Large Semantic Stores in Soar

Nate Derbinsky, John E. Laird, Bryan Smith

University of Michigan

Semantic Knowledge

Some knowledge can be useful independent of the context in which it was initially learned

Lexical	Word meanings, synonyms, …
Mathematical	Arithmetic facts, function/relation definitions, ….
Geographical	Capitals, bodies of water, …
Historical	Wars, discoveries, reigns of power, …
Ontological	Biology, technology, art, …
Commonsensical	"Tables typically have four legs"
…	

Large Semantic Stores

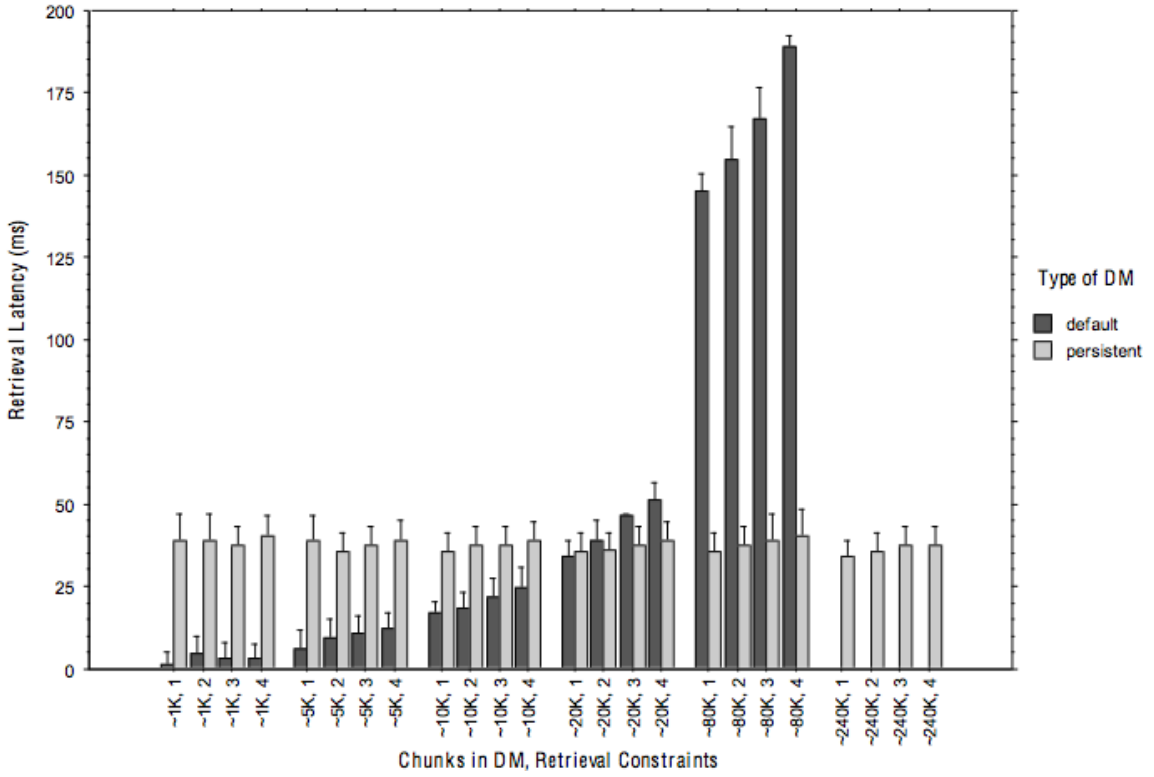
Agents contending with complex tasks will need access to large amounts of semantic knowledge

WordNet	Cyc	SUMO
Lexical	Common Sense	Ontological
> 212K senses >820K facts	>500K concepts >5M facts >1M rules	>4.5K classes >250K facts

Scaling to Large Semantic Stores

To remain reactive to dynamic environments, an agent's cognitive architecture must support bounded search over large stores of knowledge

Scaling: ACT-R DM



Douglass, S., Ball, J., & Rodgers, S. (2009). Large declarative memories in ACT-R. ICCM

Scaling: Soar (Data Chunking)

Approach

Productions condition upon all possible combinations of cue structures

Results in creation, matching, and updating an exponential number of rules

Example

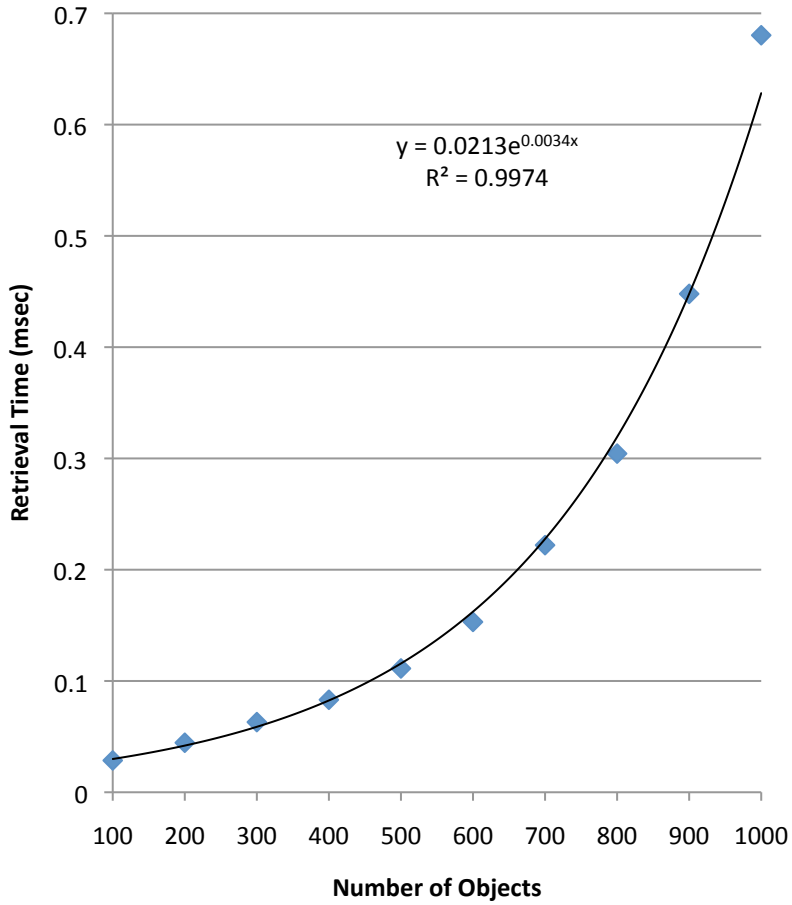
1. (a AND b AND c AND d)
2. (a AND b AND c)
3. (a AND b AND d)
4. (a AND c AND d)
5. (b AND c AND d)

…

Scaling: Soar (Working Memory)

On-demand procedural rule matching per-cue

Match time grows with the number of objects



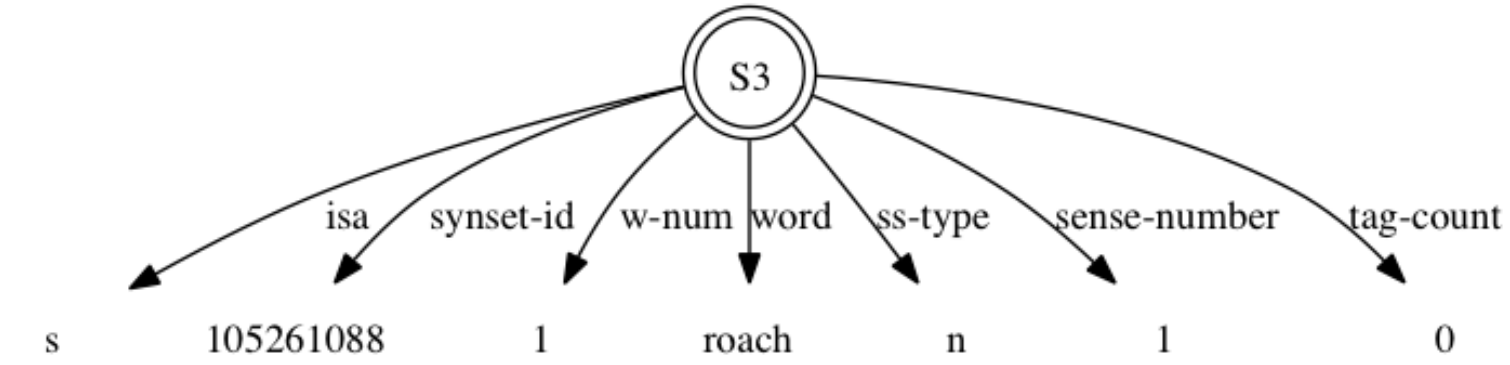
Large Semantic Stores in Soar

- Defining semantic knowledge/retrieval
- Retrieval mechanism description
- Evaluation
  - WordNet
  - Synthetic

Semantic Knowledge

Long-term identifier and augmentations
- Same representation as working memory

Similar to ACT-R declarative chunks
- No pre-defined chunk types
- Supports multi-valued attributes



Semantic Retrievals

Given

- Conjunctive set of attribute [/value pairs]
- Semantic store

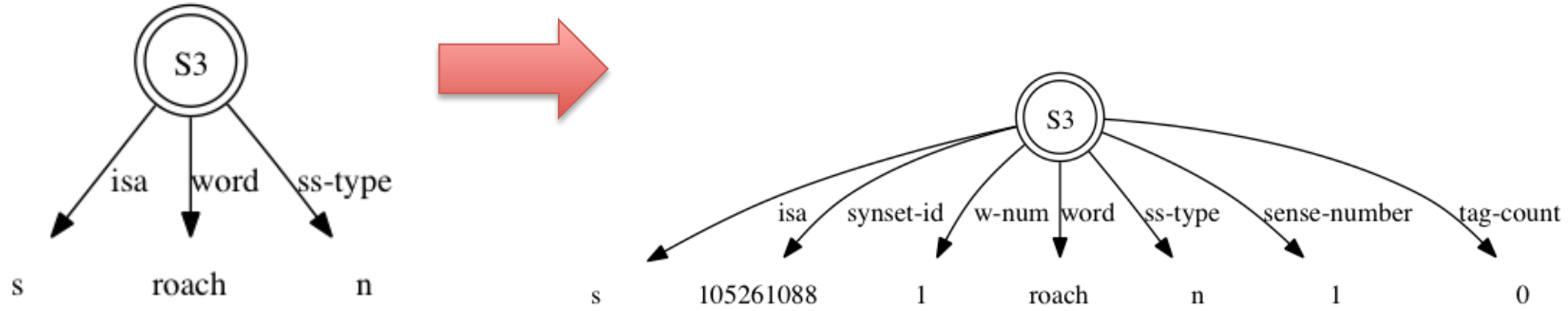
Find

- Most recently accessed identifier that contains cue structures

Does NOT Support

- Inference
- Partial matching
- Spreading activation

…



Retrieval Mechanism

Algorithm

1. Sort cue elements by rarity of occurrence
   a) *candidates* = identifiers satisfying rarest cue element
2. While *candidates* remain
   a) c = next *candidate*
   b) if c contains remaining cue elements, done

Notes

B+-trees support…

- log-time incremental updates to attribute/value frequency statistics during storage
- log-time access to incrementally updated linked list of candidates
- log-time query of candidate-cue element containment

Comparison to Other Approaches

Database Management System (DBMS)

- Constrained storage/retrieval interface yields opportunities for optimization

CYC

- Strict separation of problem/knowledge search delegates expensive inference to control knowledge

Evaluation: WordNet

WN-LEXICAL WordNet 3 data conversion

Edmond, B. (2006). WN-LEXICAL: An ACT-R Module Built from the WordNet Lexical Database. ICCM

Soar-SMem import

Instructions on wiki (~ 5 minutes)

~821.5K nodes, ~3.76M edges

~400MB on disk

WordNet Retrievals

Cue Type	Cue Size	Avg. Retrieval Time	Std. Deviation
Underspecified	1	0.1887 msec	0.0216 msec
Full word sense	7	0.2973 msec	0.0108 msec

100x faster retrievals on a comparable set of cues scaling to a 3x larger semantic store

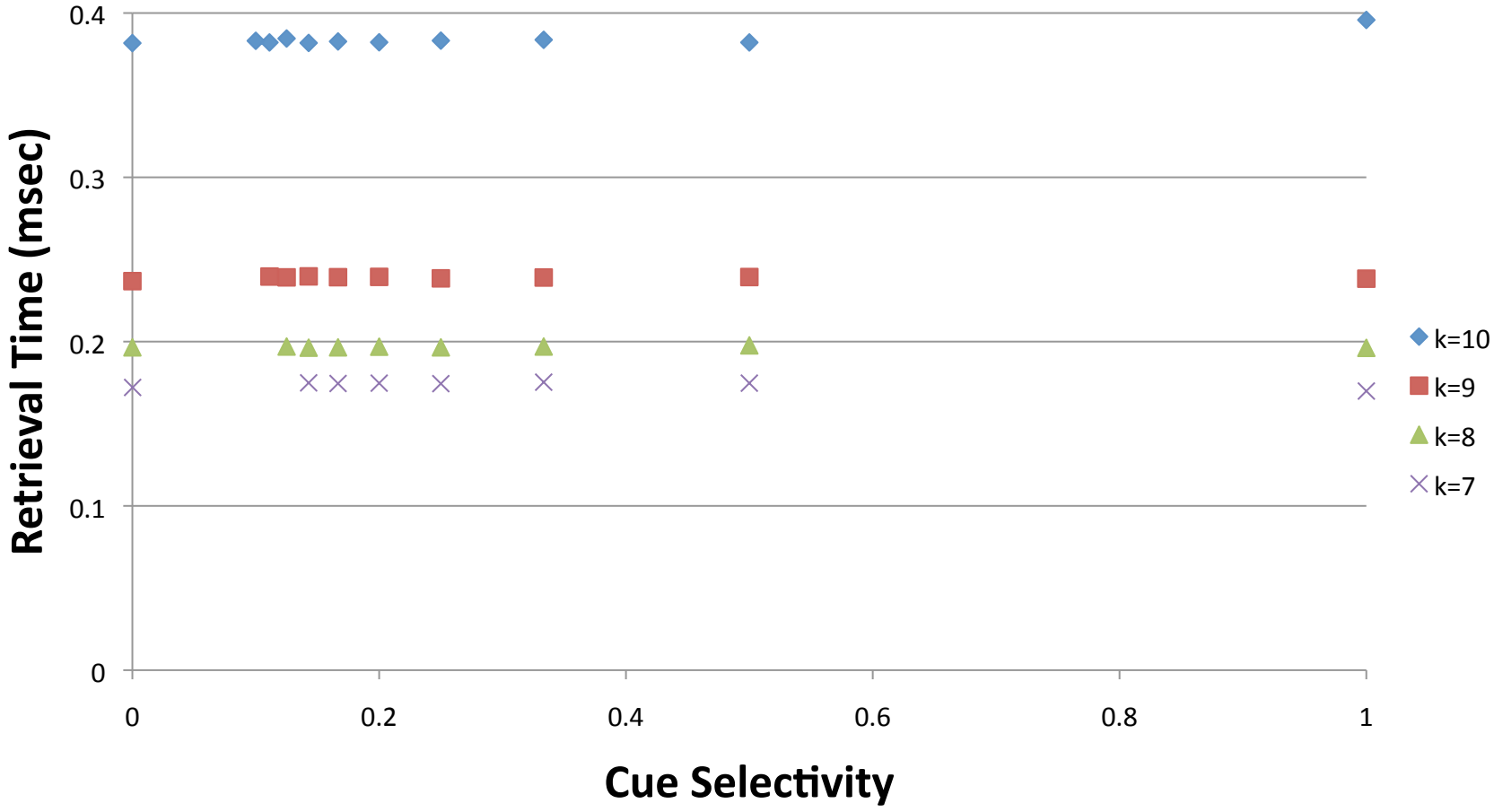
Evaluation: Synthetic

Focus on scaling

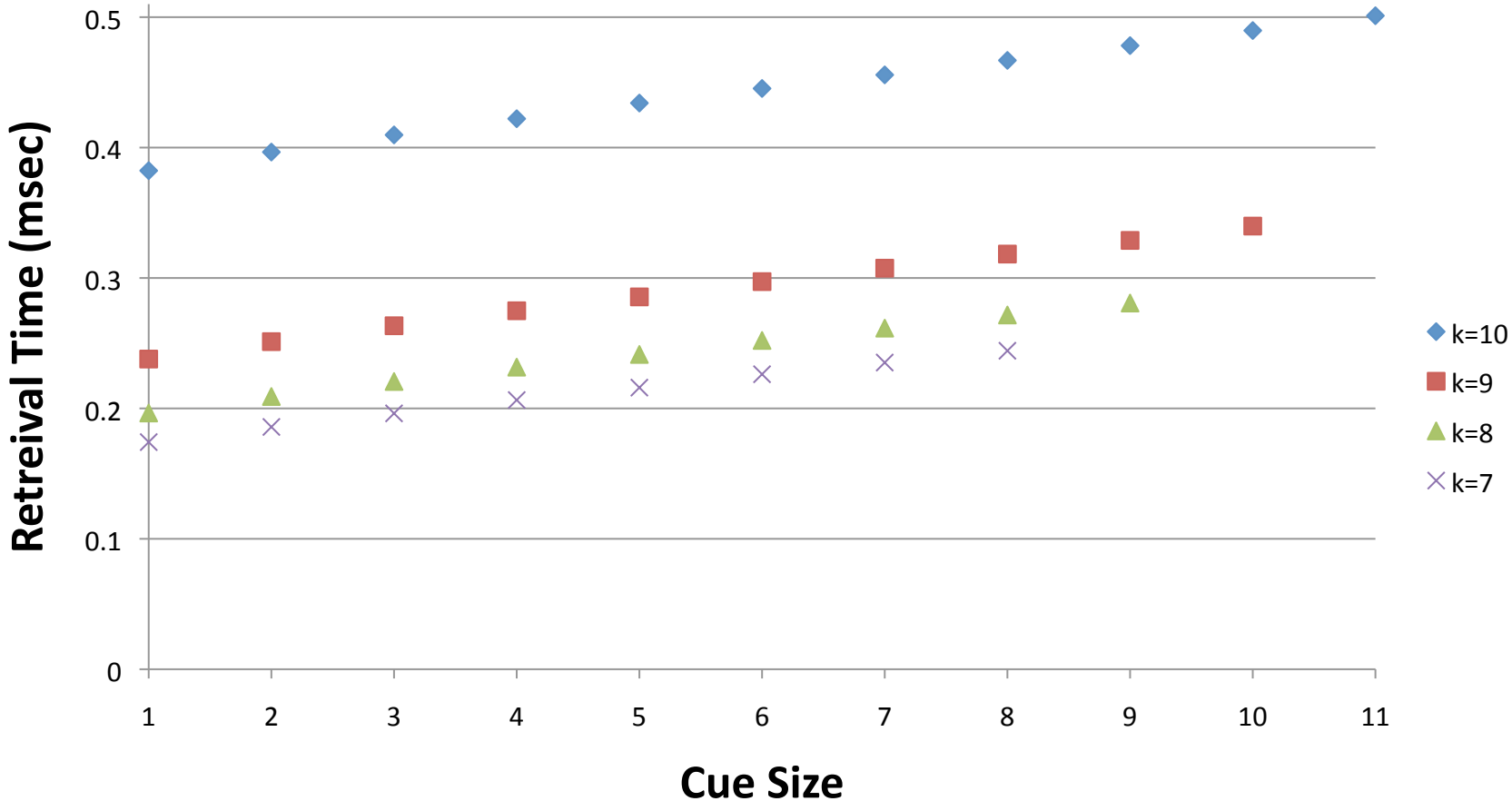
- Generator with single size parameter (k)
- Easily construct queries to control for selectivity

k	Nodes: k!	Edges: [k+1]!	Store Size (MB)
7	5K	40K	3.00
8	40K	362K	27.81
9	362K	3.6M	291.95
10	3.6M	39.9M	2048.00

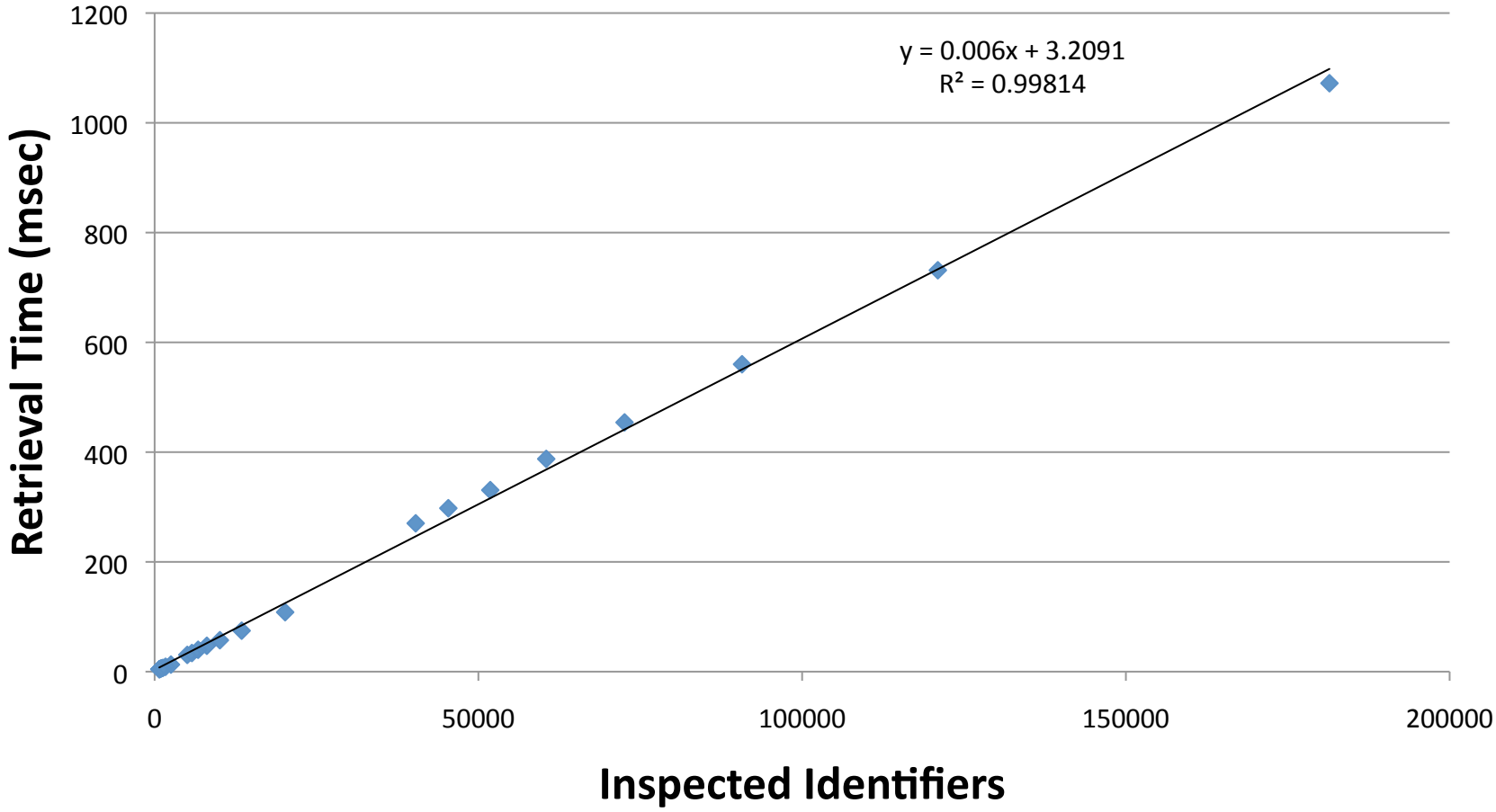
Underspecified Cues (individual features)



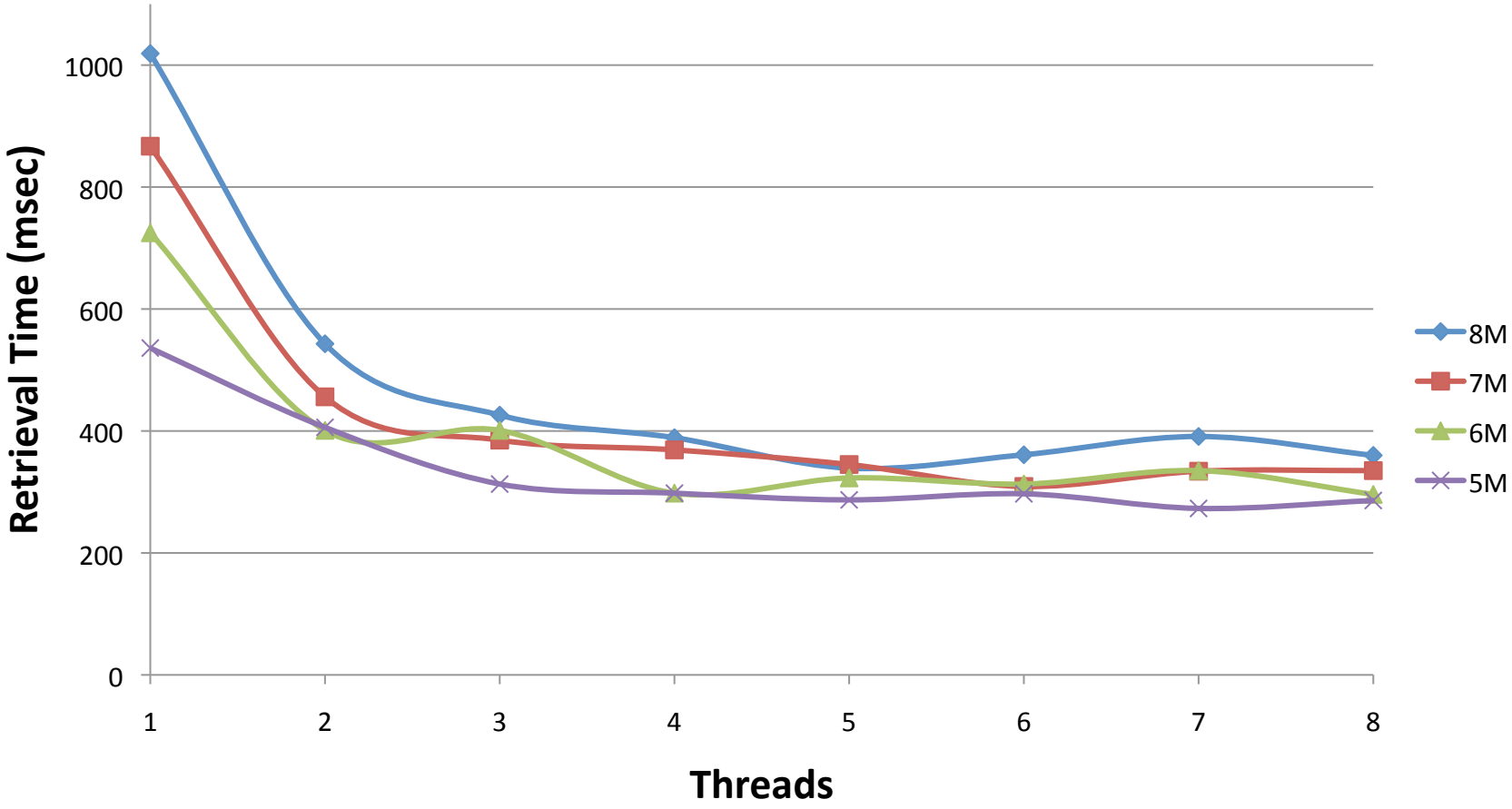
Successful Cues



Worst-Case Failure



Benefits of Parallelism



Evaluation

Nuggets

- Demonstration of semantic retrievals scaling to very large, synthetic and real-world stores
- Available in Soar 9.3.0

Coal

- Worst-case performance is not tractable
  - Parallelism
  - Heuristic search
- No model of typical cues and how they could affect performance of retrievals
- Limited retrieval bias