A normal form for first-order logic over physically feasible data structures

Steven Lindell
Haverford College
Department of Computer Science

Questions

1. What is the simplest kind of vocabulary retaining the full expressive power of first-order logic?
2. What (infinite) classes of finite data structures are physically realizable (i.e. scale feasibly)?

- **Simplify**: find a “natural” more elegant approach which does not use actual relations in the classical sense – just predicates and functions!

Singulary logic

Vocabulary: one-place symbols (no commas allowed)

$\mathbf{S} = \langle S, P_1, \ldots P_m, f_1, \ldots f_k \rangle$

Monadic predicates:	$P(\_)$	$P \subseteq S$	*unary* relation
Monadic functions:	$f(\_)$	$f: S \to S$	*unary* mapping

Atomic forms:	$P(F(x))$	$F = f \circ \ldots$
involving <u>equality</u>:	$F(y) = G(z)$	[not a true relation]

Transformations

<u>L-structure of arity k</u>	<u>singulary L^*-structure</u>
$\mathbf{S} = \langle S, R_1, \ldots, R_m \rangle$	$\mathbf{S^*} = \langle S^*, P_1, \ldots, P_m, f_1, \ldots, f_k \rangle$

a *k-ary* <u>fact</u> $R_i(s_1, \ldots s_j \ldots, s_k)$ where $s_j \in S$ $\mapsto$ $\langle s_1, \ldots s_j \ldots, s_k \rangle \in P_i$, $\downarrow f_j$, $\langle s_j, \ldots, s_j \rangle \in S^k$

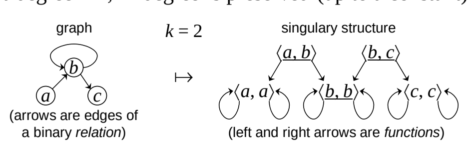
In a sense, it preserves the *true size* of **S** (as a database):

$|\mathbf{S}| = |S| + |R_1| + \ldots + |R_m| \sim |\mathbf{S^*}| = O(|S^*|)$.

Theorem: For each first-order L-sentence θ there is a first-order L^*-sentence θ^* such that $\mathbf{S} \models \theta$ iff $\mathbf{S^*} \models \theta^*$.

Bounded-degree classes

out-degree = k, in-degree is preserved (up to a constant)



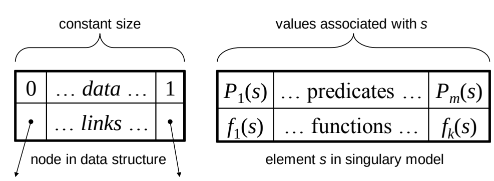
Definition: The degree of **S** is that of its *Gaifman graph*, relating a and b iff they appear jointly in **S**.

Theorem: A class of structures is of bounded degree iff the corresponding class of singulary structures is also.

Models for data structures

Information is stored inside nodes of fixed width.

- The number of nodes determines *size* of structure.
- The links between nodes determines its *shape*.



Physical feasibility

Assume actual implementations use resources:

- *matter* (quantized) occupies *space* at limited density

Total amount of resources $\propto$ size (no. of nodes).

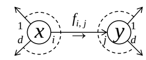
Nodes are uniform (no *overcrowding* allowed): $O(1)$ *bits* (mass) and $O(1)$ *connections* (space)

Unbounded in-degree is physically untenable, even though no bound on lengths of connections.

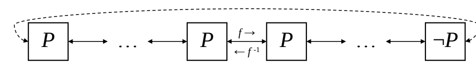
- Conclusion: feasible structures have reversible links

Reversible $\mapsto$ bijections

Construct partial injections $f_{i,j}^{-1} = f_{j,i}$ (where defined):



Convert (finite) f-chains to cycles using *wrap-around*:



Removes *nils* – results in an *bijective* singular model:

$$\langle D, P_1, \ldots, P_m, f_1, \ldots, f_k, f_1^{-1}, \ldots, f_k^{-1} \rangle$$

Normal form result

Definition: For each $n \geq 1$, the *numerical quantifier* $\exists^n x \ldots$ means "there are (at least) n x's such that …".

Theorem: Any formula $\theta(x_1, \ldots x_l)$ in bijective singular logic can be rewritten as a Boolean combination of:

formulas:	$\alpha(x_1, \ldots x_l)$	where α is atomic,
sentences:	$\exists^n x\ \beta(x)$	β is quantifier-free

Proof: syntactic – the key idea is to solve equations:

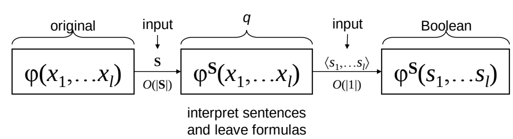
$$F(x) = G(y) \quad \mapsto \quad x = F^{-1} \circ G(y)$$

Linear-time evaluations

Given: a first-order formula $\varphi(x_1, \ldots, x_l)$.

Problem: to compute $\{\langle s_1, \ldots, s_l \rangle : \mathbf{S} \models \varphi(s_1, \ldots, s_l)\}$.

Solution: Calculate a *partial interpretation* query $q = \varphi^{\mathbf{S}}(x_1, \ldots, x_l)$ using time linear in $|\mathbf{S}|$, so that $\varphi^{\mathbf{S}}(s_1, \ldots, s_l)$ can be computed in constant time (i.e. q demonstrates an immediate knowledge of the answer). The method:



Answers

Singular vocabularies are the simplest kind of signature which retain the full power of FOL.

- Any structure can be mapped to a singular one, in which size and degree are preserved uniformly.

Data structures are really just singular models.

- Physically implementable ones are reversibly-linked, in which every function can be made bijective.

Elementary definability over asymptotically scalable classes is much simpler than it appears.

- Using numerical quantifiers, no nesting is required, and all formulas can be evaluated in linear-time.

References

[Church] *Introduction to Mathematical Logic*, 1956.

[Hanf] 'Model-theoretic methods in the study of elementary logic' 1965.

[Gaifman] 'On local and non-local properties' 1982.

[Seese] "Linear time computable problems and first-order descriptions" 1996.

[Immerman] *Descriptive Complexity*, 1999.

[Libkin] *Elements of Finite Model Theory*, 2005.

A full paper is available from my web page.