Stanford CS224v Course
Conversational Virtual Assistants with Deep Learning

Lecture 11

Satisfying Natural Language Constraints with SMT

A Case Study in Clinical Trials Matching

Monica Lam and Cyrus Zhou

Large-Scale Constraint Satisfaction

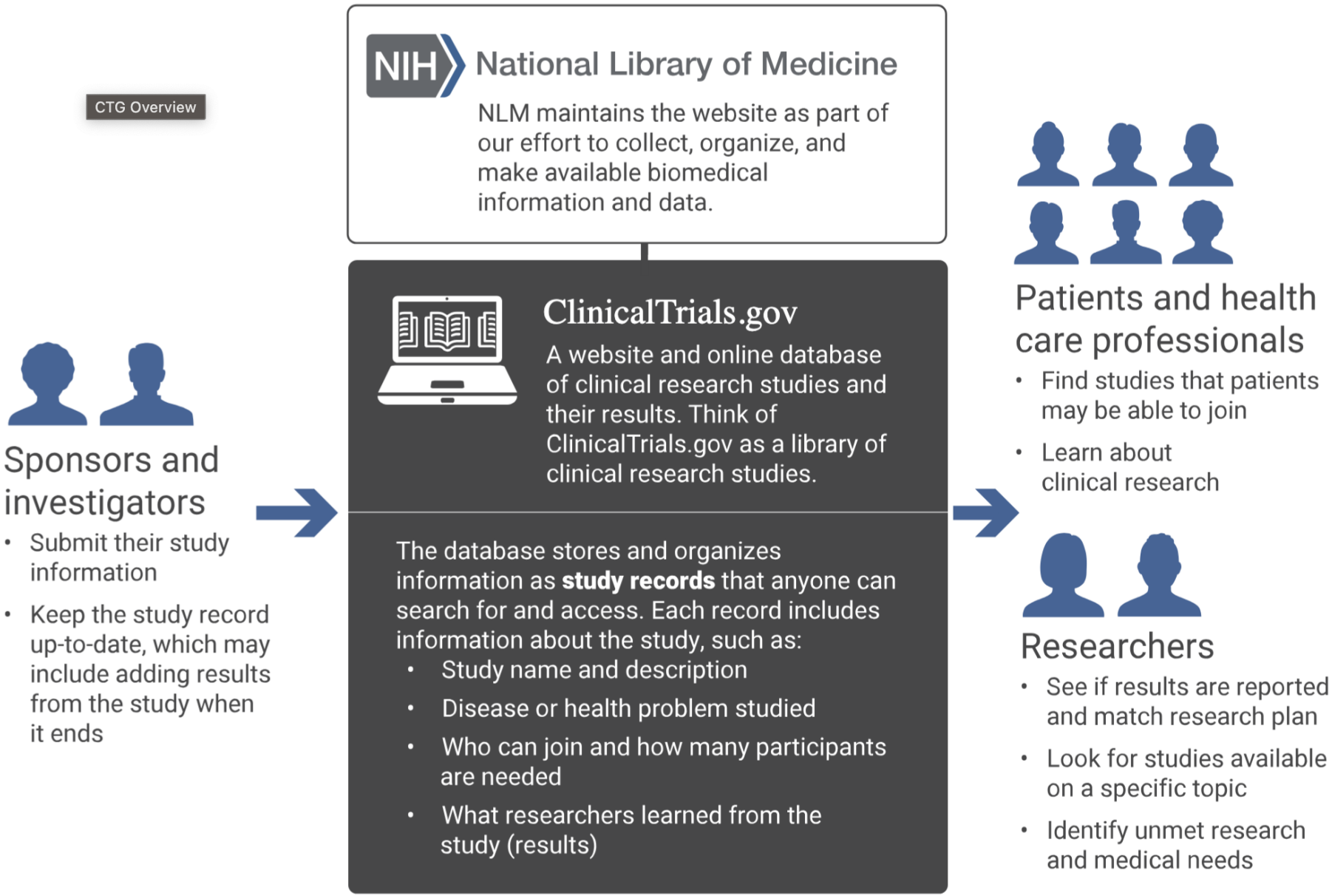
Many applications

- Education: Degree satisfaction
- Finance: Government regulations
- Medical: Clinical Trial matching

Why is Clinical Trial Matching Important?

- Clinical trials are essential to modern medical progress.
  - Test out new treatments
    (drugs, procedures, behavioral therapies)
- Nearly **$1.9 billion** is spent annually on recruitment efforts
- But still,
  around **80%** of trials fail to meet the initial enrollment target/timeline
  - Causing monetary loss
  - Blocking scientific progress

Trials are in ClinicalTrials.gov!



ClinicalTrials.gov are Used Widely

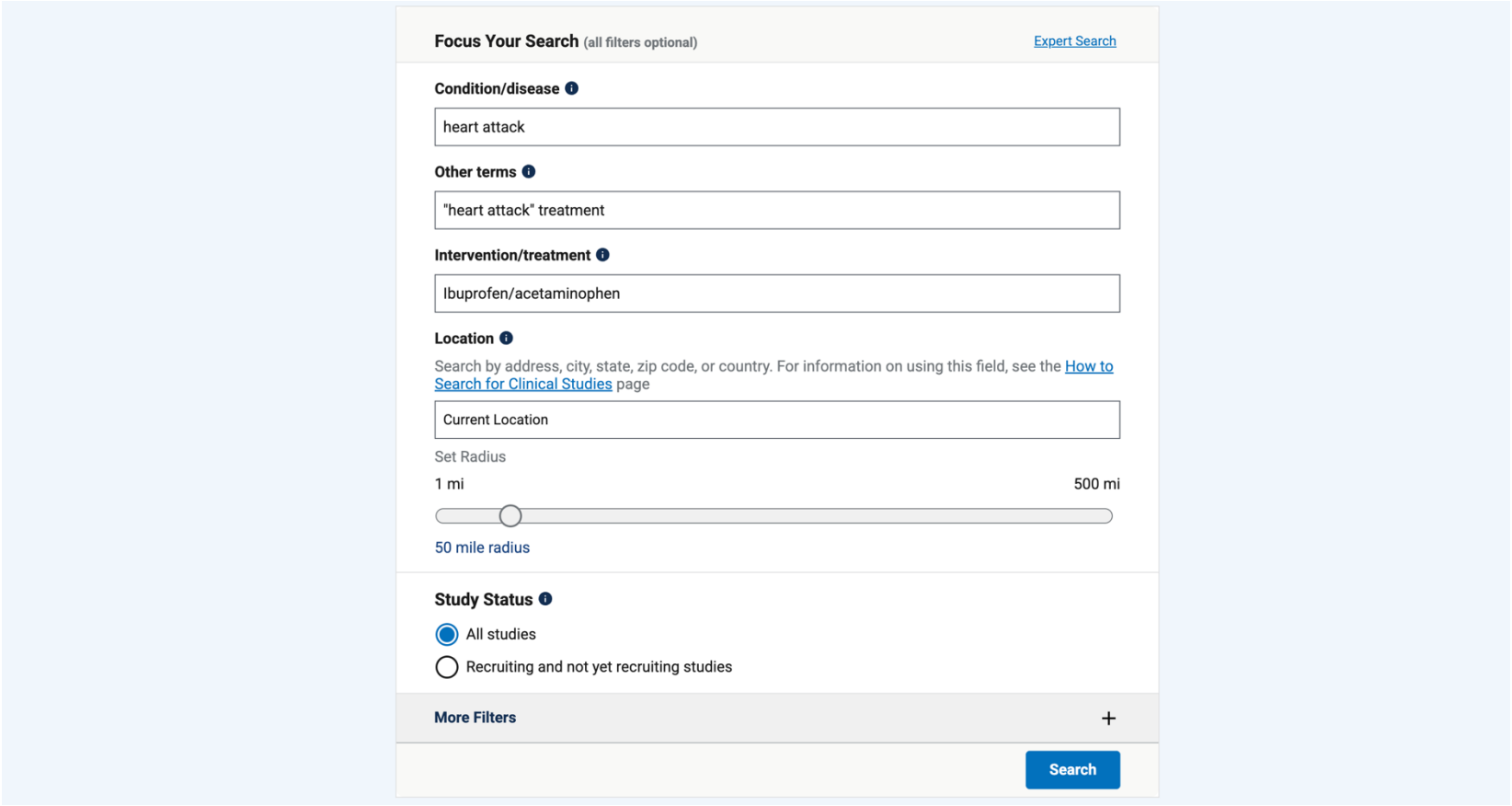
- 485,000 studies listed from all 50 states and 223 countries (3/2024)
  - Good coverage of trials: e.g.
    NIH-funded clinical trials are expected to register and submit results

- ClinicalTrials.gov receives about 4.5 million visitors monthly.

https://www.bumc.bu.edu/ohra/clinicaltrials-gov/clinicaltrials-gov-what-why-which-studies-when

What is the Current Practice?

- Doctor referring patients to trials that they know of
- Patients looking for trials ..



Motivation for Our Project

"I'm a Physician,
and I can't Find a Trial for Myself"

Let's Take a Look at a Clinical Trial

Study Overview

Brief Summary

The purpose of the study is to assess the frequency and intensity of intramyocardial haemorrhage in patients with primary STEMI and different reperfusion strategies.

Detailed Description

The study non-randomized, opened, controlled. In half of patients despite on carried in-time reperfusion therapy intramyocardial haemorrhage determined after a long-term period of severe ischemia. Earlier, definition of intramyocardial haemorrhage was possible only by autopsy. Nowaday, cardiac contrast MRI is the best diagnostic method, which allows to assess the regional and global function of the LV, structural changes in myocardial tissue and also in T2 mode it became assessable to reveal intramyocardial haemorrhage.

Taking into account the results of previous researches, it can be concluded that the intramyocardial haemorrhage was determined in half of patients with primary PCI [1]. An influence of fibrinolytic therapy to the intramyocardial haemorrhage was conducted in small group of patients in one trial, and therefore further data will be actual and useful [2].

It is planned to study 60 patients with primary STEMI using standard therapy. The patients will be divided into 2 groups. Patients of the 1st group will be treated by pharmaco-invasive strategy. The 2nd group will be treated by primary PCI. Patients in all groups after reperfusion strategies will be conducted cardiac contrast MRI for detection intramyocardial haemorrhage within 2 days onset. At day 7 and through 3 months, the clinical condition of the patients will be assessed and cardiac ultrasound will be performed for the evaluation of myocardial contractile function and 2D global longitudinal strain. Also, the incidence rate of secondary endpoints will be evaluated.

Study Start (Actual)

2018-01-25

Primary Completion (Actual)

2018-10-30

Study Completion (Actual)

2019-03-20

Enrollment (Actual)

60

Study Type

Interventional

Phase

Not Applicable

Eligibility Criteria

Description

Inclusion Criteria:

- Age ≥ 18 years at time of randomization (18 years and older);
- Acute myocardial infarction;
- Reperfusion of the infarct-related coronary artery in terms within 12 h of symptom onset;
- Written the informed consent to participate in research;

Exclusion Criteria:

- Inability to obtain informed consent;
- Patients previously undergone endovascular / surgical revascularization of coronary artery;
- Severe comorbidity;
- History of myocardial infarction;
- History of intracranial haemorrhage;
- Pulmonary edema, cardiogenic shock;
- Creatinine clearance <30 mL/min or dialysis;
- Unable to undergo or contra-indications for MRI;
- Allergy for contrast agent;
- Indication or use of oral anticoagulant therapy;
- Major bleeding;
- Atrio-ventricular block II and III degree;
- Active gastroduodenal ulcer;
- Aortic dissection;
- Acute psychotic disorders

Ages Eligible for Study

18 Years and older (Adult, Older Adult)

Sexes Eligible for Study

All

Accepts Healthy Volunteers

No

What does a Real Patient Record Look Like?

Hospital Medicine — Admission H&P

Patient: J.D. (F, 52) • **MRN:** 1××××× • **Room:** 6W-6213
Date/Time: 2025-09-21 18:42 PT • **Admitting Service:** Hospital Medicine-B
Attending: Alex Kim, MD • **Resident:** Priya Shah, MD • **PCP:** Unknown
Code Status: Full
Isolation: None
Language: English
Insurance: Commercial PPO

Chief Complaint

"Chest pressure and nausea."

History of Present Illness

52-year-old female with HTN and HLD presenting with acute central chest pressure starting ~12:30 PT while at rest, radiating to L arm, associated with diaphoresis and nausea, no syncope. Pain 8/10, non-pleuritic, non-positional. EMS EKG reportedly ST elevations anteriorly; ASA given by EMS? **unclear**. In ED: EKG with 2 mm STE V2–V4; initial trop-I 0.42 ng/mL ↑. Given ASA 325 mg PO, ticagrelor 180 mg PO, heparin 4,000 U IV bolus. Cath lab activated; patient transferred for emergent PCI at 13:35.

Cath Lab (brief op note): Proximal LAD 99% thrombotic lesion → drug-eluting stent ×1 with TIMI-3 flow restoration. LVEDP mildly elevated. No complications. Total contrast 90 mL. Fluoro time 12.8 min.

Post-PCI pain resolved (0–1/10). Admitted to CCU then transferred to floor on 6W at 18:10 in stable condition on DAPT.

Prior episodes: Intermittent exertional chest tightness over last 2–3 months climbing stairs, self-limited; never evaluated.

Precipitating factors today: At rest; possible recent work stress; slept 4–5 h/night for past week. No recent illness, immobilization, or surgery.

ED Course (chronological): See nursing and ED MD notes copied below under *Copied Forward Content.*

Problem List (active)

1. **Acute anterior STEMI s/p PCI to proximal LAD** (09/21/2025)
2. **Hypertension** (2016)
3. **Hyperlipidemia** (2018)
4. **Overweight** (BMI 29)

Allergies

- **Penicillin** — rash (non-anaphylactic)

Home Medications (patient report, med reconciliation completed by pharmacy)

- Lisinopril 10 mg PO daily (last fill 2 months ago; reports good adherence)
- Atorvastatin 40 mg PO nightly
- OTC: multivitamin daily
- Denies herbals/supplements

Social History / SDOH

- Lives with spouse in apartment (elevator access); independent in ADLs/IADLs.
- Work: retail manager; high stress; on feet most of day.
- Tobacco: never.
- Alcohol: 1–2 drinks/week.
- Illicits: denies.
- Transportation: drives.
- Food security/housing security: stable.
- Prefers afternoon appointments due to work schedule.

Family History

- Father MI at 58; deceased at 72 from stroke.
- Mother T2DM, HTN.

Review of Systems (template — positives in bold)

- **Constitutional: diaphoresis** at onset; no fever/chills, no weight loss.
- **Cardiac: chest pressure**, **radiation L arm**; no palpitations; no LE edema.
- **Respiratory:** no SOB at rest, no cough, no wheeze.
- **GI: nausea**; no vomiting/diarrhea, no melena/hematochezia.
- **GU:** no dysuria, no hematuria.
- **Neuro:** no syncope, no focal weakness, no speech changes.
- **MSK:** no calf pain, no recent trauma.
- **Derm:** no rashes currently.
- **Endo:** no heat/cold intolerance.
- **Psych:** baseline stress; no SI/HI.

Auto-imported ROS block (system default) — 14-pt ROS negative except as noted above.

Vitals (selected)

ED Triage (12:48): BP 162/94, HR 104, RR 20, Temp 36.8 °C, SpO_2 98% RA, Wt 78 kg, Ht 164 cm (BMI 29.0)
Post-PCI (15:00): BP 138/82, HR 88, RR 16, SpO_2 98% RA
On transfer to floor (18:15): BP 134/79, HR 82, RR 16, Temp 36.7 °C, SpO_2 98% RA

Intake/Output (last 24 h): Intake 450 mL; Output 300 mL; Net +150 mL.

Physical Exam (on floor 18:25)

- **General:** alert, oriented, comfortable, no distress.
- **HEENT:** PERRL, anicteric; MMM.
- **Neck:** supple; no JVD at 30°.
- **CV:** RRR, no murmurs/rubs/gallops.
- **Lungs:** CTAB, normal WOB.
- **Abd:** soft, NT/ND, +BS.
- **Ext:** warm, well-perfused; no edema; R radial access site clean/dry/intact with TR band removed at 17:30; distal pulses 2+.
- **Neuro:** AO×3, non-focal; speech clear; moves all extremities.
- **Skin:** no rash.

Labs (auto-imported)

CBC
09/21 13:10 — WBC 8.2, Hgb 13.6, Hct 40.3, Plt 248
BMP
09/21 13:10 — Na 139, K 4.2, Cl 103, CO_2 24, BUN 15, Cr 0.86, Glu 146
Mg/Phos
09/21 13:10 — Mg 1.9, Phos 3.4
Cardiac markers
09/21 13:10 — Troponin-I **0.42** (↑)
09/21 17:40 — Troponin-I **2.3** (↑ ↑)
Lipid panel (fasting pending) ordered for 09/22 06:00
HbA1c ordered for 09/22 06:00

> **Note:** Full lab history available in chart. Above values auto-pulled; see results tab for reference ranges.

Microbiology

- None pending. COVID/flu not indicated.

Imaging / Studies

EKG (ED 13:04): Sinus tachy 102. 2 mm ST elevations V2–V4 with reciprocal depressions II/III/aVF.

Clinical Matching is Complex

1. Basic matching problem: Is a patient eligible for a trial?
   - LLMs seem to do a reasonable job for most cases

2. From a pharmaceutical company's perspective:
   - Recruit from **millions** of patients for a trial
   - How to match against millions?

3. From a patient's perspective
   - Find eligible trials among **500K clinical trials**

Key Questions: How accurate is the LLM approach?
How to handle the scale?

Outline

- Motivation
- **Prior Research**
- SMT Approach
- SMT-Based Matching Algorithm
- Preliminary Results

LLM-Based Matching

- TrialGPT: Matching patients to clinical trials with LLMs,
  Jin et al., Nov, 2024, Nature Communications
  - Synthetic patients
- Zero-Shot Clinical Trial Patient Matching with LLMs
  Wornow et al., Dec. 2024, New England Journal of Medicine
  - Automatic qualitative coding from patient records
- Cohort Discovery: A Survey on LLM-Assisted Clinical Trial Recruitment
  Ghosh, et al. June 2025, arXiv:2506.15301v1

Datasets

TrialGPT dataset

- Koopman, B. & Zuccon, G.
  A test collection for matching patients to clinical trials. In Proc. 39th International ACM SIGIR Conference on Research and Development in Information Retrieval 669-672 (2016).
- Roberts, K., Demner-Fushman, D., Voorhees, E. M., Bedrick, S. & Hersh, W. R.
  Overview of the TREC 2022 Clinical Trials Track. In Proc. Thirty-First Text REtrieval Conference (TREC 2022) (2022).
- Roberts, K., Demner-Fushman, D., Voorhees, E. M., Bedrick, S. & Hersh, W. R.
  Overview of the TREC 2021 Clinical Trials Track. In Proc. Thirtieth Text REtrieval Conference (TREC 2021) (2021).

Warnow's paper datasets

- Stubbs, et al.,,
  Cohort selection for clinical trials: n2c2 2018 shared task track 1.
  Journal of the American Medical Informatics Association, 26(11):1163–1171, 2019

nature communications

Article

https://doi.org/10.1038/s41467-024-53081-z

Matching patients to clinical trials with large language models

Received: 18 January 2024

Accepted: 1 October 2024

Published online: 18 November 2024

Qiao Jin[1], Zifeng Wang[2], Charalampos S. Floudas[3], Fangyuan Chen[4], Changlin Gong[5], Dara Bracken-Clarke[3], Elisabetta Xue[3], Yifan Yang[1,6], Jimeng Sun[2] & Zhiyong Lu[1]

TrialGPT Dataset

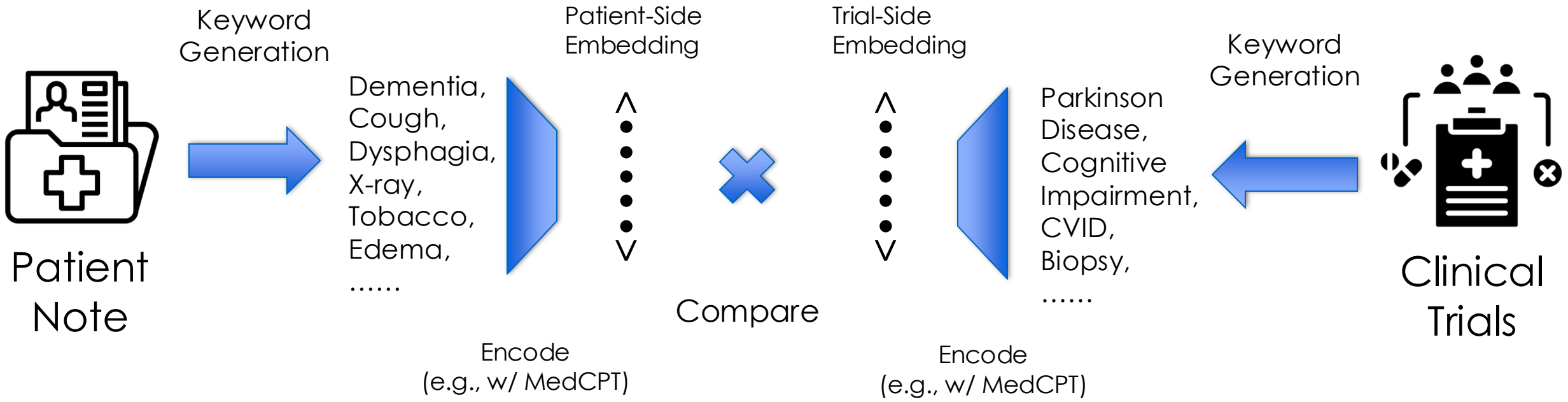
- Number of trials: 75,000
- Synthesized patient notes: 183

Cohort	SIGIR	TREC 2021 CT	TREC 2022 CT
N	58	75	50
Age (year)	38.5 ± 23.7	41.6 ± 19.4	35.3 ± 20.2
Sex (male: female)	29: 29	38: 37	28: 22
Note length (words)	88.7 ± 36.8	156.2 ± 45.4	109.9 ± 21.6
Eligible trials/patient	7.3 ± 6.7	74.3 ± 49.0	78.8 ± 67.3
Potential trials/ patient	11.7 ± 10.2	None	None
Excluded trials/ patient	None	80.3 ± 60.3	60.7 ± 65.5
Irrelevant trials/ patient	47.1 ± 19.5	323.2 ± 93.2	568.4 ± 164.1
Considered initial trials	3621	26149	26581

We show the mean ± standard deviation for applicable variables. "None" denotes there is no such eligibility label in the corresponding cohort. SIGIR: the patient-trial matching cohort published at the Special Interest Group on Information Retrieval (SIGIR).
TREC the Text REtreival Conference (TREC), *CT* the clinical trials track at TREC.

TrialGPT: Finding Trials for Patients

1. Retrieval based on embedding similarity



2. Matching: Use LLM to patch patient record with retrieved trials

Recall of Dense Retrieval

	Recall	#Retrieved
SIGIR	93%	500
TREC 2021	92%	1000
TREC 2022	88%	1000

What if the life-saving trial is not retrieved?

False negatives can be fatal

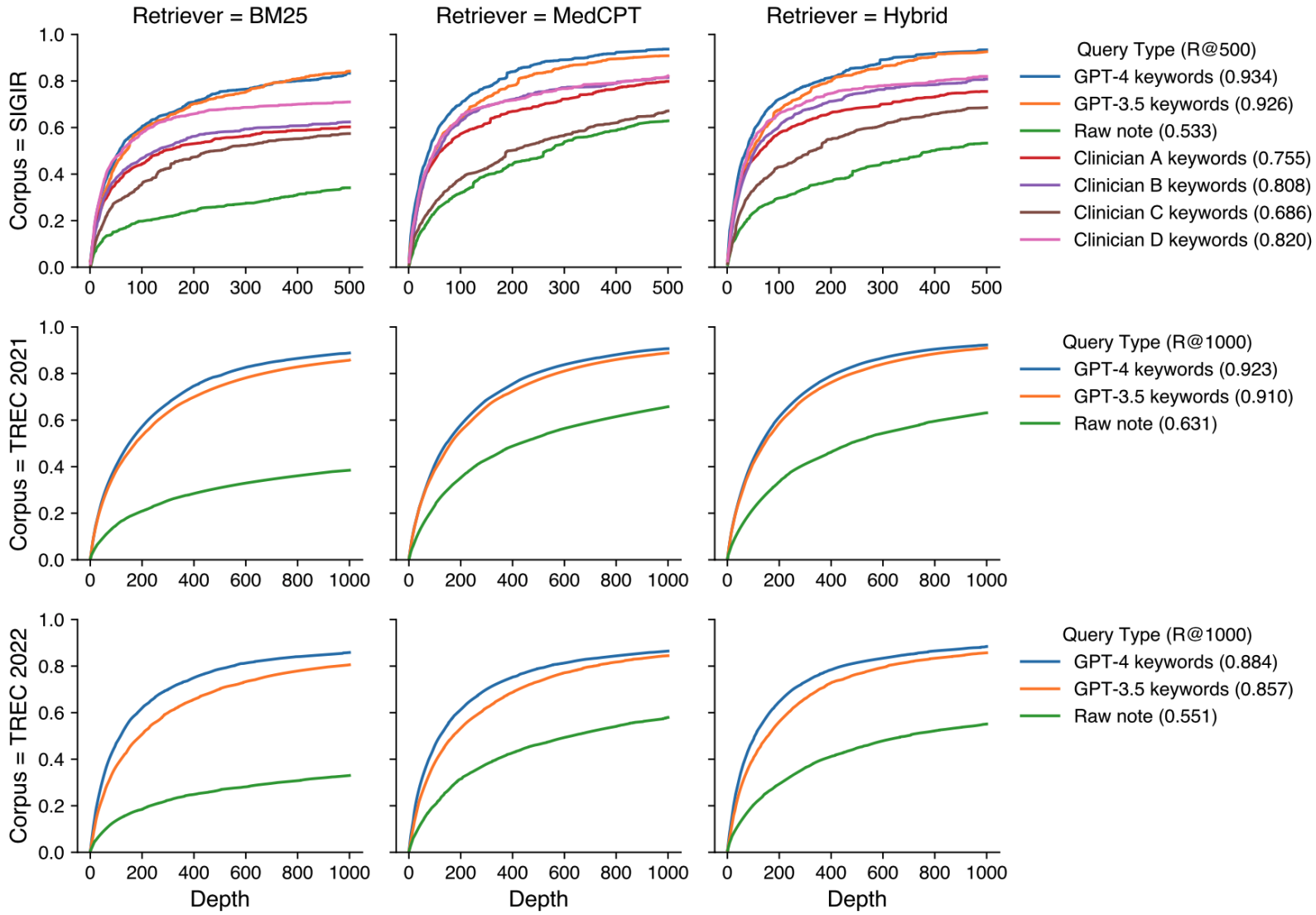
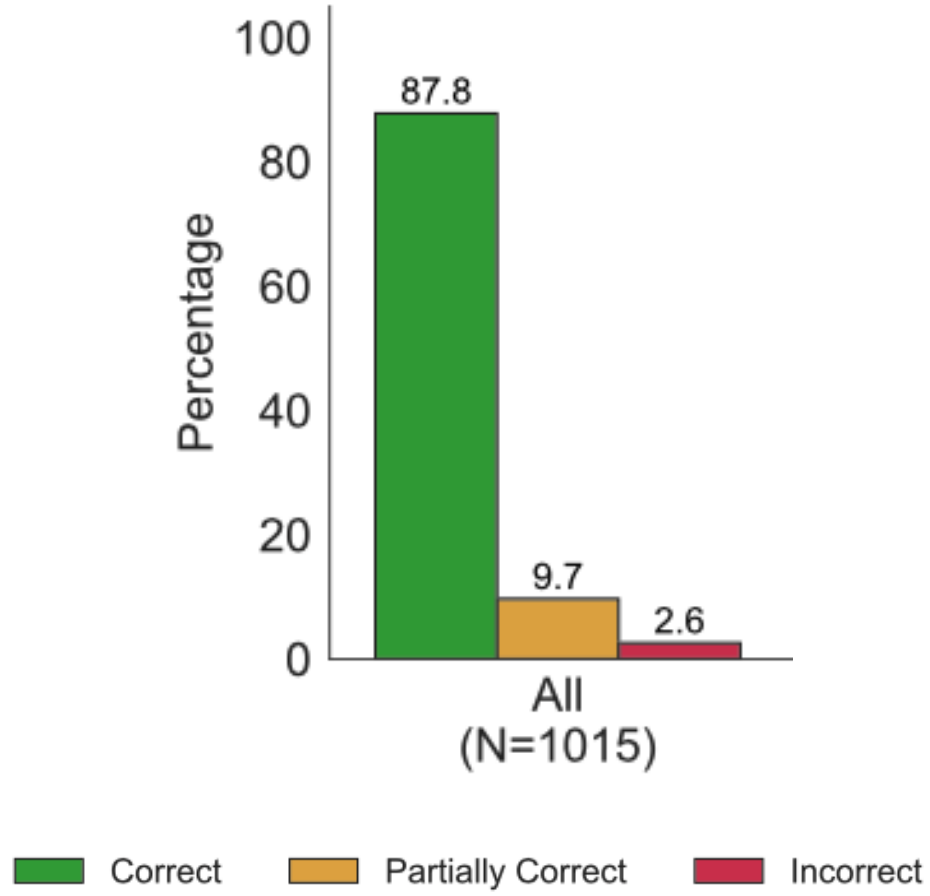


Fig. 2 | First-stage retrieval results. a Overview of TrialGPT-Retrieval. LLMs first generate a list of keywords for a given patient note. These keywords are used to derive the keyword-level relevant clinical trials, which are then fused to generate a final ranking. **b** Recalls of relevant clinical trials at different depths for various query types and retrievers. The hybrid retriever combines the results of the BM25 (lexical matching) and the MedCPT (semantic matching) retrievers. Source data are provided as a Source Data file.

Precision of Dense Retrieval
(Embedding-Similarity Based)

- Imprecise – many false positives
- Based on "Semantic Similarity" not "Logical Compatibility"
- E.g. If a patient has diabetes
  - Retrieved trials:
    - Those that *include* diabetes
    - *Those that exclude* diabetes

LLM's Precision in Matching



- Is LLM precise in the match?
- Human evaluators also differ

Summary of Errors

- Dense retrieval
  may have false negatives
- LLM matching may be erroneous

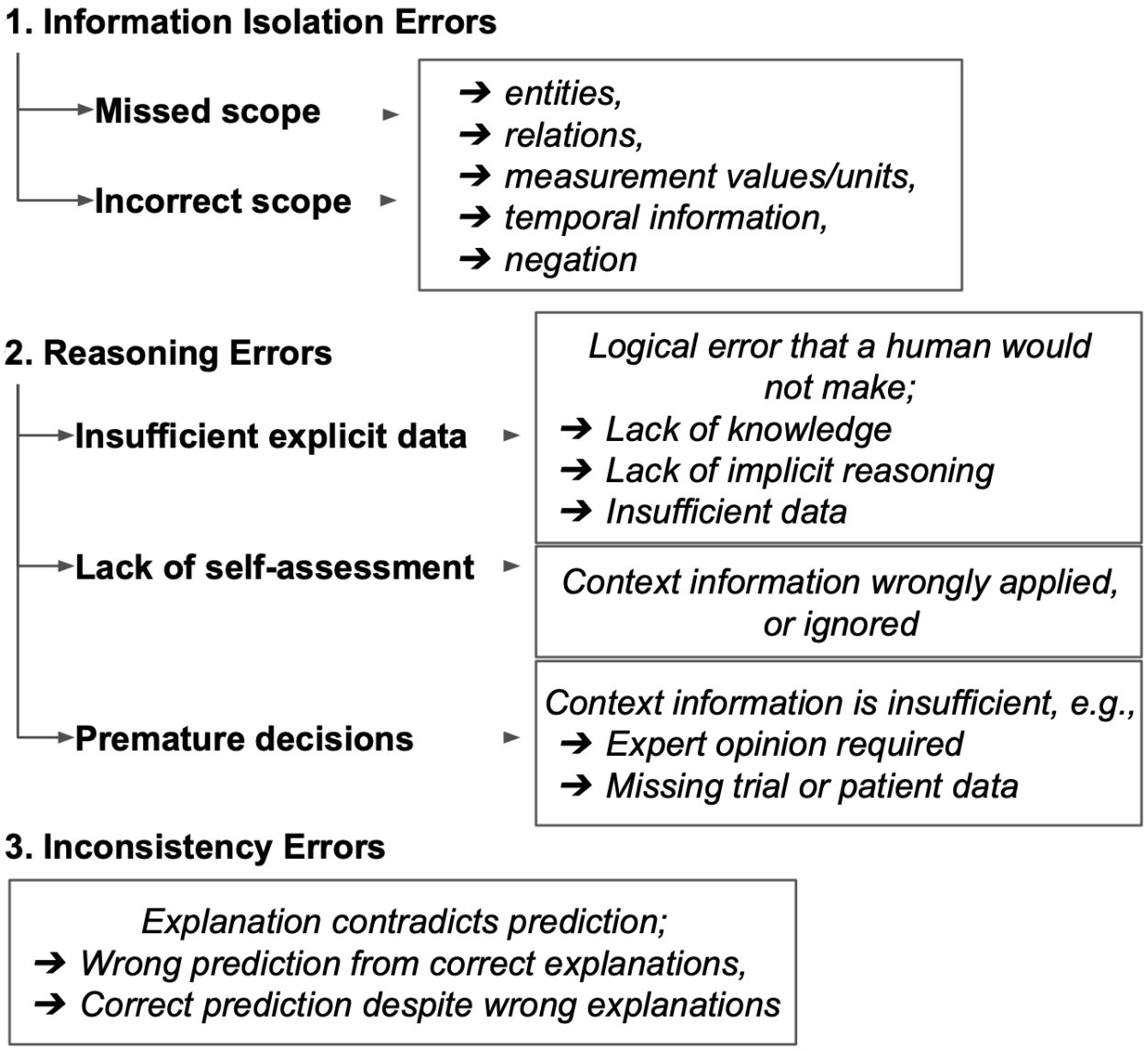


Figure 2: Taxonomy of errors in LLM-generations.

Cohort Discovery: A Survey on LLM-Assisted Clinical Trial Recruitment
Ghosh, et al. June 2025, arXiv:2506.15301v1

Outline

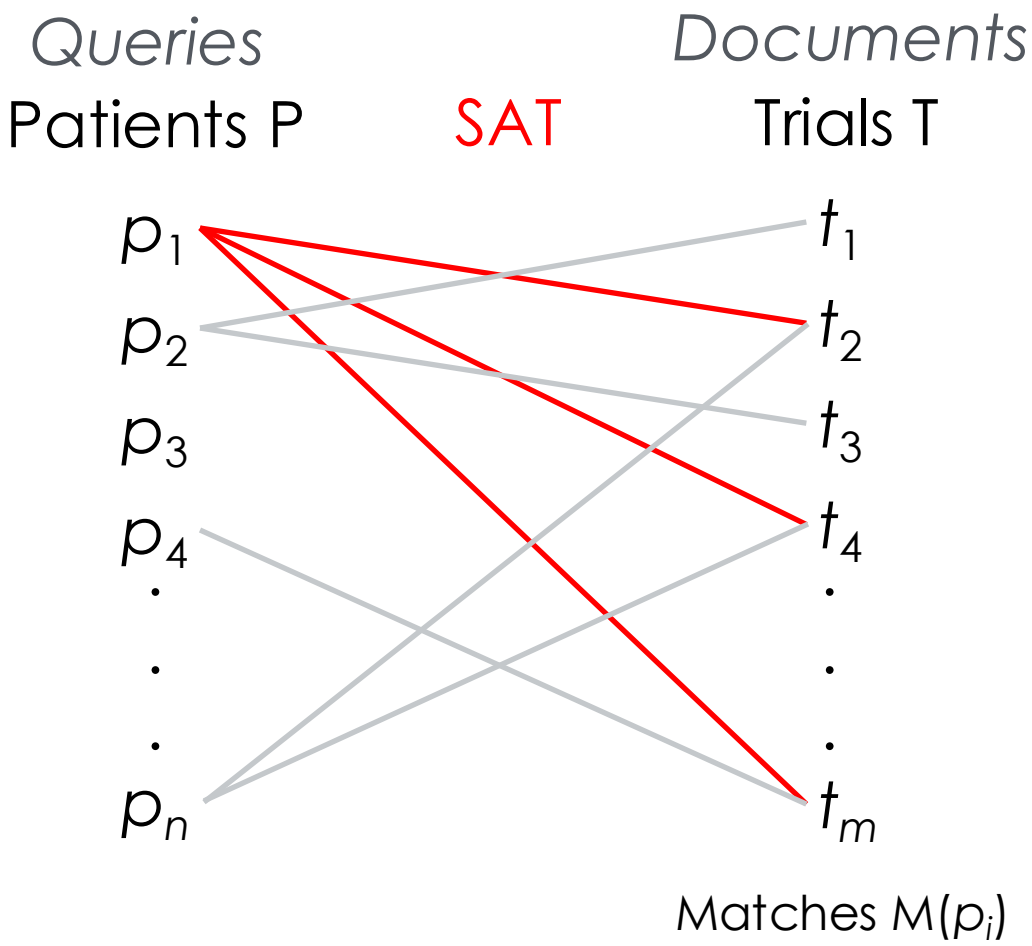
- Motivation
- Prior Work
- **SMT Approach**
- SMT-Based Matching Algorithm
- Preliminary Results

Goal: No False Negatives & High Precision

Aim to prevent:

- Patients miss life-saving opportunities
- Trials miss recruit targets
  - Stall scientific progress
  - Incur monetary loss
  - Waste researchers' time

Problem Statement



For every patient,
find all trials the patient satisfies (SAT)
→ pairwise operation $O(N^2)$

Two-step approach:

1. Retrieve
   - With **no false negatives**
   - As few false positives as possible
   - Quickly
2. Match patient with retrieved trials
   - Accurately and quickly

Lecture Objective

Can We Leverage

Formal Notations and a Theorem Solver

To Improve Precision and Recall?

2 Formal Notations

- **Propositional Logic (PL)**
  - Variables: Boolean
  - Operators: and, or, not
  - Conjunctive Normal Form: A and (B or C) and (A or D)…
  - SAT solver: checks if a set of PL statements are satisfiable

- **SMT (Satisfiability Modulo Theories)** ← More expressive than PL, SMT solver more expensive than SAT solver
  - Variables: Boolean, number, …
  - Operators: and, or, not, arithmetic, numeric comparison
  - Conjunctive Normal Form:
    A and (B > 10 or C + B > D) and (A or E)…
  - SMT solver: checks if a set of SMT statements are satisfiable

Example

https://www.clinicaltrials.gov/study/NCT00092885

An Approved Drug to Study a New Indication for Seasonal Allergic Rhinitis in Patients With Asthma (0476-269)

ClinicalTrials.gov ID NCT00092885

Sponsor Organon and Co

Information provided by Organon and Co (Responsible Party)

Last Update Posted 2024-08-15

Eligibility

Description

Inclusion Criteria:

- Non-smoker
- A 2-year documented history of seasonal allergic rhinitis
- A 1-year documented history of chronic asthma
- Positive allergy testing

Exclusion Criteria:

- Medical history of a lung disorder (other than asthma) or a recent upper respiratory tract infection.

Ages Eligible for Study

15 Years to 85 Years (Child, Adult, Older Adult)

Sexes Eligible for Study

All

Accepts Healthy Volunteers

No

Inclusion Constraints in SMT

Free-text Requirements

Non-smoker
A 2-year documented history of seasonal allergic rhinitis
A 1-year documented history of chronic asthma
Positive allergy testing

Requirements as SMT Constraints

```
(declare-const patient_has_finding_of_tobacco_smoking_behavior_now Bool)
(assert (not patient_has_finding_of_tobacco_smoking_behavior_now))

(declare-const patient_has_finding_of_seasonal_allergic_rhinitis_inthehistory Bool)
(assert (and patient_has_finding_of_seasonal_allergic_rhinitis_inthehistory
      (>= duration_of_documented_history_of_seasonal_allergic_rhinitis_in_years 2.0)))

(declare-const duration_of_chronic_asthma_in_years Real)
(declare-const patient_has_diagnosis_of_asthma_inthehistory Bool)
(declare-const patient_has_diagnosis_of_asthma_inthehistory@@chronic Bool)
(declare-const patient_has_diagnosis_of_asthma_inthehistory@@duration_at_least_1_year Bool)
(assert (=> patient_has_diagnosis_of_asthma_inthehistory@@chronic
        patient_has_diagnosis_of_asthma_inthehistory))
(assert (=> patient_has_diagnosis_of_asthma_inthehistory@@duration_at_least_1_year
                patient_has_diagnosis_of_asthma_inthehistory))
(assert (= patient_has_diagnosis_of_asthma_inthehistory@@duration_at_least_1_year
       (>= duration_of_chronic_asthma_in_years 1.0)))
(assert (and patient_has_diagnosis_of_asthma_inthehistory@@chronic
      patient_has_diagnosis_of_asthma_inthehistory@@duration_at_least_1_year))
```

Patient Record as SMTs (Example 1)

A 58-year-old African-American woman presents to the ER with episodic pressing/burning anterior chest pain that began two days earlier for the first time in her life.
The pain started while she was walking, radiates to the back, and is accompanied by nausea, diaphoresis and mild dyspnea, but is not increased on inspiration.
The latest episode of pain ended half an hour prior to her arrival.

(assert patient_has_undergone_emergency_room_admission_inthehistory)

(assert patient_has_symptoms_of_anterior_chest_wall_pain_now)

(assert patient_has_finding_of_anterior_chest_wall_pain_inthepast2days)

(assert patient_has_symptoms_of_dyspnea_inthepast2days)

Patient Record as SMTs (Example 2)

She is known to have hypertension and obesity.
She denies smoking, diabetes, hypercholesterolemia, or a family history of heart disease.
She currently takes no medications.
Physical examination is normal.
The EKG shows nonspecific changes.

(assert patient_has_diagnosis_of_hypertensive_disorder_now)

(assert (not patient_has_diagnosis_of_hypercholesterolemia_now))

(assert patient_has_undergone_electrocardiographic_procedure)

(assert (not patient_has_finding_of_diabetes_procedure_now))

Trial Matching as a Satisfiability Problem

- Given

  (A) Trial Requirements in SMT

  (B) Patient Fact Value Assertions in SMT

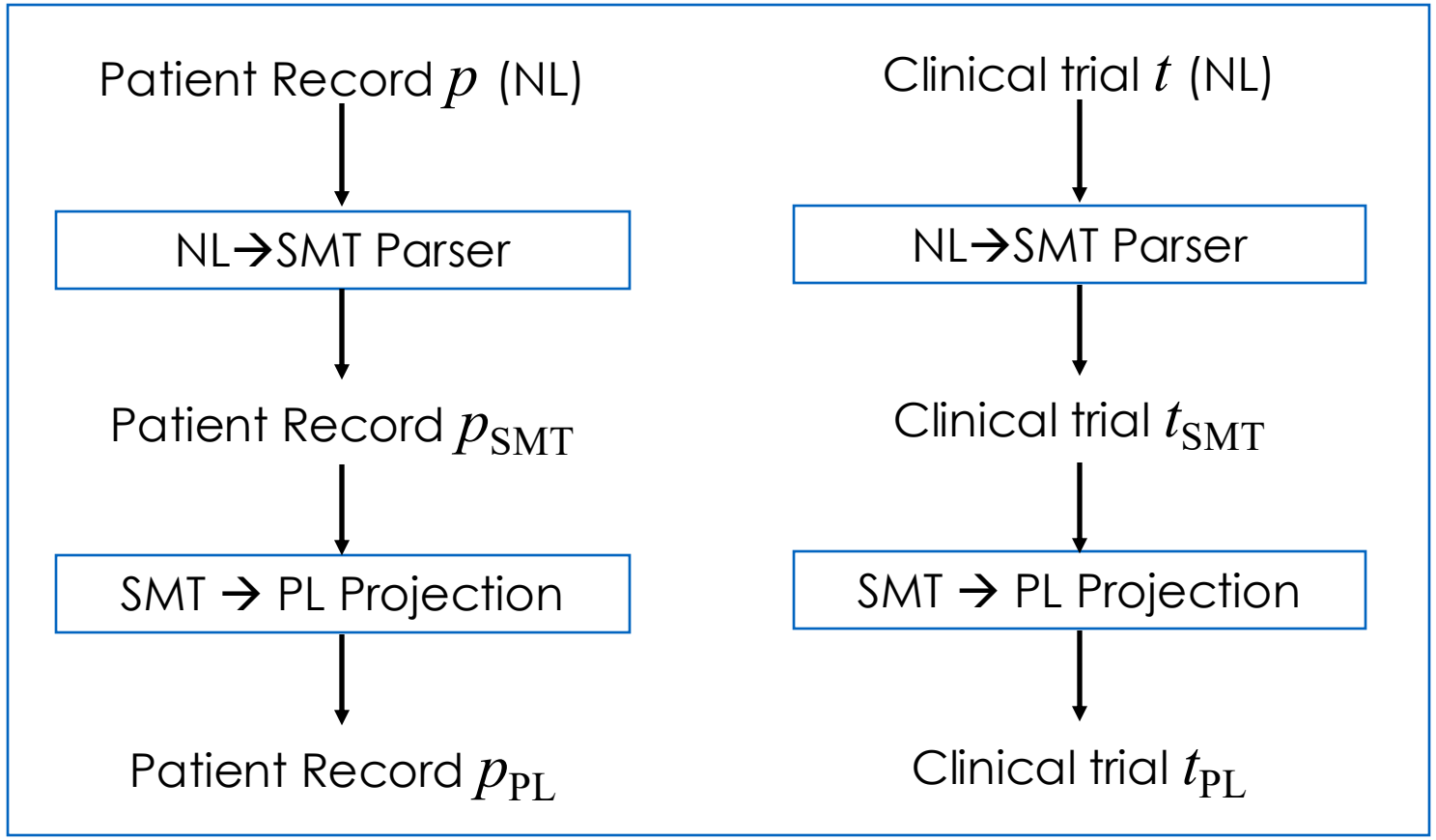
- There is a match if and only if (A) + (B) is satisfiable
- Use SMT Solver (Z3)

Z3 theorem prover, Microsoft, https://github.com/Z3Prover/z3

How Do We Retrieve the Potential Trials?

- Dense retrieval has inadequate recall and precision
- Use Propositional Logic (PL)
- Represent PL constraints and assertions in databases
  - Match all records quickly!

Clinical Trail Matching using SMT

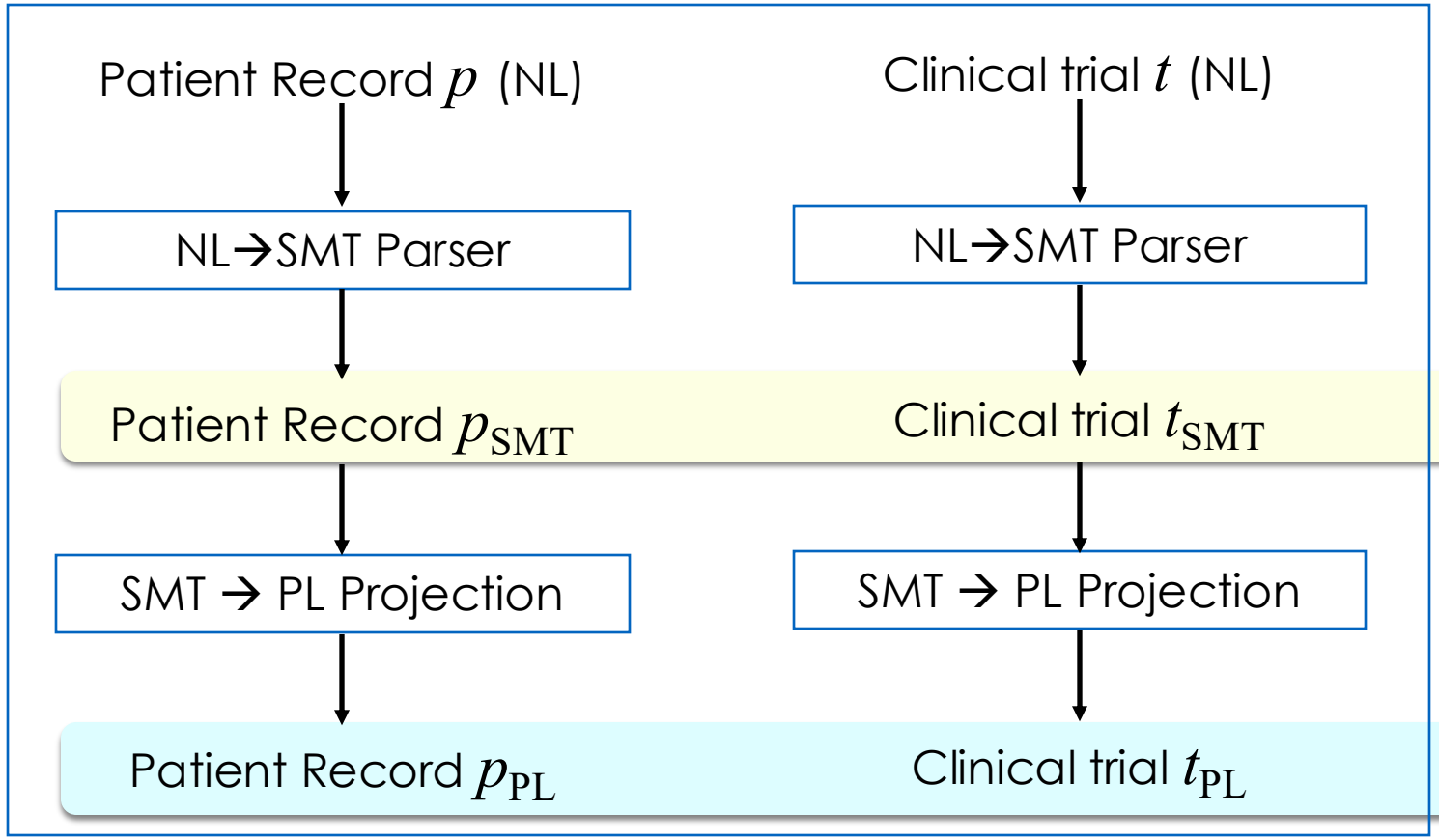


NL: Natural language
SMT: Satisfiability Modulo Theory
PL: Propositional Logic

A ⊨ B means
if A is satisfiable, B is satisfiable

$$p_{\text{SMT}} \models p_{\text{PL}}$$
$$t_{\text{SMT}} \models t_{\text{PL}}$$
$$p_{\text{SMT}} \wedge t_{\text{SMT}} \models p_{\text{PL}} \wedge t_{\text{PL}}$$

Clinical Trail Matching using SMT



NL: Natural language
SMT: Satisfiability Modulo Theory
PL: Propositional Logic

SMT Solver
a pair at a time

DB ops to match
all trials & patients

$$p_{\text{SMT}} \models p_{\text{PL}}$$
$$t_{\text{SMT}} \models t_{\text{PL}}$$
$$p_{\text{SMT}} \wedge t_{\text{SMT}} \models p_{\text{PL}} \wedge t_{\text{PL}}$$

SAT and SMT Solvers Are Provably Correct

If Semantics of the Patient Records and Trials is Captured Perfectly in SMT

➔ 100% Precision and Recall

But errors may be introduced in the NL→SMT parse

Advantages of SMT-PL-Approach

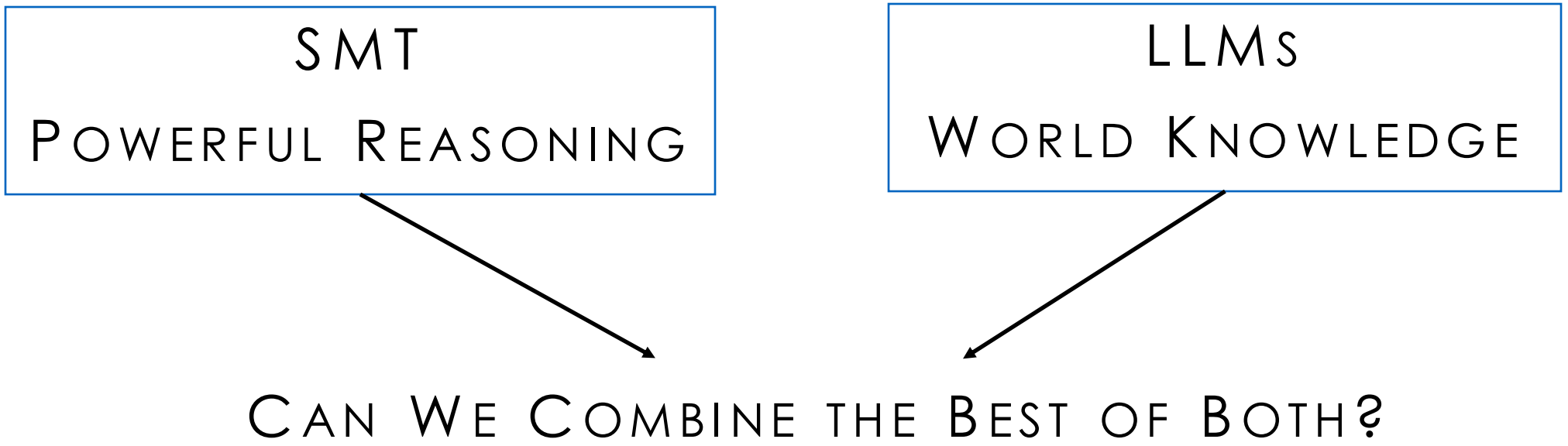
- LLM is used to **parse** text to SMT **once** per document
  - **Accuracy**:
    Parsing is easier than reasoning
    - Defer reasoning to SMT solvers
    - Easier to improve
    - Contexts are smaller
  - **Efficiency**: O(n), not O(n^2)
  - **Interpretability**
    - Can be checked & corrected

- **Retrieval** of satisfiable trials from **millions** of trials/patients in PL is **efficient & accurate**
  - Expressiveness is close to SMT for trials small % that sat t_{PL} and not sat t_{SMT}
    → **few false positives**

- **Matching** of satisfiable trials in **SMT** on **hundreds** of pairwise trials/patients is **efficient and accurate**

Disadvantages of SMT Approach

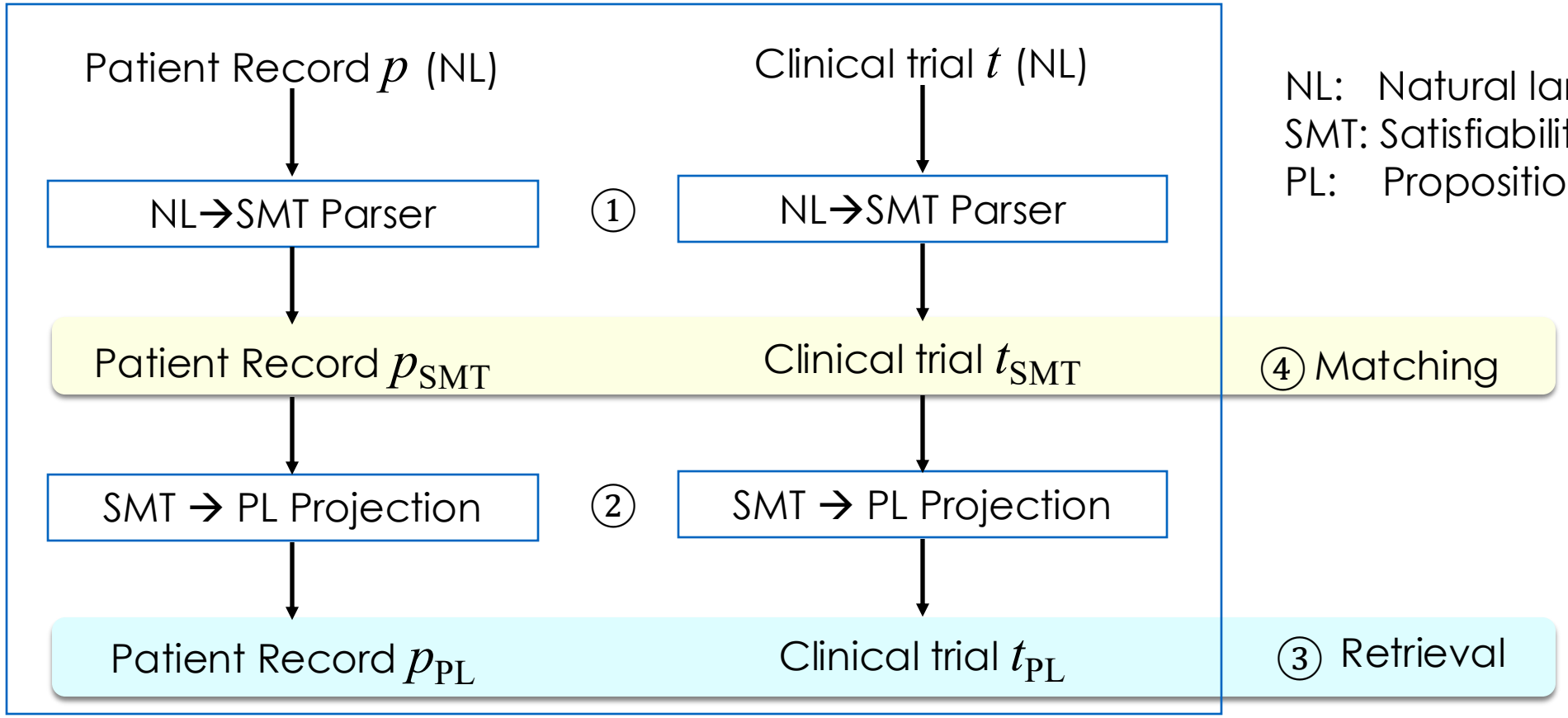
- SMT solvers operate on variables – they have no meaning
- All information useful for SMT solvers must be included formally
  - **Taxonomy**: same concepts → same variable name — **Canonicalization**
  - **Implication relations** between concepts
    e.g. A ⇒ B means A satisfies B — **Include implication relations**
  - **Judgment**: Distinguish between hard and soft constraints
    e.g. "Patient must have an Xray" can always be satisfied — **Requirement classification**
  - **Interpretation**: Require common knowledge
    e.g. "A heavy smoker" ⇒ daily-cigarettes >= 10 — **Not handled yet**


SMT
Powerful Reasoning
LLMs
World Knowledge
Can We Combine the Best of Both?

Outline

- Motivation
- TrialGPT Dataset and Prior Work
- SMT Approach
- **SMT-Based Matching Algorithm**
- Preliminary Results

Clinical Trail Matching using SMT



NL: Natural language
SMT: Satisfiability Modulo Theory
PL: Propositional Logic

$$p_{\text{SMT}} \models p_{\text{PL}}$$
$$t_{\text{SMT}} \models t_{\text{PL}}$$
$$p_{\text{SMT}} \wedge t_{\text{SMT}} \models p_{\text{PL}} \wedge t_{\text{PL}}$$

Algorithm

1. NL → SMT

Challenge 1: Canonicalization

Theorem provers operate on variables – they have no meaning

- Same concepts must be represented by the same variables

Example

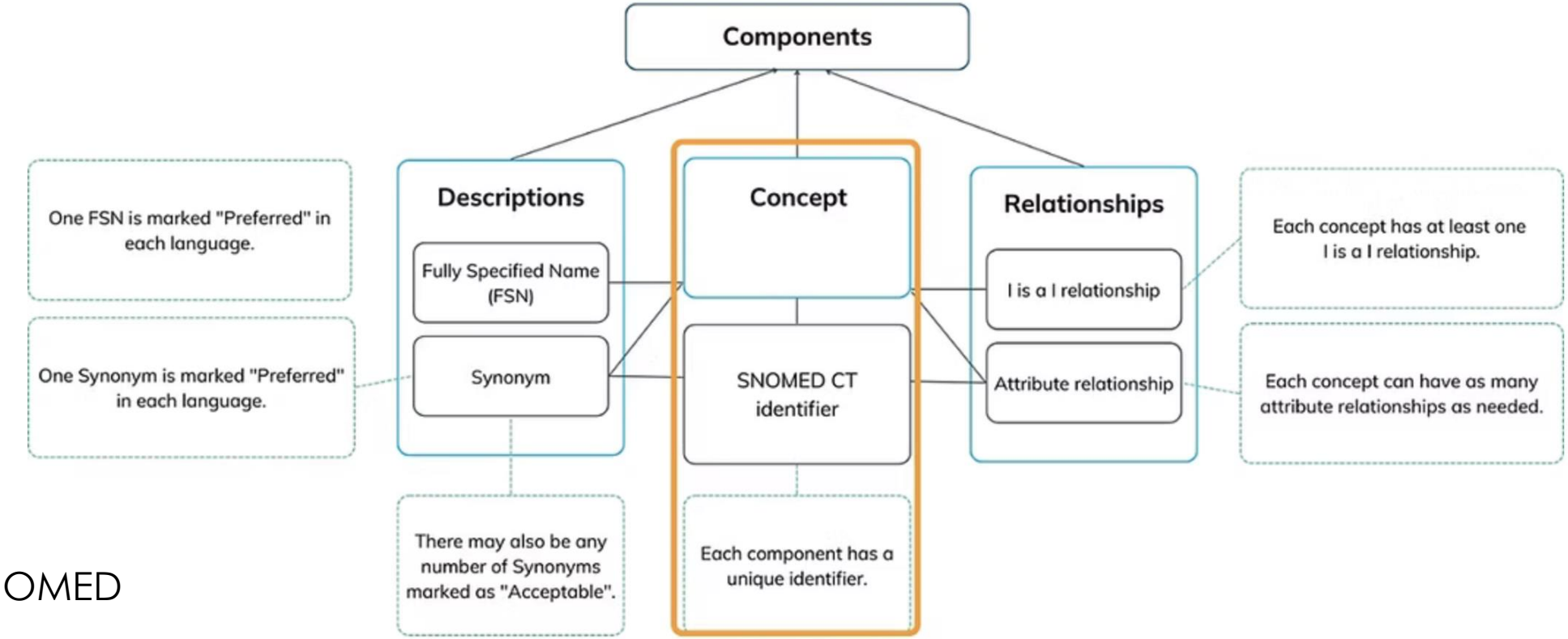
	Patient p	**Trial** t	$p \wedge t$
English	Has dizzy spells	Has syncope	SAT
Semantic parse	Has-dizzy-spell	Has-syncope	UNSAT
Canonicalized	Has-syncope	Has-syncope	SAT

Canonicalization is complex! So many clinical terms!

Standardization in medicine are important → SNOMED

SNOMED

Systematized Nomenclature of Medicine – Clinical Terms

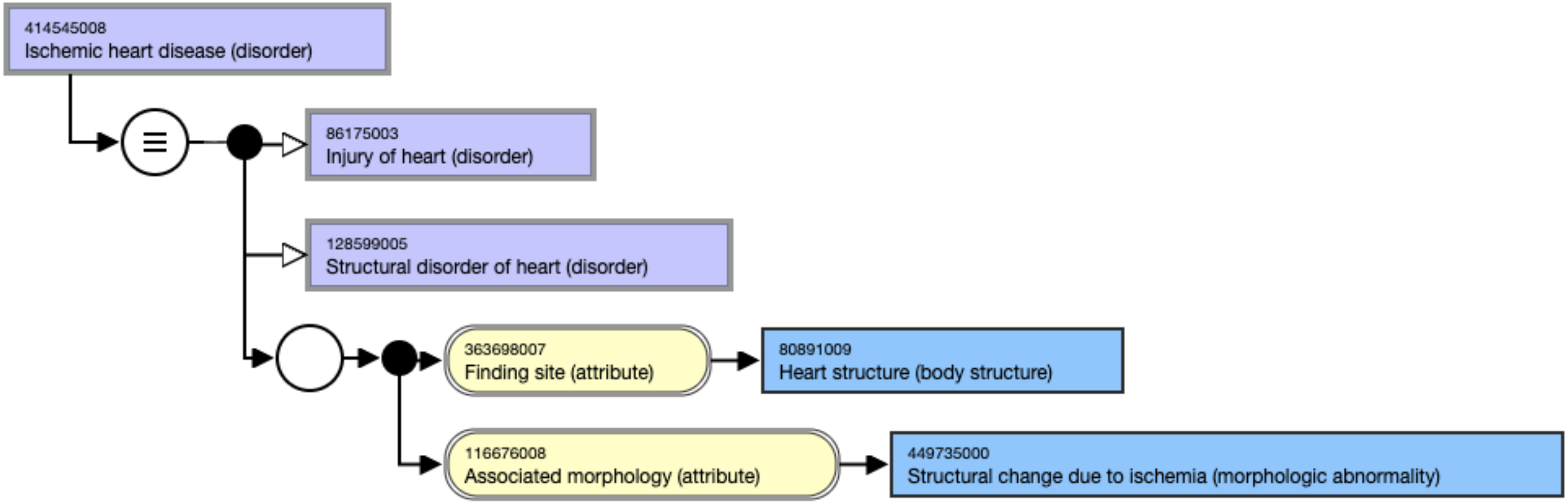


US Edition of SNOMED

- 300,000 unique concepts
- Over 1,000,000 descriptions
- 903,000 links or semantic relationships between concepts.

SNOMED Example

- Top-level classes: Clinical Findings, Procedures, Observable Entities, Substances…

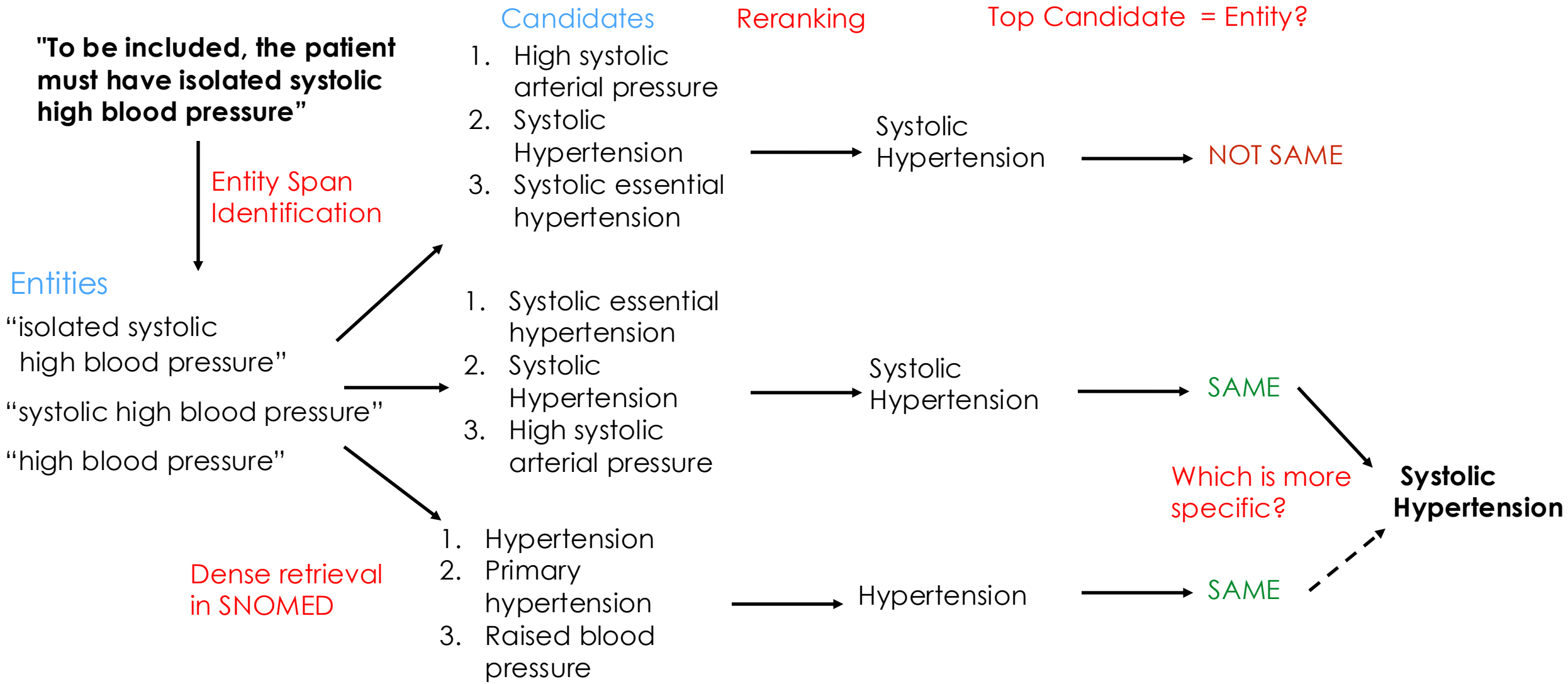


Canonicalization Algorithm

- Problem: NER (Named entity recognizer)
  - Turn the informal English to the closest term in SNOMED

- Challenges and solutions:
  - Many similar terms in SNOMED → Dense retrieval
  - Find the most precise entity in SNOMED →
    Retrieve, rerank, ensure equivalence, pick the most precise

How to Canonicalize?



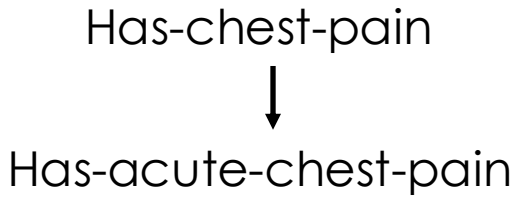
Challenge 2: Relationships Between Terms

Example

	Patient p	**Trial** t	$p \wedge t$
Canonicalized	Has-acute-chest-pain	Has-chest-pain	UNSAT
With implications	Has-acute-chest-pain (Has-chest-pain)	Has-chest-pain	SAT

- **SNOMED includes implication relations between terms:**

  ∃ an edge (x,y) if y ⟹ x



Has-acute-chest-pain ⟹ Has-chest-pain

¬ Has-chest-pain ⟹ ¬ Has-acute-chest-pain

- Codifying the patient
  - If patient has x, includes ancestors of x
  - If patient has ¬ x, includes the ¬ of all descendants of x

Challenge 3: How to Handle Missing Information?

- Properties that can be satisfied (can be ignored)

Example

	Patient p	**Trial** t	$p \wedge t$
Semantic parse		Has-a-chest-Xray	SAT

- Use LLM for classification

- Required properties, Expected to be reported: Default to "False"

Example

	Patient p	**Trial** t	$p \wedge t$
Semantic parse		Has-chest-pain	UNSAT

Algorithm

1. NL → SMT
2. **SMT → PL**

SMT is More Expressive than PL

Example:

At least three of the following disorders: A, B, C, D, E, F, G

SMT

```
(assert (>= (+
            (ite A 1 0)
            (ite B 1 0)
            (ite C 1 0)
            (ite D 1 0)
            (ite E 1 0)
            (ite F 1 0)
            (ite G 1 0))
        3))
```

Ite: If This Then

PL

(A∧B∧C) ∨ (A∧B∧D) ∨ (A∧B∧E) ∨ (A∧B∧F)
∨ (A∧B∧G) ∨(A∧C∧D) ∨ (A∧C∧E) ∨ (A∧C∧F)
∨ (A∧C∧G) ∨(A∧D∧E) ∨ (A∧D∧F) ∨ (A∧D∧G)
∨ (A∧E∧F) ∨(A∧E∧G) ∨ (A∧F∧G) ∨ (B∧C∧D)
∨ (B∧C∧E) ∨(B∧C∧F) ∨ (B∧C∧G) ∨ (B∧D∧E)
∨ (B∧D∧F) ∨(B∧D∧G) ∨ (B∧E∧F) ∨ (B∧E∧G)
∨ (B∧F∧G) ∨(C∧D∧E) ∨ (C∧D∧F) ∨ (C∧D∧G)
∨ (C∧E∧F) ∨(C∧E∧G) ∨ (C∧F∧G) ∨ (D∧E∧F)
∨ (D∧E∧G) ∨(D∧F∧G) ∨ (E∧F∧G)

SMT is direct and succinct

Example:
At least two procedures within 30 days before index day

SMT

```
; Uninterpreted sort of procedure events for a single patient
(declare-sort Proc 0)

; Attributes
(declare-fun kind (Proc) String)
(declare-fun day  (Proc) Int)    ; measured in days
(declare-const index_day Int)

; The target procedure type e.g., percutaneous coronary intervention
(define-fun is_target ((p Proc)) Bool (= (kind p) "PCI"))

; "At least two distinct target procedures in [index-30, index]"
(assert (exists ((p1 Proc) (p2 Proc))
  (and (distinct p1 p2)
       (is_target p1) (is_target p2)
       (<= 0 (- index_day (day p1))) (<= (- index_day (day p1)) 30)
       (<= 0 (- index_day (day p2))) (<= (- index_day (day p2)) 30)))))

(check-sat)
```

PL

Cannot be expressed!

SMT → PL

- Goal of translation from SMT → PL:
  - t_{PL} represents the tightest constraints of t_{SMT} expressible in PL
  - Remove constraints not satisfiable in PL

- Many constraints include non-canonical variables
  - Can we project away the non-canonical variables and retain the rest of the constraints?
  - Use Z3 Quantifier Elimination to handle the projection accurately

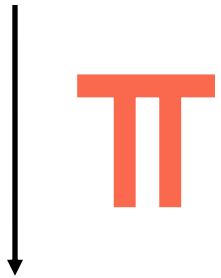
Z3 theorem prover, Microsoft, https://github.com/Z3Prover/z3

Example of Projection SMT → PL
-- removing non-canonical variables

(assert (or **has_diagnosis_of_squamous_cell_carcinoma_of_esophagus**

(and **has_diagnosis_of_adenocarcinoma_of_esophagus**

eligible_for_potentially_curative_radiotherapy)))



Canonical variables are marked in bold, other variables are projected away

(assert (or **has_diagnosis_of_squamous_cell_carcinoma_of_esophagus**

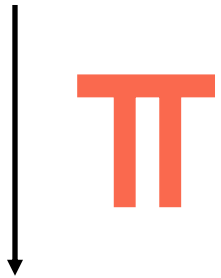
has_diagnosis_of_adenocarcinoma_of_esophagus))

Examples of Projection

(assert (or **has_diagnosis_of_squamous_cell_carcinoma_of_esophagus**

has_diagnosis_of_adenocarcinoma_of_esophagus

eligible_for_potentially_curative_radiotherapy))



Canonical variables are marked in bold, other variables are projected away

NOTHING!

Algorithm

1. NL → SMT
2. SMT → PL
3. **Retrieval (PL)**

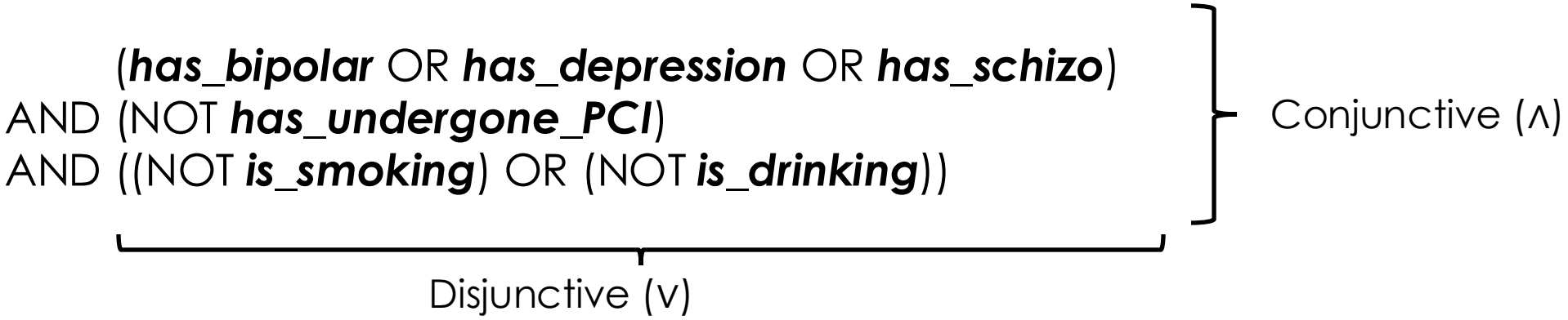
Retrieve Potential Trials from Corpus

1. Represent large sets of PL constraints in DB

2. Evaluate satisfaction as a DB queries

Constraints in a Trial

- All PL formulas can be expressed in Conjunctive Normal Form
- Example of a constraint



Variables: ***has_bipolar, has_depression, has_schizo, has_undergone_PCI, is_smoking, is_drinking***

Representing PL formulas in DB: Example

c1 { (***has_bipolar*** OR ***has_depression*** OR ***has_schizo***)
AND (NOT ***has_undergone_PCI***)
AND ((NOT ***is_smoking***) OR (NOT ***is_drinking***)) } a1 a2 a3

C	∧
c1	a1
c1	a2
c1	a3
c2	…
c2	…
c3	…

∧	∨
a1	o1
a1	o2
a1	o3
a2	o4
a3	o5
a3	o6

∨	Variables	Is Pos
o1	has_bipolar	1
o2	has_depression	1
o3	has_schizo	1
o4	has_undergone_PCI	0
o5	is_smoking	0
o6	is_drinking	0

Constraints ∧ (and) ∨ (or)

Matching Patient 1

has_bipolar AND ***has_depression***

(***has_bipolar*** OR ***has_depression*** OR ***has_schizo***)
AND (NOT ***has_undergone_PCI***)
AND ((NOT ***is_smoking***) OR (NOT ***is_drinking***))

pid	Variable	Val
123	has_bipolar	1
123	has_depression	1
123	has_schizo	0
123	has_undergone_PCI	0
123	is_smoking	0
123	Is_drinking	0

Patient Fact Table

∨	Variable	Is Pos
o1	has_bipolar	1
o2	has_depression	1
o3	has_schizo	1
o4	has_undergone_PCI	0
o5	is_smoking	0
o6	is_drinking	0

∨ (or)

∧	∨
a1	o1
a1	o2
a1	o3
a2	o4
a3	o5
a3	o6

∧ (and)

C	∧
c1	a1
c1	a2
c1	a3

Constraints

Matching Patient 1

has_bipolar AND ***has_depression***

(***has_bipolar*** OR ***has_depression*** OR ***has_schizo***)
AND (NOT ***has_undergone_PCI***)
AND ((NOT ***is_smoking***) OR (NOT ***is_drinking***))

pid	Variable	Val
123	has_bipolar	1
123	has_depression	1
123	has_schizo	0
123	has_undergone_PCI	0
123	is_smoking	0
123	Is_drinking	0

Patient Fact Table

Eval v

v	Variable	Is Pos
o1	has_bipolar	1
o2	has_depression	1
o3	has_schizo	1
o4	has_undergone_PCI	0
o5	is_smoking	0
o6	is_drinking	0

v (or)

∧	v
a1	o1
a1	o2
a1	o3
a2	o4
a3	o5
a3	o6

∧ (and)

C	∧
c1	a1
c1	a2
c1	a3

Constraints

Matching Patient 1

has_bipolar AND ***has_depression***

(***has_bipolar*** OR ***has_depression*** OR ***has_schizo***)
AND (NOT ***has_undergone_PCI***)
AND ((NOT ***is_smoking***) OR (NOT ***is_drinking***))

pid	v	Variable	Is_pos	v-Eval
123	o1	has_bipolar	1	1
123	o2	has_depression	1	1
123	o3	has_schizo	1	0
123	o4	has_undergone_PCI	0	1
123	o5	is_smoking	0	1
123	o6	is_drinking	0	1

v-Evaluated

Eval ∧

∧	v
a1	o1
a1	o2
a1	o3
a2	o4
a3	o5
a3	o6

∧ (and)

C	∧
c1	a1
c1	a2
c1	a3

Constraints

Matching Patient 1

has_bipolar AND ***has_depression***

(***has_bipolar*** OR ***has_depression*** OR ***has_schizo***)
AND (NOT ***has_undergone_PCI***)
AND ((NOT ***is_smoking***) OR (NOT ***is_drinking***))

∧	∧_eval
a1	1
a2	1
a3	1

C	∧
c1	a1
c1	a2
c1	a3

C	C_eval
C1	**1**

SAT!

∧ (and) Evaluated  Constraints

Eval Constraints

Matching Patient 2

has_schizo AND ***is_smoking*** AND ***is_drinking***

(***has_bipolar*** OR ***has_depression*** OR ***has_schizo***) AND (NOT ***has_undergone_PCI***) AND ((NOT ***is_smoking***) OR (NOT ***is_drinking***))

pid	Variable	val
123	has_bipolar	0
123	has_depression	0
123	has_schizo	1
123	has_undergone_PCI	0
123	is_smoking	1
123	Is_drinking	1

Patient Fact Table

v	Variable	Is Pos
o1	has_bipolar	1
o2	has_depression	1
o3	has_schizo	1
o4	has_undergone_PCI	0
o5	is_smoking	0
o6	is_drinking	0

v (or)

∧	v
a1	o1
a1	o2
a1	o3
a2	o4
a3	o5
a3	o6

∧ (and)

C	∧
c1	a1
c1	a2
c1	a3

Constraints

Matching Patient 2

has_schizo AND ***is_smoking*** AND ***is_drinking***

(***has_bipolar*** OR ***has_depression*** OR ***has_schizo***)
AND (NOT ***has_undergone_PCI***)
AND ((NOT ***is_smoking***) OR (NOT ***is_drinking***))

pid	Variable	val
123	has_bipolar	0
123	has_depression	0
123	has_schizo	1
123	has_undergone_PCI	0
123	is_smoking	1
123	Is_drinking	1

Patient Fact Table

Eval v

v	Variable	Is Pos
o1	has_bipolar	1
o2	has_depression	1
o3	has_schizo	1
o4	has_undergone_PCI	0
o5	is_smoking	0
o6	is_drinking	0

v (or)

∧	v
a1	o1
a1	o2
a1	o3
a2	o4
a3	o5
a3	o6

∧ (and)

C	∧
c1	a1
c1	a2
c1	a3

Constraints

Matching Patient 2

has_schizo AND ***is_smoking*** AND ***is_drinking***

(***has_bipolar*** OR ***has_depression*** OR ***has_schizo***) AND (NOT ***has_undergone_PCI***) AND ((NOT ***is_smoking***) OR (NOT ***is_drinking***))

pid	v	Variable	Is_pos	v-Eval
123	o1	has_bipolar	1	0
123	o2	has_depression	1	0
123	o3	has_schizo	1	1
123	o4	has_undergone_PCI	0	1
123	o5	is_smoking	0	0
123	o6	is_drinking	0	0

∧	v
a1	o1
a1	o2
a1	o3
a2	o4
a3	o5
a3	o6

C	∧
c1	a1
c1	a2
c1	a3

v-**Evaluated**

Eval ∧

∧ (and)

Constraints

Matching Patient (unsat)

has_schizo AND ***is_smoking***
AND ***is_drinking***

(***has_bipolar*** OR ***has_depression*** OR ***has_schizo***)
AND (NOT ***has_undergone_PCI***)
AND ((NOT ***is_smoking***) OR (NOT ***is_drinking***))

∧	∧_eval
a1	1
a2	1
a3	0

C	∧
c1	a1
c1	a2
c1	a3

C	C_eval
C1	**0**

UNSAT!

∧ (and) Evaluated  Constraints

Eval Constraints

Speed and Accuracy of Retrieval

- The above example shows that
  - A few DB calls can match 1 patient with 1 constraint
  - The procedure **handles all constraints in all trials**
  - Adding one more level of eval handles **eligibility of all trials**

- Very fast and efficient

Relevance

- Eligibility is inadequate
  - Patients seeking clinical trials have a medical concern
- Solution: Add a **relevance filter** to retrieval

Algorithm

1. NL → SMT
2. SMT → PL
3. Retrieval (PL)
4. **Matching (SMT)**

Trial Matching as a Satisfiability Problem

- Given

  (A) Trial Requirements in SMT

  (B) Patient Fact Value Assertions in SMT

- There is a match if and only if (A) + (B) is satisfiable
- Use SMT Solver (Z3)

Z3 theorem prover, Microsoft, https://github.com/Z3Prover/z3

Matching with SMT

Trial Req (SMT)

(assert (or has_cough has_fever))

(assert (>= age 18))

(assert (not is_pregnant))

Patient Fact (SMT)

(assert (= age 67)

(assert has_cough)

(assert (not is_pregnant))

SAT!

SMT Solver

Matching with SMT (Using Z3)

Trial Req (SMT)

(assert (or has_cough has_fever))

(assert (>= age 18))

(assert (not is_pregnant))

Patient Fact (SMT)

(assert (= age 67)

(assert (not has_cough)

(assert (not has_fever)

(assert (not is_pregnant))

SMT Solver

UNSAT

Outline

- Motivation
- TrialGPT Dataset and Prior Work
- SMT Approach
- SMT-Based Matching Algorithm
- **Preliminary Results**

Experiment: Based on TrialGPT

- Definition
  - Eligible trials:
    Patient satisfies all constraints, with diseases targeted by trials
  - Relevant trials: Trials address the patient's major disease
  - Potential trials: Trials for similar diseases
- Experiment setup
  - Number of trials: 3621
  - Represented in SMT and PL

Very Preliminary Results on Retrieval

- On 8 patients:
  - **Relevant and not excluded explicitly: 105 (average)**
  - **Our Recall: 100%**
    (Checked the disagreement with gold by hand)
- For 4 out of the 8 patients (Based on sampling 10%)
  - The gold is full of errors! Checked by hand
  - **Our Precision: 89%**
  - Use TrialGPT to retrieve the same number of trials
    - **Their Precision: 56%**

Speed: 2.95s per patient against 3621 trials (M2 MacBook, SQLite)

Conclusions

- **SMT: A new approach to large-scale hard reasoning tasks**
  - Important application: clinical trials matching
- **PL DB:** Retrieval with **100% recall** and **high precision**
- **LLMs encodes the logic of the constraints ONCE and FOR ALL**
  - Interpretable!
  - **Requires canonicalization: SNOMED**

Future research

- Can we apply to other domains? Use of SMTs requires canonicalization
  - Medicine has SNOMED
  - **Can we use AI to canonicalize other fields?**
- Can we improve LLMs reasoning skills?
  - **By fine-tuning LLMs to learn the formal representation?**