Faithful Temporal Question Answering over Heterogeneous Sources

Zhen Jia[1], **Philipp Christmann[2]**, and **Gerhard Weikum[2]**

[1]Southwest Jiaotong University, China   [2]Max Planck Institute for Informatics, Germany

FAITH is a faithful system for temporal QA providing explanatory evidence for answers

★ Lookups of *dates*

> When was Bohemian Rhapsody recorded? → *1975*

★ Questions with *explicit temporal constraints*

> Record company of Queen *in 1975*? → *EMI*
>
> Record company of Queen *in 1990*? → *Parlophone*

★ Questions with *implicit temporal constraints*

> Queen's record company *when recording Bohemian Rhapsody*? → *EMI*
>
> Queen's record company *after Freddie Mercury's death*? → *Parlophone*

⇒ Many QA systems falsely reply ***EMI*** by their popularity prior, being unfaithful to the temporal constraint in the question

⇒ Answer should satisfy **temporal constraints**

LIMITATIONS OF THE STATE-OF-THE-ART

☆ **Temporal constraints** matched solely in **latent space**

☆ **Answer** derivation **not explainable** for end users

☆ Operate only over a **single information source**, KB **or** text, inherently limiting the answer coverage of the QA system

FAITH

★ **Faithful answering** of temporal questions with time constraints, using novel mechanism for handling challenging **implicit questions**

★ **First temporal QA system** tapping into **heterogeneous sources** (Wikidata KB, Wikipedia text, tables, and infoboxes)

★ **Explainable answer derivations** and **explanatory evidence for answers**: answer traceable through entire system

FAITHFULNESS

★ **Structured modelling** of temporal information need

★ Transforming **implicit temporal constraints** into **explicit conditions** via the implicit question resolver

★ Dedicated **temporal pruning** of temporally incompatible evidence via **temporal rules**

> After managing FC Nantes, which football club did Antoine Raab take on next?

Understanding

Model for Time-aware Structured Frame for generating *question entity, question relation, expected answer type, temporal category,* and *temporal signal*	**Explicit Temporal Annotation** to identify explicit temporal values

Implicit Question Resolver for generating intermediate question(s) and recursively calling FAITH for answering them

Intermediate questions for resolving implicit constraint	**Explicit temporal values**
"When Antoine Raab managed FC Nantes start date?"	*1946*
"When Antoine Raab managed FC Nantes end date?"	*1949*

> ***Question entity:*** *Antoine Raab, FC Nantes*
> ***Question relation:*** After managing which football club did take on next
> ***Expected answer type:*** association football club
> ***Temporal signal:*** after
> ***Temporal category:*** implicit
> ***Temporal value:*** ***[1946, 1949]***
> **Structured Frame**

Retrieval

Heterogeneous Retrieval from Wikidata KB, Wikipedia text, tables and infoboxes → **Evidences**

Temporal Pruning to prune out temporally irrelevant evidence

Evidence Scoring to prune out semantically irrelevant evidence

Top-*k* Faithful Evidences   **Structured Frame**

Answering

Graph Construction → GNN ... (Iteration 1) → GNN (Iteration i-1) → GNN (Iteration i)

[EXPLAIGNN, Christmann et al., SIGIR 2023]
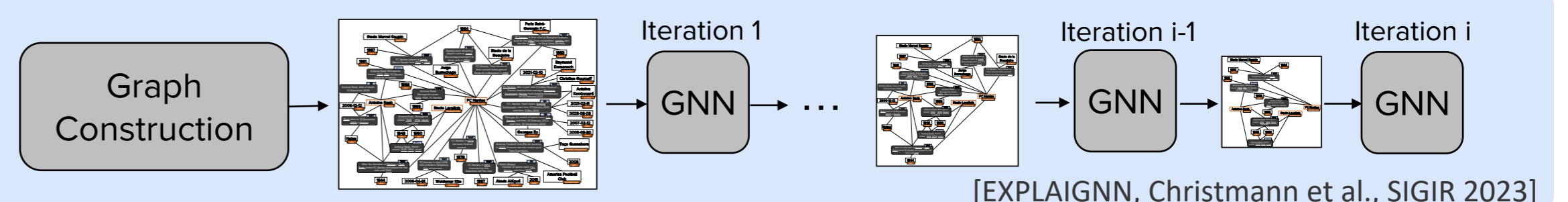
Stade Lavallois

Supporting evidence for answers of intermediate questions (*1946* and *1949*)

- "*Antoine Raab, managerial career, 1946-1949, FC Nantes*" [Wikipedia infobox]
- "*Antoine Raab, after the liberation of Nantes in 1944, Raab joined FC Nantes and played for the club until 1949*" [Wikipedia text]

Supporting evidence for final answer (**Stade Lavallois**)

- "*Antoine Raab, managerial career, 1949-1950, **Stade Lavallois**"* [Wikipedia infobox]

FAITH outperforms state-of-the-art on implicit questions, providing more faithful answers

🕐 TIQ Benchmark Construction
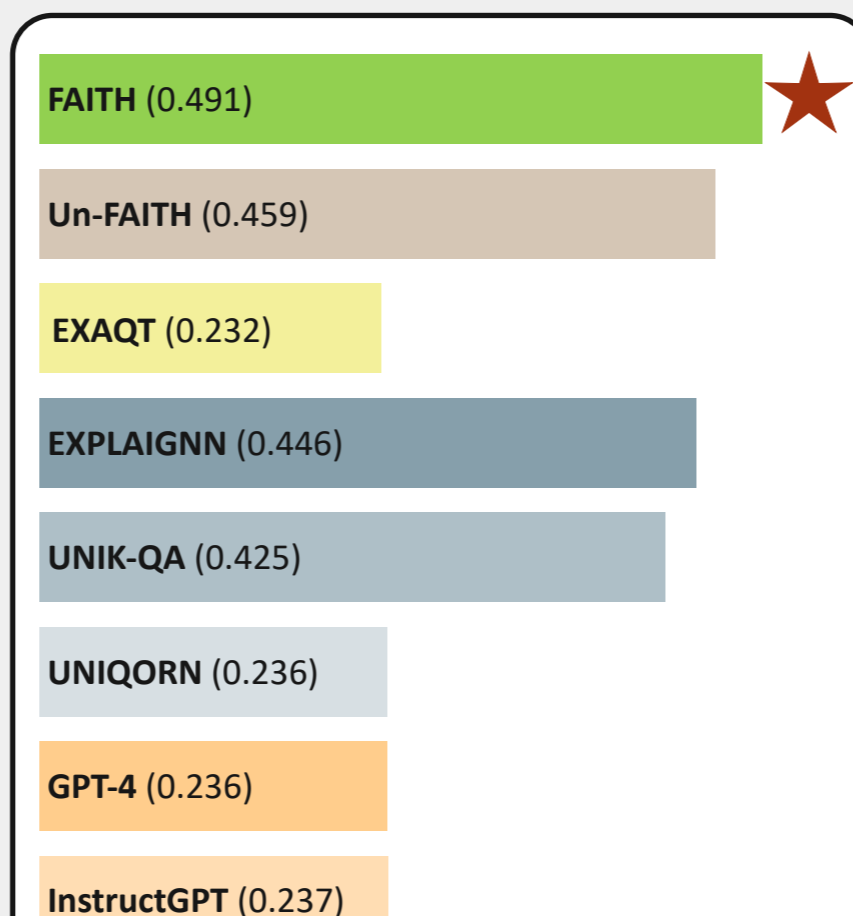
★ Automatic construction of implicit questions

★ Questions derived from multiple sources: Wikipedia text, infoboxes, and Wikidata KB

★ Configurable among various dimensions

🕐 TIQ Benchmark Characteristics

★ 10,000 implicit questions

★ 2,542 long-tail entities (<20 KB-facts)

★ 2,613 prominent entities (>500 KB-facts)

★ 2.45 question entities per question

★ 17.96 words per question

P@1 on 🕐 TIQ
(P@1: Precision at 1)

- FAITH (0.491) ★
- Un-FAITH (0.459)
- EXAQT (0.232)
- EXPLAIGNN (0.446)
- UNIK-QA (0.425)
- UNIQORN (0.236)
- GPT-4 (0.236)
- InstructGPT (0.237)

Faithfulness Evaluation

Comparing the faithfulness of FAITH and Un-FAITH for correct answers, and how often temporal constraints are violated or ignored.

Benchmark →	🕐 TIQ		TimeQuestions	
Method ↓	Faithful	Temporally Unfaithful	Faithful	Temporally Unfaithful
FAITH ★	0.95	0.00	0.94	0.01
Un-FAITH	0.90	0.08	0.87	0.13

Example: What movies starring Taylor Lautner *in 2011*? → *Abduction*

Evidence provided by FAITH (faithful):

- "*Taylor Lautner, Year is 2011, Title is **Abduction**, Role is Nathan Harper*" [Wikipedia table]

Evidence provided by Un-FAITH (temporal constraint not matched → temporally unfaithful):

- "***Abduction**, cast member, Taylor Lautner*" [Wikidata KB]