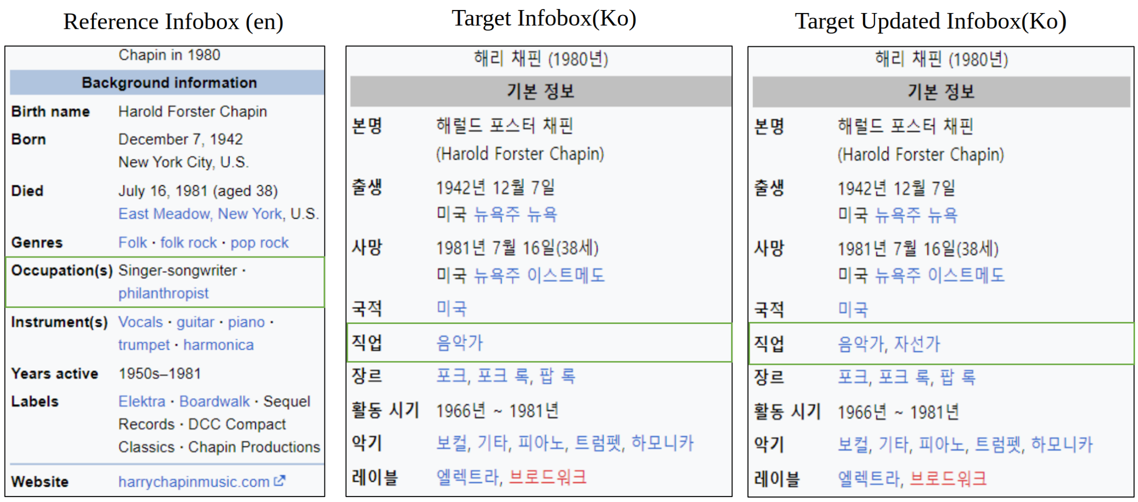
French Infobox

Ouverture	17 juillet 1955
Domaine	Disneyland Resort
Superficie	34 ha (340 000 m | 2 |)
Pays	États-Unis
État	Californie
Ville	Anaheim
Propriétaire	Disneyland Inc. | The Walt Disney Company
Type de parc	Parc à thèmes
Nombre d'attractions	Total : 39 | Nb de montagnes russes : 4 | Nb d'attractions aquatiques : 9
Nombre de visiteurs	18 666 millions | (2018)
Site web	disneyland.disney.go.com
Coordonnées	33° 48′ 44″ nord, 117° 55′ 08″ ouest

Updated French Infobox

Ouverture	July 17, 1955 | ; 66 years ago
Domaine	Disneyland Resort
Superficie	34 ha (340,000 m | 2 |)
Ville	Disneyland Resort, 1313 Disneyland Dr, Anaheim, Californie, États-Unis
Propriétaire	Parcs, Expériences et Produits Disney | (The Walt Disney Company)
Type de parc	Parc à thèmes
Nombre d'attractions	Total : 39 | Nb de montagnes russes : 4 | Nb d'attractions aquatiques : 9
Nombre de visiteurs	18,666 million | (2018)
Site web	disneyland.disney.go.com
Coordonnées	33°49′N | 117°55′W | / | 33.81°N 117.92°W | Coordinates | 33°49′N | 117°55′W | / | 33.81°N 117.92°W
Thème	Contes de fées et personnages de Disney
Noms précédents	Disneyland (1955–1998)
Saison d'exploitation	Toute l'année
Statut	en fonctionnement
Slogan	L'endroit le plus heureux sur terre

Figure 5: **Example From Update Algorithm Proposed**: Update English Infobox is obtained by using French Infobox as a reference and vice versa. Multiple keys are updated in both infoboxes "Opened," "Location," "Owner," "Coordinates," "Operating Season," and "Slogan" in French, and "number of visitors," "surface," "Type of park," number of attractions" are updated in English infobox.



Reference Infobox (en)

Chapin in 1980

Background information	
Birth name	Harold Forster Chapin
Born	December 7, 1942 New York City, U.S.
Died	July 16, 1981 (aged 38) East Meadow, New York, U.S.
Genres	Folk · folk rock · pop rock
Occupation(s)	Singer-songwriter · philanthropist
Instrument(s)	Vocals · guitar · piano · trumpet · harmonica
Years active	1950s–1981
Labels	Elektra · Boardwalk · Sequel Records · DCC Compact Classics · Chapin Productions
Website	harrychapinmusic.com

Target Infobox(Ko)

해리 채핀 (1980년)

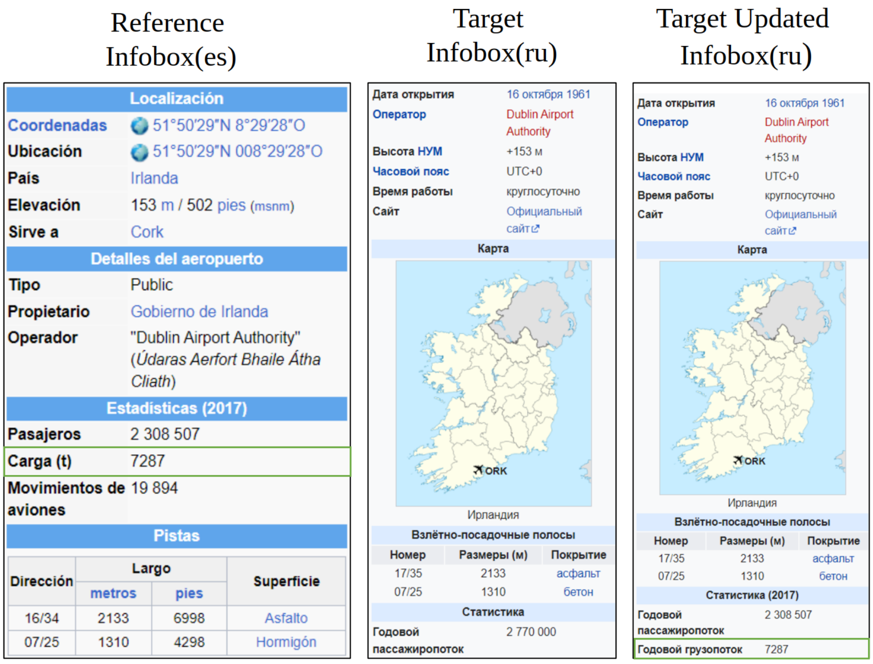
기본 정보	
본명	해럴드 포스터 채핀 (Harold Forster Chapin)
출생	1942년 12월 7일 미국 뉴욕주 뉴욕
사망	1981년 7월 16일(38세) 미국 뉴욕주 이스트메도
국적	미국
직업	음악가
장르	포크, 포크 록, 팝 록
활동 시기	1966년 ~ 1981년
악기	보컬, 기타, 피아노, 트럼펫, 하모니카
레이블	엘렉트라, 브로드워크

Target Updated Infobox(Ko)

해리 채핀 (1980년)

기본 정보	
본명	해럴드 포스터 채핀 (Harold Forster Chapin)
출생	1942년 12월 7일 미국 뉴욕주 뉴욕
사망	1981년 7월 16일(38세) 미국 뉴욕주 이스트메도
국적	미국
직업	음악가, 자선가
장르	포크, 포크 록, 팝 록
활동 시기	1966년 ~ 1981년
악기	보컬, 기타, 피아노, 트럼펫, 하모니카
레이블	엘렉트라, 브로드워크

Figure 6: **Example From Live Updates**: In the above figure, the Target infobox needs to be updated using Reference infobox(available in English version) as extra/grounding information. The updated infobox is shown in column 3, where the key 'job' is updated. This is an example of "Value substitution," as in Table 8. The red box highlights the updated information.



Reference Infobox(es)

Localización	
Coordenadas	51°50′29″N 8°29′28″O
Ubicación	51°50′29″N 008°29′28″O
País	Irlanda
Elevación	153 m / 502 pies (msnm)
Sirve a	Cork
Detalles del aeropuerto	
Tipo	Public
Propietario	Gobierno de Irlanda
Operador	"Dublin Airport Authority" (*Údaras Aerfort Bhaile Átha Cliath*)
Estadísticas (2017)	
Pasajeros	2 308 507
Carga (t)	7287
Movimientos de aviones	19 894

Pistas

Dirección	Largo		Superficie
	metros	pies	
16/34	2133	6998	Asfalto
07/25	1310	4298	Hormigón

Target Infobox(ru)

Дата открытия	16 октября 1961
Оператор	Dublin Airport Authority
Высота НУМ	+153 м
Часовой пояс	UTC+0
Время работы	круглосуточно
Сайт	Официальный сайт

Карта

Взлётно-посадочные полосы

Номер	Размеры (м)	Покрытие
17/35	2133	асфальт
07/25	1310	бетон

Статистика	
Годовой пассажиропоток	2 770 000

Target Updated Infobox(ru)

Дата открытия	16 октября 1961
Оператор	Dublin Airport Authority
Высота НУМ	+153 м
Часовой пояс	UTC+0
Время работы	круглосуточно
Сайт	Официальный сайт

Карта

Взлётно-посадочные полосы

Номер	Размеры (м)	Покрытие
17/35	2133	асфальт
07/25	1310	бетон

Статистика (2017)	
Годовой пассажиропоток	2 308 507
Годовой грузопоток	7287

Figure 7: **Example From Live Updates**: In the above figure, the target infobox needs to be updated using a reference infobox as extra/grounding information. The updated Infobox is shown in column 3, where the 'Load/Cargo Traffic' key is updated. This is an example of Row Addition, as referred to in Table 8. The red box highlights the updated information.

	Match			UnMatch		
Alignment	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Corpus-based	75.68	45.38	56.74	58.9	91.71	71.73
+ Key Only	74.45	58.62	62.14	62.44	89.37	73.52
+ Key-Value-Bi	82.78	85.66	84.2	82.53	88.73	85.52
+ Key-Value-Uni	82.2	86.58	84.33	82.94	88.05	85.42
+ Multi-Key	82.16	86.68	84.36	83.05	88.01	85.46

Table 18: $T_x \leftrightarrow T_y$ alignment performance on Human-Annotated Test Data.

	Match			UnMatch		
Alignment	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Corpus-based	94.81	41.1	57.34	37.55	96.19	71.73
+ Key Only	92.04	61.04	73.4	46.6	94.81	73.52
+ Key-Value-Bi	87.65	85.89	90.07	77.37	88.73	85.52
+ Key-Value-Uni	88.59	86.52	90.34	78.53	88.05	85.42
+ Multi-Key	91.15	88.59	90.14	78.52	88.01	85.46

Table 19: $T_{en} \overset{*}{\leftrightarrow} T_{hi}$ alignment performance on Human-Annotated Test Data.

	Match			UnMatch		
Alignment	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Corpus-based	89.94	56.41	69.33	47.19	95.26	63.11
+ Key Only	88.78	64.43	74.67	51.74	91.99	66.23
+ Key-Value-Bi	92.38	93.7	93.04	86.73	91.81	89.2
+ Key-Value-Uni	92.13	94.13	93.12	86.75	90.58	88.62
+ Multi-Key	91.51	94.13	92.8	86.73	89.66	88.17

Table 20: $T_{en} \overset{*}{\leftrightarrow} T_{zh}$ alignment performance on Human-Annotated Test Data.

ACL 2023 Responsible NLP Checklist

A For every submission:

☑ A1. Did you describe the limitations of your work?
After conclusion before the ethic statement

☑ A2. Did you discuss any potential risks of your work?
In the limitation section

☑ A3. Do the abstract and introduction summarize the paper's main claims?
In the abstract and introduction in section 1

☒ A4. Have you used AI writing assistants when working on this paper?
Left blank.

B ☑ Did you use or create scientific artifacts?

section 2 (Dataset) and section 4 (Model)

☑ B1. Did you cite the creators of artifacts you used?
Yes for models (Section 5)

☑ B2. Did you discuss the license or terms for use and / or distribution of any artifacts?
Non commercial academic use (dataset and models) discussed in the ethic statement section

☑ B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
In ethics statement section

□ B4. Did you discuss the steps taken to check whether the data that was collected / used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect / anonymize it?
Not applicable. Left blank.

☑ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
Section 3 and appendix

☑ B6. Did you report relevant statistics like the number of examples, details of train / test / dev splits, etc. for the data that you used / created? Even for commonly-used benchmark datasets, include the number of examples in train / validation / test splits, as these provide necessary context for a reader to understand experimental results. For example, small differences in accuracy on large test sets may be significant, while on small test sets they may not be.
Section 3 and appendix

C ☑ Did you run computational experiments?

Section 5

☑ C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
Section 5

The Responsible NLP Checklist used at ACL 2023 is adopted from NAACL 2022, with the addition of a question on AI writing assistance.

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
Section 5

☑ C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?
Section 5

☑ C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation), did you report the implementation, model, and parameter settings used (e.g., NLTK, Spacy, ROUGE, etc.)?
Section 5

D ☑ Did you use human annotators (e.g., crowdworkers) or research with human participants?
Section 3

□ D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?
Not applicable. Left blank.

□ D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?
Not applicable. Left blank.

□ D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating? For example, if you collected data via crowdsourcing, did your instructions to crowdworkers explain how the data would be used?
Not applicable. Left blank.

□ D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
Not applicable. Left blank.

□ D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
Not applicable. Left blank.