Case Study: Natural Language Interface for Patients' Electronic Health Records (EHR)

Revanth Korrapolu
Madhumitha Sivaraj

3,243

Number of medication errors linked to EHR-usability issues at three pediatric hospitals from 2012 to 2017

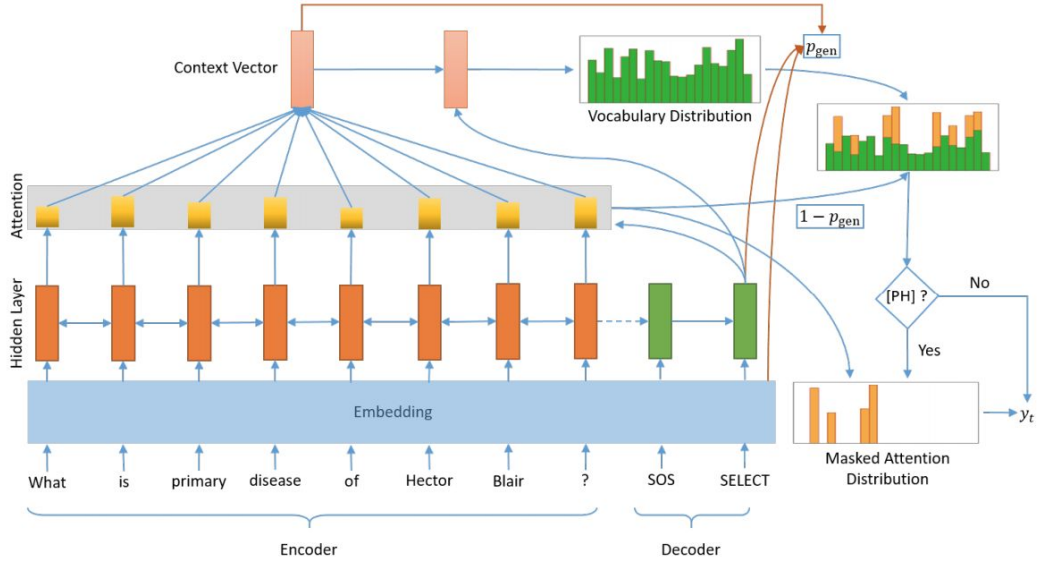
5.9 HOURS

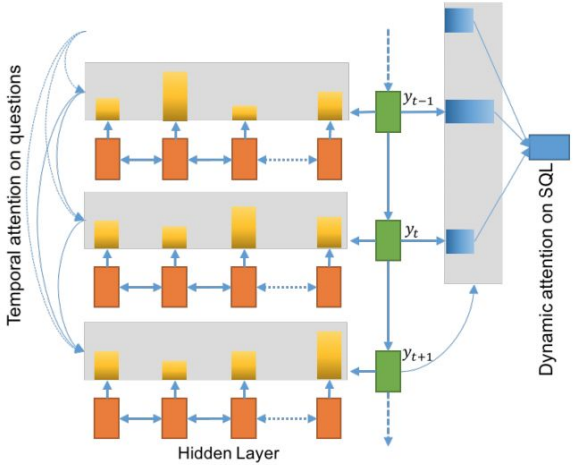
Average time doctors spend on EHRs per day

4,000

Clicks per a shift on EHR

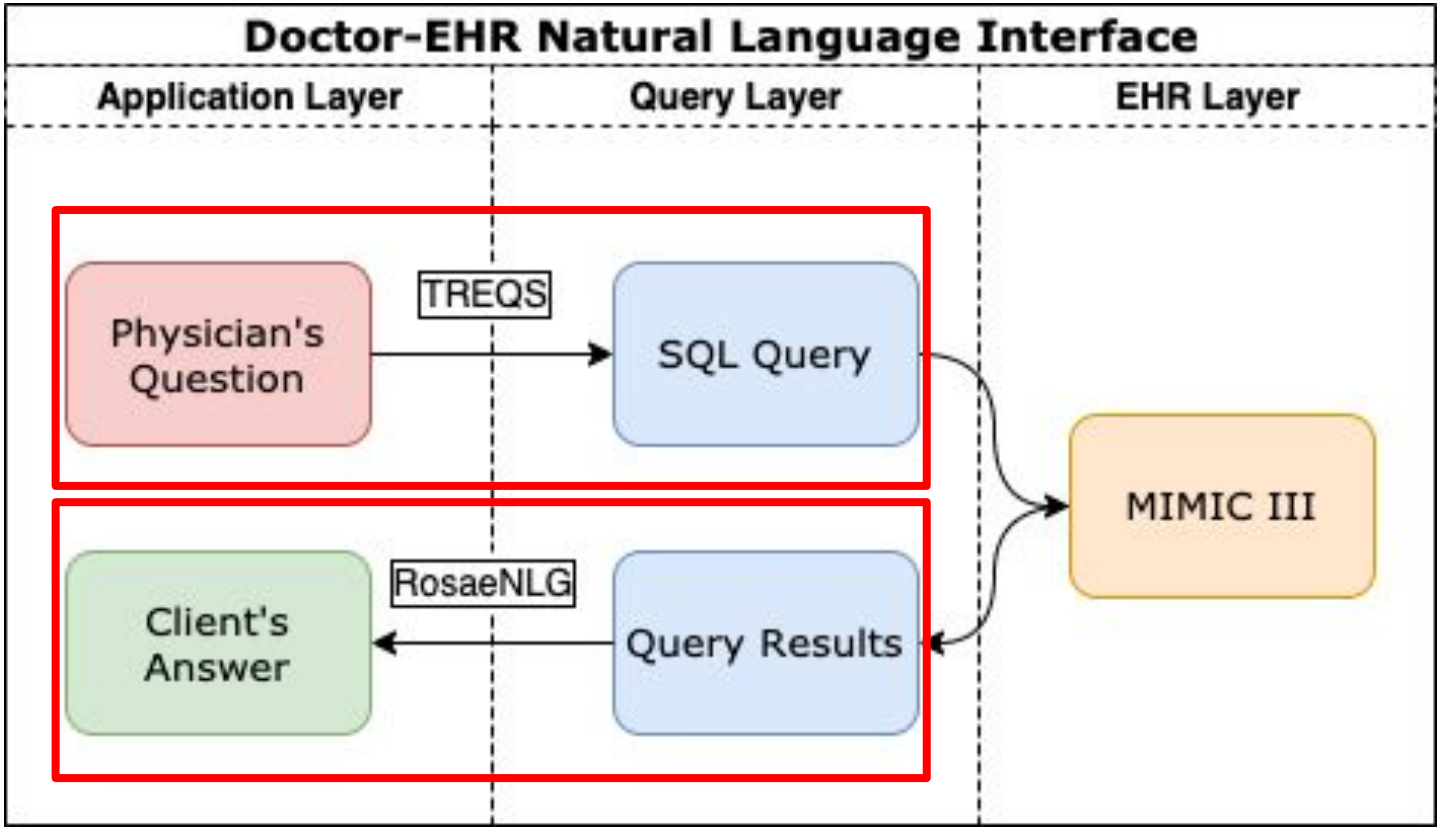
Motivation: TRanslate-Edit Model for Question-to-SQL (TREQS)





* from TREQS: Ping Wang, Tian Shi, Chandan Reddy

Question	Generated SQL
provide the number of patients less than 83 years of age who were diagnosed with pneumococcal pneumonia .	select count (distinct demographic.subject_id) from demographic inner join diagnoses on demographic.hadm_id = diagnoses.hadm_id where demographic.age < 83 and diagnoses.short_title = pneumococcal pneumonia
provide the number of patients whose diagnoses is lap surg convert to open and they are female .	select count (distinct demographic.subject_id) from demographic inner join diagnoses on demographic.hadm_id = diagnoses.hadm_id where demographic.gender = f and diagnoses.long_title = lap surg convert to open they
among patients treated with amitriptyline , calculate the number of female patients .	select count (distinct demographic.subject_id) from demographic inner join prescriptions on demographic.hadm_id = prescriptions.hadm_id where demographic.gender = f and prescriptions.drug = amitriptyline
get me the number of elective hospital admission patients who had coronary artery primary disease .	select count (distinct demographic.subject_id) from demographic where demographic.admission_type = elective and demographic.diagnosis = coronary artery disease
give the number of patients whose admission type is elective and primary disease is abdominal abscess .	select count (distinct demographic.subject_id) from demographic where demographic.admission_type = elective and demographic.diagnosis = abdominal abscess
how many patients aged below 36 years have stayed in the hospital for more than 14 days ?	select count (distinct demographic.subject_id) from demographic where demographic.age < 36 and demographic.days_stay > 14
what is the number of patients whose admission location is emergency room admit and with primary disease t5 fracture ?	select count (distinct demographic.subject_id) from demographic where demographic.admission_location = emergency room admit and demographic.diagnosis = t5 fracture


Doctor-EHR Natural Language Interface
Application Layer
Query Layer
EHR Layer
Physician's Question
TREQS
SQL Query
MIMIC III
Client's Answer
RosaeNLG
Query Results

Step 1: Natural Language Question

Ex: Who are male smokers over the age 45?

Step 2: Translate into SQL

Method	Example 1	Example 2
Question	how many female patients underwent the procedure of abdomen artery incision?	how many patients admitted in emergency were tested for ferritin?
Ground truth	**select** count (distinct demographic."subject_id") **from** demographic inner join procedures on demographic.hadm_id = procedures.hadm_id **where** demographic."gender" = "f" and procedures."short_title" = "abdomen artery incision"	**select** count (distinct demographic."subject_id") **from** demographic inner join lab on demographic.hadm_id = lab.hadm_id **where** demographic."admission_type" = "emergency" and lab."label" = "ferritin"
M-SQLNET	**select** count (distinct demographic."subject_id") **from** demographic inner join procedures on demographic.hadm_id = procedures.hadm_id **where** demographic."gender" = "f" and procedures."short_title" = "parent infus nutrit sub"	**select** count (distinct demographic."subject_id") **from** demographic inner join lab on demographic.hadm_id = lab.hadm_id **where** demographic."admission_type" = "emergency" and lab."label" = "po2"
Seq2Seq	**select** count (distinct demographic."subject_id") **from** demographic inner join procedures on demographic.hadm_id = procedures.hadm_id **where** demographic."gender" = "m" and procedures."long_title" = "other abdomen"	**select** count (distinct demographic."subject_id") **from** demographic inner join lab on demographic.hadm_id = lab.hadm_id **where** demographic."admission_location" = "phys referral/normal deli" and lab."itemid" = "ferritin"
Seq2Seq+recover	**select** count (distinct demographic."subject_id") **from** demographic inner join procedures on demographic.hadm_id = procedures.hadm_id **where** demographic."gender" = "m" and procedures."long_title" = "other bronchoscopy"	**select** count (distinct demographic."subject_id") **from** demographic inner join lab on demographic.hadm_id = lab.hadm_id **where** demographic."admission_location" = "phys referral/normal deli" and lab."itemid" = "51200"
PtrGen	**select** count (distinct demographic."subject_id") **from** demographic inner join procedures on demographic.hadm_id = procedures.hadm_id **where** demographic."gender" = "f" and procedures."long_title" = "spinal abdomen artery"	**select** count (distinct demographic."subject_id") **from** demographic inner join lab on demographic.hadm_id = lab.hadm_id **where** demographic."admission_type" = "emergency" and lab."label" = "troponin i"
PtrGen+recover	**select** count (distinct demographic."subject_id") **from** demographic inner join procedures on demographic.hadm_id = procedures.hadm_id **where** demographic."gender" = "f" and procedures."long_title" = "spinal tap"	**select** count (distinct demographic."subject_id") **from** demographic inner join lab on demographic.hadm_id = lab.hadm_id **where** demographic."admission_type" = "emergency" and lab."label" = "troponin i"
TREQS	**select** count (distinct demographic."subject_id") **from** demographic inner join procedures on demographic.hadm_id = procedures.hadm_id **where** demographic."gender" = "f" and procedures."short_title" = "abdomen artery abdomen"	**select** count (distinct demographic."subject_id") **from** demographic inner join lab on demographic.hadm_id = lab.hadm_id **where** demographic."admission_type" = "emergency" and lab."label" = "ferritin"

*Chart from TREQS: Ping Wang, Tian Shi, Chandan Reddy

Step 3: Query Database

MIMIC-III

List of tables

The following tables are used to define and track patient stays:

- **ADMISSIONS**: Every unique hospitalization for each patient in the database (defines `HADM_ID`)
- **CALLOUT**: Information regarding when a patient was cleared for ICU discharge and when the patient was actually discharged
- **ICUSTAYS**: Every unique ICU stay in the database (defines `ICUSTAY_ID`)
- **PATIENTS**: Every unique patient in the database (defines `SUBJECT_ID`)
- **SERVICES**: The clinical service under which a patient is registered
- **TRANSFERS**: Patient movement from bed to bed within the hospital, including ICU admission and discharge

MIMICSQL

Physicians Questions → SQL QUERY

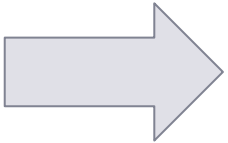
Q: "Who are male smokers over the age 45?",

"sql": "SELECT DISTINCT(p.id), p.resource#>'{name,0}' FROM patient p JOIN observation o ON o.resource#>>'{subject,id}' = p.id WHERE (o.resource @> '{"code": {"coding": [{"code": "72166-2"}]}}'::jsonb) — LOINC: Smoking status AND ((o.resource @> '{"value": {"CodeableConcept": { "coding": [{"code": "449868002"}]}}}') — SNOMED: Current every day smoker OR (o.resource @> '{"value": {"CodeableConcept": { "coding": [{"code": "8517006"}]}}}')) — SNOMED: Former smoker AND (extract(year from age(now(), (p.resource->>'birthDate')::date)) > 45) — patient's age > 45 AND (p.resource->>'gender' = 'male'); — patient gender is male"

Step 4: NLG of Query Results

SQL Output:


```
{  "use": "official",
 "given": [
  "Adolfo777"
 ],
 "family": "Stokes453",
 "prefix": [
  "Mr."
 ]
},
{ "use": "official",
 "given": [
  "Alvaro283"
 ],
 "family": "Zulauf375",
 "prefix": [
  "Mr."
 ]
}
```

Natural Language Output:

```
madhu@Madhumithas-MBP templates_malesmokers % npx rosaenlg-cli -l en_US -w male_smokers_over45.pug

  watching male_smokers_over45.pug
The male smokers over the age 45 are Adolfo777 Stokes453 and Alvaro283 Zulauf375.
```

Conclusion

- Execution Accuracy
- Logic Form Accuracy

$$Acc_{EX} = N_{EX}/N$$

$$Acc_{LF} = N_{LF}/N$$

	Template Questions				NL Questions			
Method	**Development**		**Testing**		**Development**		**Testing**	
TREQS + recover	**0.853**	**0.924**	**0.912**	**0.940**	**0.562**	**0.675**	**0.556**	**0.654**

- NLG Efficacy = limited

Questions