

Stanford CS224V Course
Agentic AI

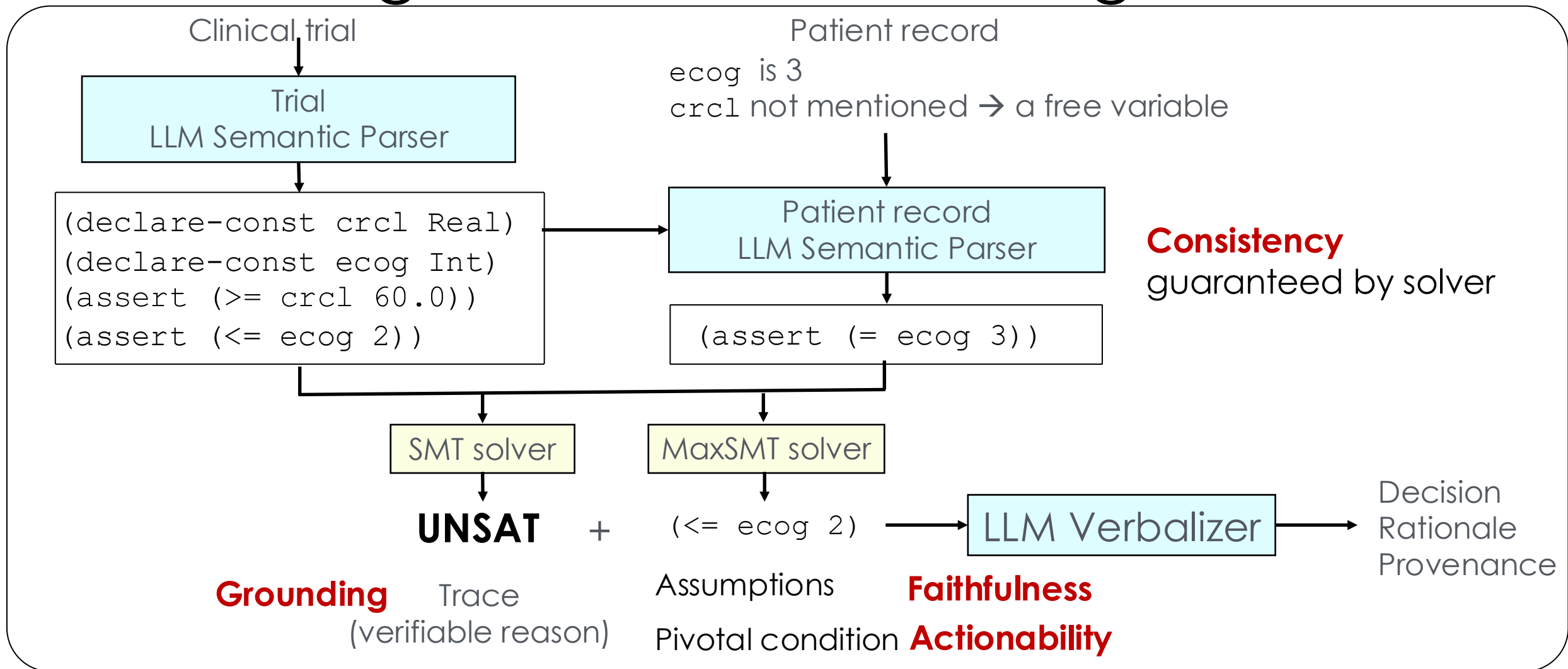
Lecture 4

Constraint-Satisfaction Based Retrieval

Case study: Clinical Trial Matching

Cyrus Zhou and Monica Lam

Formulating clinical trial matching



Cost of VERDICT

m: number of patients; n: number of trials	#times
• Trial parsed only once	$O(n)$
• Patient record is parsed differently for each trial	$O(mn)$
• SMT, MaxSMT calls	$O(mn)$

Too expensive for 500,000 clinical trials per patient!

Can we use a cheaper retrieval method to come up with a short list for matching?

Embedding-based retrieval has poor recall – may miss life-saving trials.

Goal: (1) Perfect recall

(2) High accuracy

New idea: **SatIR, constraint-satisfaction based retrieval**

Key Idea in SatIR

m: number of patients; n: number of trials

- Trial parsed only once
- Patient record is parsed differently for each trial
- SMT, MaxSMT calls

Too expensive for 500,000 clinical trials per patient!

#times

$O(n)$

$O(mn)$

$O(mn)$

→ $O(n)$

→ $O(m)$

→ 2.5K trials:
0.14ms/patient

SatIR

KEY: Patient records
are coded once
for all trials

SatIR: Use relational algebra (Databases): **less expressive**, but **much cheaper**

Project SMT formulas to relational algebra

An important math concept

Patient p

S: Eligible trials for p
(using SMT)

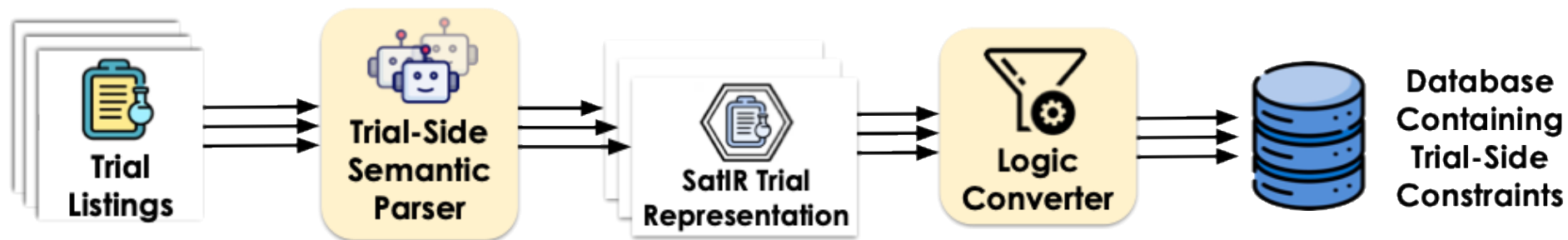
D: Eligible trials for p (using DBses)

PERFECT RECALL $D \subseteq S$

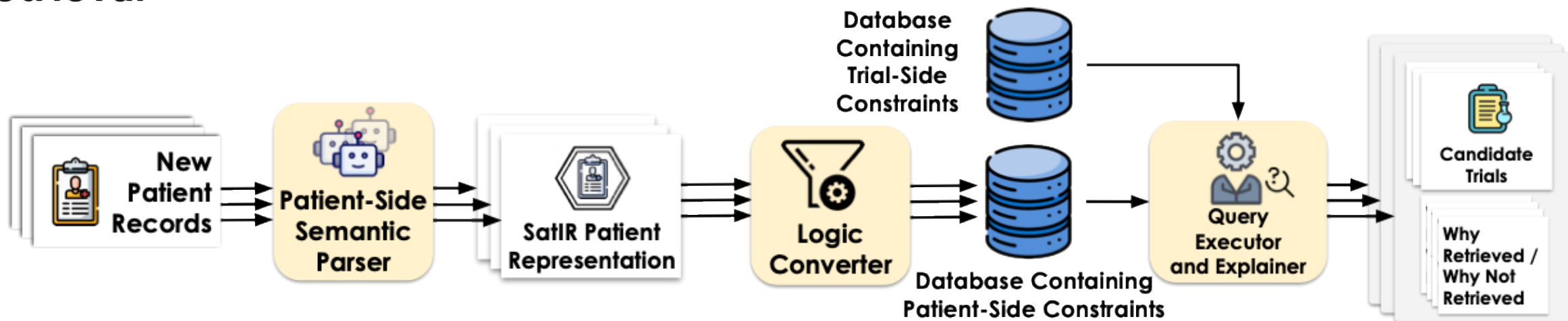
INACCURACY $D - S$

SatIR: Satisfiability-based information retrieval

Indexing



Retrieval



The constraints are stored in the databases

Why is this difficult?

1. SMTs symbols have no meaning!
 - Every nuance and inference needs to be spelled out
2. How to turn SMT satisfaction into database operations?

Outline of this Lecture

1. High-level concept in SatIR
2. Semantic parsing: NL to SMT for clinical trials
 - Basic concepts
 - Advanced concepts
3. Fast constraint-satisfaction based retrieval
4. Experimental Results

Outline of this Lecture

1. High-level concept in SatIR
2. Semantic parsing: NL to SMT for clinical trials
 - Basic concepts
 - Advanced concepts
3. Fast constraint-satisfaction based retrieval
4. Experimental Results

What is a Semantic Parser?

Translates NL into some formal representation



Formal Semantics: SMT Representation

An SMT formula: a set of atoms in a propositional formula (and, or, not)

For clinical trial matching:

```
atom = (variable, cmp, target)
variable = (relation, concept, qualifier)
( (hasCondition, Type2Diabetes, within5y), =, true)
( (hasMeasurement, HbA1c, current), <, 7.0)
```

Examples of Qualifiers

"steroid-induced osteonecrosis of the jaw" →

X steroid_induced, osteonecrosis, jaw

✓ base: Osteonecrosis

@@ qualifier: steroid-induced

@@ qualifier: finding site = jaw

patient: type2_diabetes diagnosed three years ago

fact: type2_diabetes @@ [t-3y, now]

trial: type2_diabetes @@within5y

constraint: interval $\cap$ [now-5y, now] $\neq \emptyset$

A parsed trial (Kawasaki disease, SatIR Fig. 3)

Kawasaki syndrome
with coronary artery dilation or aneurysm,
and ≥ 1 of:
 coronary Z-score ≥ 2.5 in any major segment;
 Persistent or recrudescent fever
 ≥ 38.0 °C
 for ≥ 36 h after initial therapy;
 or evidence of active inflammation.

```
(assert (and kawasaki_now  
  (or coronary_dilatation_now  
       coronary_aneurysm_now)))  
(assert (or  
  (and z_score_recorded_now@@any_major_segment  
       ( $\geq$  z_score 2.5))  
  (and (or fever_now@@persistent  
           fever_now@@recrudescent)  
       ( $\geq$  body_temp_celsius 38.0)  
       fever_now@@36h_after_initial_therapy)  
  acute_inflammation_now))
```

SMT programs are complex

Each trial is represented by
a non-trivial SMT program.

Typical number of lines:

~360

One-shot text-to-SMT is brittle

Problems

- Variables redeclared inconsistently
- Qualifiers dropped or misattached
- Polarity flips (esp. exclusions)
- Valid syntax, wrong meaning

Techniques

- Split cohorts
- Reverse exclusion polarity → one list
- Make entity spans contiguous
- Make and/or precedence explicit
- Split into self-contained units

Outline of this Lecture

1. High-level concept in SatIR
2. Semantic parsing: NL to SMT for clinical trials
 - Basic concepts
 - **Advanced concepts**
3. Fast constraint-satisfaction based retrieval
4. Experimental Results

Disadvantages of SMT Approach

- SMT solvers operate on variables – they have no meaning
- All information useful for SMT solvers must be included formally

1. **Taxonomy**: same concepts $\rightarrow$ same variable name **Canonicalization**

2. **Implication relations** between concepts

e.g. $A \Rightarrow B$ means A satisfies B

**Include
implication relations**

3. **Missing Data**: most conditions in trials are missing

Data Imputation

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-spells	Has-syncope	UNSAT
Canonicalized	Has-syncope	Has-syncope	SAT

Canonicalization is complex! So many clinical terms!

Standardization in medicine are important → SNOMED

SNOMED

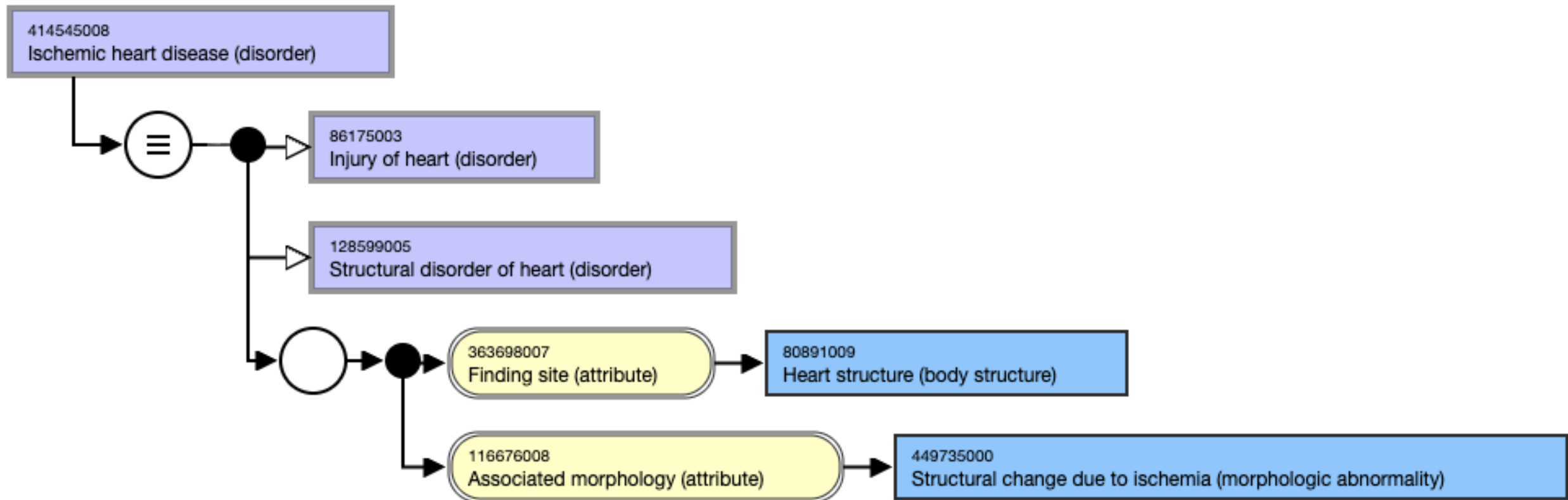
Systematized Nomenclature of Medicine – Clinical Terms

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...



Using SNOMED in the SMT representation

An SMT formula: a set of atoms in a propositional formula (and, or, not)

For clinical trial matching:

```
atom = (variable, cmp, target)
variable = (relation, concept, qualifier)
```

Relations: (we defined them empirically)

- e.g. hasFindingOf, hasCondition, hasMeasurement, hasUndergone

Canonical concepts

- Use canonical terms from SNOMED in clinical trials and patient records

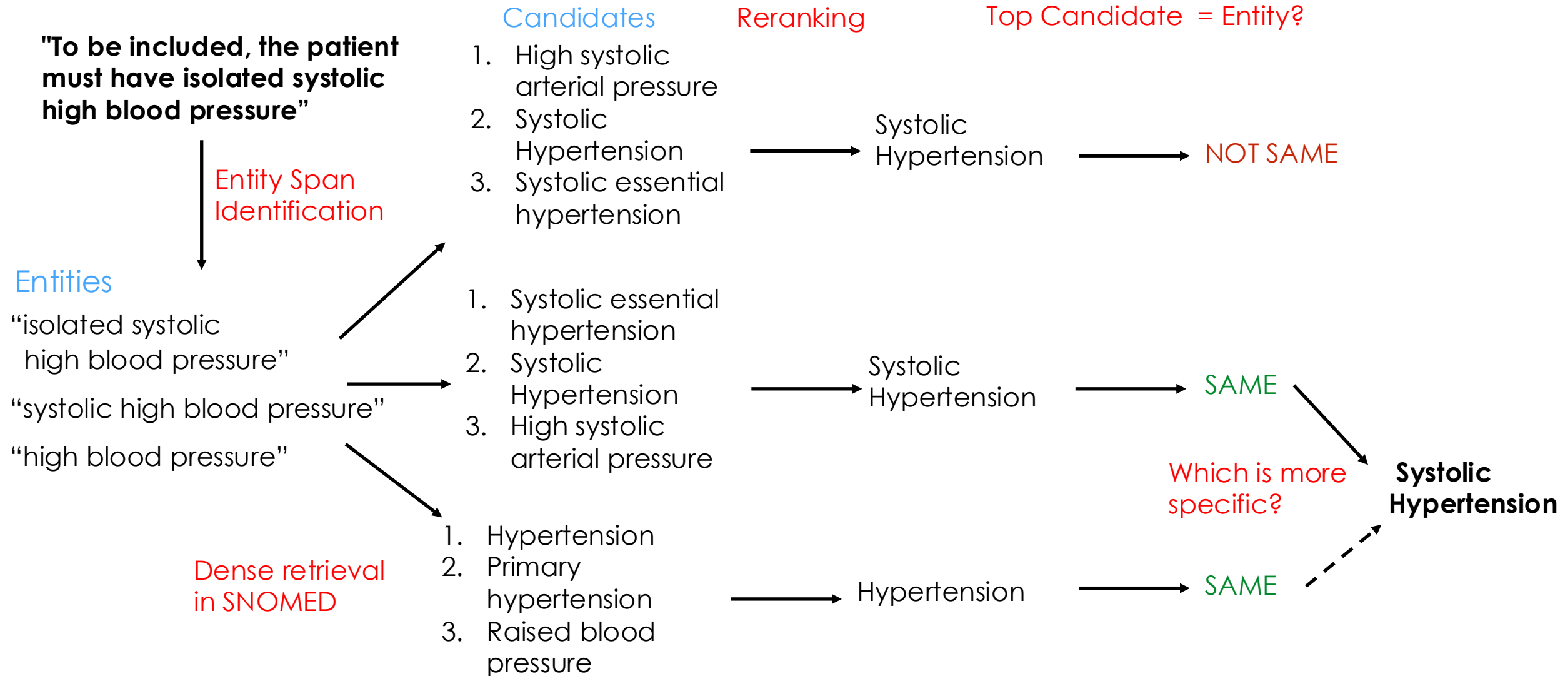
Non-canonical concepts

- Many concepts are trials/patient records are NOT in SNOMED
- Make up non-canonical variables with a description
- E.g. HasAdequateSocialSupportForFollow

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?



A Full Trial Example

The patient must have Kawasaki syndrome with either coronary artery dilation or aneurysm and meet one of the following conditions:

- coronary artery involvement, defined as Z-score ≥ 2.5 in any major coronary segment;
- persistent or recrudescent fever, defined as temperature $\geq 38.0^{\circ}\text{C}$ for ≥ 36 hours after initial therapy; or
- evidence of active inflammation,

To be included, the patient must have Kawasaki syndrome with (coronary artery dilation) **OR** (coronary artery aneurysm).

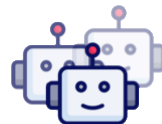
To be included, the patient must satisfy ≥ 1 of the following conditions:

- (coronary artery involvement, defined as Z-score ≥ 2.5 in any major coronary segment) **OR**
- (persistent fever **OR** recrudescent fever, defined as temperature $\geq 38.0^{\circ}\text{C}$ for ≥ 36 hours after initial therapy) **OR**
- (evidence of active inflammation)

SNOMED-Based Medical Ontology



ii. Canonicalization of Medical Concepts



Dilation:

- Dilatation of Blood Vessel
- ID: 248493003

Coronary Artery Aneurysm:

- Aneurysm of Coronary Vessels
- ID: 50570003

Coronary Artery:

- Coronary Artery
- ID: 41801008I

Fever:

- Fever
- ID: 386661006

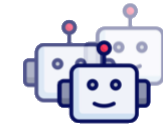
Body Temperature:

- Body Temperature
- ID: 386725007

Active inflammation:

- Acute Inflammation
- ID: 4532008

iii. Programming of Constraints



```
(assert (and
  acute_febrile_mucocutaneous_lymph_node_syndrome_now
  (or
    dilatation_of_blood_vessel_now@@finding_site_coronary_artery
    aneurysm_of_coronary_vessels_now)))
(assert (or (and
  coronary_artery_z_score_value_recorded_now_with_unit_unitless@@in_any_major_coronary_segment
  (>=
    coronary_artery_z_score_value_recorded_now_with_unit_unitless 2.5))
  (and (or fever_now@@persistent
    fever_now@@recrudescent)
  (>=
    body_temperature_value_recorded_now_withunit_celsius 38.0)
    fever_now@@foradurationof36hours
    fever_now@@temporalcontext_after_initial_therapy
    acute_inflammation_now)))
```

Text Color Legend

Preprocessing Modifications

Canonicalized Concepts

Canonical SMT Variables

Qualifiers to Canonical Variables

Non-Canonical Variables

Challenge 2: Implication Relations Between Concepts

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:**

$\exists$ an edge (x,y) if $y \Rightarrow x$

Has-chest-pain
↓
Has-acute-chest-pain

Has-acute-chest-pain $\Rightarrow$ Has-chest-pain
 $\neg$ Has-chest-pain $\Rightarrow$ $\neg$ Has-acute-chest-pain

- Codifying the patient
 - If patient has x , includes ancestors of x
 - If patient has $\neg x$, includes the $\neg$ of all descendants of x

Different kinds of subsumption

- Concept: AcuteAppendicitis $\sqsubseteq$ Appendicitis
- Relation: HasDiagnosisOf $\sqsubseteq$ HasFindingOf
- Qualifier: specific $\sqsubseteq$ general
- Causal: postop hemorrhage $\sqsubseteq$ underwent surgery

The patient records include all the implications

An important concept

-42%

True positives
if we don't use the ontology

(TREC 2022, GPT-5 judge)

Challenge 3. Missing Data

Use **salience** of property in question to impute data
Assume: Important facts get written down

Trial asks about	If missing, assume:	Why
Pregnancy	false	Pregnancy is salient
Ability to walk	true	Inability to walk is salient
Hospitalized now	unconstrained	Could go either way

Per-condition LLM question

"If this were true, would it plausibly be in the note?"

1,190

stored default values (TREC)

-7 to -9%

true positives without them

Using salience in handling specificity:

when may a vague note match a specific trial?

Hierarchy of diseases

Note: "appendicitis"

Acute appendicitis

✓ match allowed: "acute" is often left unwritten

Ruptured suppurative appendicitis

✗ no match:
an emergency would be documented

Widen to a more specific trial target only if the detail could plausibly be omitted.

Other constraints

Trial constraints need to be classified as hard, soft, operational (ignored)

Explicit matching objective

Objective	A trial matches if it...
treat-chief	treats the chief complaint
treat-any	treats any documented condition
relevant-to-any	is meaningfully related to any condition

The objective is a formula too: a product decision, made explicit.

Size of patient records

49 patients **2,578** source fact lines → **48,845** SMT fact lines
patient facts after ontology enrichment (TREC, 49 patients)

Implementation Complexities

Each requirement $\leftrightarrow$ its SMT lines: errors are attributable to a sentence.

Many steps to ensure the SMT program is correct

Reuse symbols



Attach qualifiers



New variables if needed



Translate fragment



Validate syntax · sorts and logic consistency · faithfulness



Commit if valid; else roll back with error feedback

Summary: Disadvantages of SMT Approach

- SMT solvers operate on variables – they have no meaning
- All information useful for SMT solvers must be included formally

1. **Taxonomy**: same concepts $\rightarrow$ same variable name **Canonicalization**

2. **Implication relations** between concepts

e.g. $A \Rightarrow B$ means A satisfies B

**Include
implication relations**

3. **Missing Data**: most conditions in trials are missing

Data Imputation

Outline of this Lecture

1. High-level concept in SatIR
2. Semantic parsing: NL to SMT for clinical trials
 - Basic concepts
 - Advanced concepts
3. Fast constraint-satisfaction based retrieval
4. Experimental Results

Projection of SMT to Relational Algebra

SatIR: Use relational algebra (Databases): **less expressive**, but **much cheaper**

Project SMT formulas to relational algebra

Patient p



D: Eligible trials for p (using DBses)

PERFECT RECALL $D \subseteq S$

INACCURACY $D - S$

SMT vs. Propositional Logic (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) V (A∧B∧D) V (A∧B∧E) V (A∧B∧F)
V (A∧B∧G) V (A∧C∧D) V (A∧C∧E) V (A∧C∧F)
V (A∧C∧G) V (A∧D∧E) V (A∧D∧F) V (A∧D∧G)
V (A∧E∧F) V (A∧E∧G) V (A∧F∧G) V (B∧C∧D)
V (B∧C∧E) V (B∧C∧F) V (B∧C∧G) V (B∧D∧E)
V (B∧D∧F) V (B∧D∧G) V (B∧E∧F) V (B∧E∧G)
V (B∧F∧G) V (C∧D∧E) V (C∧D∧F) V (C∧D∧G)
V (C∧E∧F) V (C∧E∧G) V (C∧F∧G) V (D∧E∧F)
V (D∧E∧G) V (D∧F∧G) V (E∧F∧G)
```

SMT is direct and succinct

Some requirements representable in SMT cannot be expressed in propositional logic
(e.g. counting incidents in a time window)

Projection designed to not lose recall

sacrificing precision if necessary

```
(and T2DM@@within5y
      (>= hba1c 7.0) (<= hba1c 10.0)
      HasAdequateSocialSupportForFollowUp)
```

```
(and T2DM@@within5y
      (>= hba1c 7.0) (<= hba1c 10.0))
; non-canonical atom projected away
```

A patient without social support matches
→ no loss in recall, loss in precision

Quiz: Why do we bother to include non-canonical atoms?

- In retrieval:

```
(OR T2DM@@within5y  
(>= hba1c 7.0) (<= hba1c 10.0)  
HasAdequateSocialSupportForFollowUp)
```

- In matching:

Projecting SMT Matching to Database Queries

1. Project away non-canonical parts

2. Represent as Canonical Normal Form
(AND of ORs of atoms)

3. Match = Table joins + aggregation

Projection: every step is conservative (in the paper)

Theorem:

SatIR has perfect recall (assuming the SMT formulation is correct)

Representing PL formulas in DB: Example

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

C	Λ	Λ	V	V	Variables	Is Pos
c1	a1	a1	o1	o1	has_bipolar	1
c1	a2	a1	o2	o2	has_depression	1
c1	a3	a1	o3	o3	has_schizo	1
c2	...	a2	o4	o4	has_undergone_PCI	0
c2	...	a3	o5	o5	is_smoking	0
c3	...	a3	o6	o6	is_drinking	0

Constraints

Λ (and)
clauses

v (or)
clauses

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

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)
clauses

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

∧ (and)
clauses

C	∧
c1	a1
c1	a2
c1	a3

Constraints


+ Eval literals
(SQL query)

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	Val	Is_pos	Eval
123	o1	has_bipolar	1	1	1
123	o2	has_depression	1	1	1
123	o3	has_schizo	0	1	0
123	o4	has_undergone_PCI	0	0	1
123	o5	is_smoking	0	0	1
123	o6	is_drinking	0	0	1

literal-Evaluated

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

∧ (and)
clauses

C	∧
c1	a1
c1	a2
c1	a3

Constraints



+ Eval v
(SQL query)

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*))

$\wedge$	eval
a1	1
a2	1
a3	1

v evaluated



Eval $\wedge$
(SQL query)

C	$\wedge$
c1	a1
c1	a2
c1	a3

Constraints



C	C_eval
C1	1

SAT!

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*))

$\wedge$	$\wedge_eval$
a1	1
a2	1
a3	0

C	$\wedge$
c1	a1
c1	a2
c1	a3



C	C_eval
C1	0

UNSAT!

$\wedge$ (and) Evaluated



+Eval $\wedge$
(SQL query)

Constraints

Theorem: projection never loses recall

The price: precision

Relational constraints are less expressive to evaluate;
a downstream matcher re-checks candidates.

146 ms

per patient · 3,621 trials · laptop

Outline of this Lecture

1. High-level concept in SatIR
2. Semantic parsing: NL to SMT for clinical trials
 - Basic concepts
 - Advanced concepts
3. Fast constraint-satisfaction based retrieval
4. Experimental Results

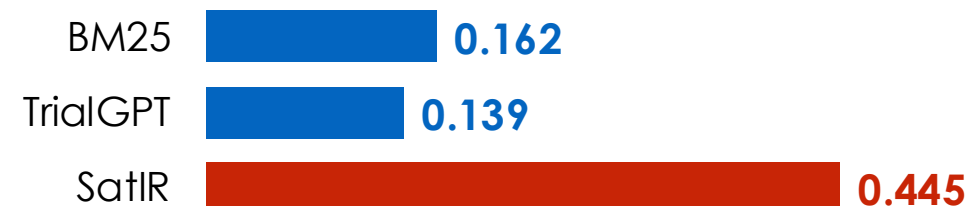
More of the right trials

Labels: GPT-5 judge; validated a sample by 2 clinicians, strong agreement

SIGIR 2016, treat-chief (59 patients, 3621 trials) Relevant & eligible per patient



TREC 2022 subset (50 patients, 2658 trials) recall, treat-chief, qrel = 2 (highly relevant)



Union of good trials across techniques



Where SatIR still fails

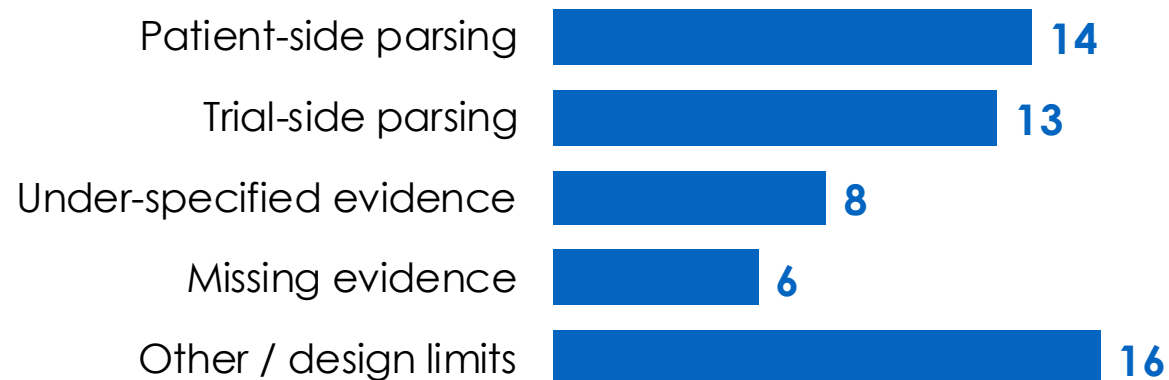
74

found only by TrialGPT
(200 candidates per patient)

Error Analysis



57 clear failures, by source



253

found only by SatIR

```
"age over 15" → (> age 15) ; patient is 15
should it be (>= age 15)? -- visible, arguable, fixable
```

SatIR: a new retriever based on **constraint satisfaction**

Constraints in documents are stored as first-class objects in databases

1. **Be conservative**: don't let unformalizable content exclude candidates
2. Compile **domain knowledge** into constraints (largest effect)
3. Store **missing-data assumptions** as data you can inspect and ablate
4. Build programs **incrementally, validated** at every step
5. Relax technique for speed if necessary
 - only with a **proof** of the properties you maintain