General Acyclicity and Cyclicity Notions for the Disjunctive Skolem Chase (Extended Abstract)

Lukas Gerlach[1], David Carral[2]

[1]Knowledge-Based Systems Group, TU Dresden, Nöthnitzer Straße 46, 01062 Dresden, Germany
[2]LIRMM, Inria, University of Montpellier, CNRS, Montpellier, France

Abstract
Query entailment over ontologies is a fundamental decision problem in the field of knowledge representation and reasoning. The disjunctive (skolem) chase is a sound and complete reasoning procedure that solves this problem for boolean conjunctive queries over the powerful first-order logic fragment of disjunctive existential rules. Yet, termination of the procedure is an undecidable problem. We develop novel acyclicity and cyclicity notions for this procedure; that is, we develop sufficient conditions to determine chase termination and non-termination. Our empirical evaluation on translated OWL ontologies shows that our novel notions are significantly more general than existing criteria.

Keywords
Decidability and Complexity of Reasoning, Rule-Based Systems, Query Answering

1. Introduction

We consider the problem of query entailment over ontologies, aka. knowledge bases (KBs) that are based on *disjunctive existential rules*. The latter form a very expressive fragment of first-order logic. The problem can then be defined as follows:

- Input: a set $\mathcal{R}$ of rules, a set $\mathcal{F}$ of facts, and a boolean conjunctive query (BCQ) γ.
- Output: yes iff γ is entailed by $\langle\mathcal{R}, \mathcal{F}\rangle$ under standard first-order semantics.

The *disjunctive skolem chase* [1, 2] can solve entailment of a BCQ γ by computing a universal model set and checking if γ is satisfied by every model in this set. Unfortunately, BCQ entailment as well as termination of the chase is undecidable [3, 4, 5]. Therefore, we study *acyclicity and cyclicity notions*; i.e., sufficient conditions for chase termination or non-termination, respectively. In this sense, a rule set $\mathcal{R}$ is terminating if the chase terminates on every KB of the form $\langle\mathcal{R}, \mathcal{F}\rangle$.

In this extended abstract, we briefly discuss our published paper [6], where we introduced *disjunctive MFA/MFC (DMFA/DMFC)* as novel (a)cyclicity notions for the disjunctive skolem chase, and show that these are more general than previous criteria in practice. We take inspiration from *model faithful (a)cyclicity (MFA/MFC)* and *restricted MFA/MFC (RMFA/RMFC)* [7, 8], which tackle (non-)termination for skolem chase and disjunctive restricted chase, respectively.

DL 2023: 36th International Workshop on Description Logics, September 2–4, 2023, Rhodes, Greece
lukas.gerlach@tu-dresden.de (L. Gerlach); david.carral@inria.fr (D. Carral)
https://kbs.inf.tu-dresden.de/lug (L. Gerlach); https://www-sop.inria.fr/members/David.Carral/ (D. Carral)
0000-0003-4566-0224 (L. Gerlach); 0000-0001-7287-4709 (D. Carral)

2. Disjunctive Skolem Chase

We define $\texttt{Cons}$, $\texttt{Vars}$, $\texttt{Funs}$, and $\texttt{Preds}$ to be mutually disjoint, finite (albeit large enough) sets of *constants*, *variables*, *function symbols*, and *predicates*, respectively, such that every $s \in \texttt{Funs} \cup \texttt{Preds}$ has an arity $\mathsf{ar}(s) \geq 1$. The set $\texttt{Terms}$ of *terms* includes $\texttt{Cons} \cup \texttt{Vars}$ and contains $f(t_1, \ldots, t_n)$ for every $n \geq 1$, $f \in \texttt{Funs}$ with $\mathsf{ar}(f) = n$, and $t_1, \ldots, t_n \in \texttt{Terms}$. A term t is *functional* if $t \notin \texttt{Cons} \cup \texttt{Vars}$. We write lists $t_1, \ldots, t_n$ of terms as $\vec{t}$. A term s is a *subterm* of a term t if $t = s$, or t is of the form $f(\vec{s})$ and s is a subterm of some term in $\vec{s}$. A term is *cyclic* if it has a subterm of the form $f(\vec{s})$ such that $f \in \texttt{Funs}(\vec{s})$ where $\texttt{Funs}(\vec{s})$ denotes the function symbols in $\vec{s}$. An *atom* is a first-order formula of the form $P(\vec{t})$ where P is a $|\vec{t}|$-ary predicate and $\vec{t}$ is a term list. A *fact* is a variable-free atom. For a formula ϕ, we write $\phi[\vec{x}]$ to indicate that $\vec{x}$ is the set of all free variables in ϕ; i.e., variables that are not explicitly quantified.

Definition 1. *A* (disjunctive existential) rule *is a constant- and function-free formula of the form*

$$\forall \vec{w}, \vec{x}. \big(\beta[\vec{w}, \vec{x}] \to \bigvee\nolimits_{i=1}^{n} \exists \vec{y}_i . \eta_i[\vec{x}_i, \vec{y}_i]\big) \tag{1}$$

where $n \geq 1$; $\vec{w}$, $\vec{x}$, $\vec{y}_i$ are pairwise disjoint lists of variables; $\bigcup_{i=1}^{n} \vec{x}_i = \vec{x}$; $\vec{x}_i$ are non-empty; and β, η_i are non-empty conjunctions of atoms (featuring exactly the denoted variables).

A rule ρ as in (1) is *deterministic* if $n = 1$, *generating* if it features at least one existential variable, and *datalog* if it is deterministic and not generating. We denote a skolemized rule (head) with $\mathsf{sk}(\rho)$ ($\mathsf{sk}(\eta_i)$), i.e. all existential variables are replaced by skolem terms.

A *(boolean conjunctive) query* γ is a first-order formula of the form $\exists \vec{y}.\beta[\vec{y}]$ with β a non-empty conjunction of function-free atoms. A *knowledge base* (KB) $\mathcal{K}$ is a pair $\langle \mathcal{R}, \mathcal{I} \rangle$ with $\mathcal{R}$ a rule set and $\mathcal{I}$ an *instance*; that is, a function-free fact set. A *(ground) substitution* σ is a partial function that maps variables to variable-free terms. We use $[x_1/t_1, \ldots, x_n/t_n]$ to denote the substitution that maps the variable x_i to the term t_i for every $1 \leq i \leq n$. For a first-order formula ϕ, let $\phi\sigma$ be the formula that results from replacing every occurrence of every variable x in the domain of σ in ϕ with $\sigma(x)$. A *trigger* λ is a pair $\langle \rho, \sigma \rangle$ with ρ a rule as in (1) and σ a substitution with domain $\vec{w} \cup \vec{x}$. The trigger λ is *loaded* for a fact set $\mathcal{F}$ if $\beta\sigma \subseteq \mathcal{F}$; it is *active* for $\mathcal{F}$ if $\mathsf{sk}(\eta_i)\sigma \nsubseteq \mathcal{F}$ for all $1 \leq i \leq n$. Let $\mathsf{out}_i(\lambda) = \mathsf{sk}(\eta_i)\sigma$ for $1 \leq i \leq n$; $\mathsf{out}(\lambda) = \{\mathsf{out}_i(\lambda) \mid 1 \leq i \leq n\}$ be the *output* of λ. A fact set $\mathcal{F}$ is *closed under* a rule ρ if no trigger with ρ is loaded and active for $\mathcal{F}$.

Definition 2. *A* (skolem) chase tree (CT) *of a KB $\langle \mathcal{R}, \mathcal{I} \rangle$ is a directed tree labelled with fact sets such that (1) the root label is $\mathcal{I}$; (2.1) for every non-leaf vertex v, there is a trigger λ with a rule in $\mathcal{R}$ that is loaded and active for the label L of v such that, for every $F \in \mathsf{out}(\lambda)$, some child of v is labelled with $L \cup F$; (2.2) if λ features a non-datalog rule, then L is closed under all datalog rules in $\mathcal{R}$; (3.1) leaf vertex labels are closed under the rules in $\mathcal{R}$; and (3.2) for a trigger λ with a rule in $\mathcal{R}$, there is a $k \geq 1$ such that λ is not loaded or not active for labels of vertices of depth at least k.*

Conditions (3.1) and (3.2) ensure *fairness.* A KB *terminates* if it only admits finite CTs. A rule set $\mathcal{R}$ *terminates* if every KB $\langle \mathcal{R}, \mathcal{I} \rangle$ terminates; it *never-terminates* if some KB $\langle \mathcal{R}, \mathcal{I} \rangle$ only admits infinite CTs. It is undecidable to determine if $\mathcal{R}$ terminates [4]. The *result* of a CT T is the set of all fact sets that can be constructed via the union of all labels in a maximal path in T.

Proposition 1. *Consider the result $\mathfrak{R}$ of some CT of a $\mathcal{K}$. Then, $\mathcal{K}$ entails a query $\gamma = \exists \vec{y}.\beta$ iff $\mathcal{F} \models \gamma$ for every $\mathcal{F} \in \mathfrak{R}$ iff for every $\mathcal{F} \in \mathfrak{R}$ there is a substitution σ with $\beta\sigma \subseteq \mathcal{F}$.*

3. Acyclicity and Cyclicity

For detecting (never-)termination, we extend MFA [7] and MFC [8] towards DMFA and DMFC [6] with ideas from RMFA and RMFC [8]. We use the latter because, opposed to the skolem chase, the disjunctive skolem chase and restricted chase both have the property that more facts can remove activeness of triggers e.g. if a head-disjunct is already present. (see Proposition 2). For *acyclicity*, we start with a naive extension of MFA for disjunctions.

Definition 3 ([6, Section 3.1])**.** *A rule set* $\mathcal{R}$ *is MFA if* $\mathsf{MFA}(\mathcal{R})$ *does not feature a cyclic term, where* $\mathsf{MFA}(\mathcal{R})$ *is the minimal fact set that contains the* critical instance $\mathcal{I}_\star = \{P(\star, \ldots, \star) \mid P \in \texttt{Preds}\}$ *(with the special constant* $\star$*) and* $\mathsf{out}_1(\langle \rho^\wedge, \sigma \rangle)$ *for every trigger* $\langle \rho, \sigma \rangle$ *that is loaded for* $\mathsf{MFA}(\mathcal{R})$*. Here,* $\rho^\wedge$ *results from* ρ *by replacing all disjunctions with conjunctions.*

Note that $\mathsf{out}_1(\langle \rho^\wedge, \sigma \rangle)$ denotes the output of the first (and only) head-disjunct of $\rho^\wedge$ with σ applied. Intuitively, $\mathsf{MFA}(\mathcal{R})$ is an overapproximation of the facts in every possible CT of any KB with $\mathcal{R}$. Without cyclic terms, the number of terms and facts is finite so $\mathcal{R}$ is terminating. MFA makes use of all triggers that are loaded without considering activeness. Indeed, deterministic triggers can always be considered active because of the following key property:

Proposition 2 ([6, Lemma 15])**.** *A deterministic trigger output occurs in a CT if it is loaded.*

This holds because a loaded deterministic trigger is only not active if its exact output is already present, which is not true for disjunctive triggers. However, it is safe to ignore disjunctive triggers that are *blocked* [6, Definition 8], ensuring the following property [6, Lemma 7]: If a trigger λ is blocked for $\mathcal{R}$, then λ is not active whenever it is loaded in any CT of a KB with $\mathcal{R}$.

Definition 4 ([6, Definitions 9,10])**.** *For a rule set* $\mathcal{R}$*, let* $\mathsf{DMFA}(\mathcal{R})$ *be defined as* $\mathsf{MFA}(\mathcal{R})$ *but only use of triggers that are not blocked. If no cyclic term occurs in* $\mathsf{DMFA}(\mathcal{R})$*, then* $\mathcal{R}$ *is* DMFA.

Theorem 1 ([6, Corollary 5, Theorem 10])**.** *If a rule set* $\mathcal{R}$ *is (D)MFA, then* $\mathcal{R}$ *terminates.*

Example 1 ([6, Example 2])**.** *The following rule set* $\mathcal{R}$*, which is a slightly simplified subset of rule set* `00007.owl` *in the Oxford Ontology Repository OXFD (see Section 4), is DMFA but not MFA:*

(1) *evidence*$(x) \rightarrow \exists w.$*Confidence*$(x, w)$ (3) *Confidence*$(x, y) \rightarrow$ *confidence*(y)

(2) *XRef*$(x, y) \rightarrow$ *evidence*$(x) \vee$ *confidence*(x) (4) *Confidence*$(x, y) \rightarrow \exists z.$*XRef*$(y, z)$

Consider a KB $\langle \mathcal{R}, \mathcal{I} \rangle$ *and suppose that* $\lambda = \langle (2), [x/f_w(t), y/f_z(f_w(t))] \rangle$ *is loaded in some CT* T *of* $\langle \mathcal{R}, \mathcal{I} \rangle$*. Then, Confidence*$(t, f_w(t))$ *occurs in* T *since* $f_w(t)$ *may only be introduced via* (1)*. Since datalog rules are prioritised (2.2 in Definition 2), confidence*$(f_w(t))$ *is added by* (3)*. But then,* λ *cannot be active, so* λ *is blocked! Hence,* $\mathsf{MFA}(\mathcal{R})$ *features a cyclic term but* $\mathsf{DMFA}(\mathcal{R})$ *does not.*

For *cyclicity*, i.e. MFC membership of a rule set $\mathcal{R}$, we compute a fact set $\mathsf{MFC}(\mathcal{R}, \rho)$ for every generating rule $\rho \in \mathcal{R}$. We can sometimes verify that ρ can be applied infinitely many times when we start on the weakest instance to which this rule can be applied [6, Section 4.1]. MFC ignores disjunctive rules in the process and relies heavily on Proposition 2. To support disjunctive rules, a key aspect is to port the notion of *unblockable triggers* from RMFC, which ensure the property from Proposition 2 [6, Definition 17]. We obtain DMFC and prove its correctness [6, Definitions 19,20, Theorem 18]. Furthermore, checking DMFA or DMFC is 2ExpTime-complete and reasoning with DMFA rule sets is coN2ExpTime-complete [6, Theorems 12,13,19].

	#∃	# tot.	# fin.	MFA	DMFA	DMFA^2	MFC	DMFC^s
OXFD	1–19	37	36	21	28	28	4	8
	20–99	18	17	3	3	3	10	14
	100+	82	26	4	6	6	14	19
	1+	**137**	**79**	**28 (35%)**	**37 (46%)**	**37 (46%)**	**28 (35%)**	**41 (51%)**
ORE15	1–19	103	98	51	66	66	18	31
	20–99	119	105	32	33	35	54	69
	100–999	278	219	5	6	119	89	100
	1–999	**500**	**422**	**88 (20%)**	**105 (24%)**	**220 (52%)**	**161 (38%)**	**200 (47%)**
MOWL	1–19	1361	1283	676	725	732	173	515
	20–99	894	740	104	114	121	301	610
	100–299	448	254	25	25	111	103	143
	1–299	**2703**	**2277**	**805 (35%)**	**864 (37%)**	**964 (42%)**	**577 (25%)**	**1268 (55%)**

Table 1
Skolem Chase Termination: Rule Sets with at least one Disjunctive and one Generating Rule

4. Evaluation & Outlook

We refine some introduced notions for the evaluation: We consider DMFA^2, looking for cyclic terms of higher nesting depth, and DMFC^s; a simplified version of DMFC. We obtain the rule sets via normalization and translation of OWL ontologies [7, Section 6] from OXFD, ORE15, and MOWL.[1] We drop OWL axioms that require equality ("at-most restrictions" and "nominals"). The tools, rule sets, and results of the evaluation are available online.[2]

We set a timeout of 30 minutes for each check and only consider rule sets for which all checks finished; we indicate the number of attempted (# tot.) vs finished (# fin.) rule sets. We group results by the number of generating rules (#∃). The percentage of finished rule sets that are fully classified by MFA and MFC for OXFD, ORE15, and MOWL are 70%, 58%, and 60%, respectively. With DMFA^2 and DMFC^s, we achieve 97%, 99%, and 97%.

For future work, we would like to develop a normalisation procedure that preserves both query entailment and chase termination. As a long term goal, we would like to adapt our notions so they can be applied in other areas of knowledge representation and reasoning. For instance, we believe that we can use our ideas to *(i)* show if an ASP program with function symbols does or does not admit a finite solution or *(ii)* determine if DPLL(T) algorithms used in automated theorem proving will terminate or not for many real-world inputs.

Acknowledgments

Lukas is funded by DFG grant 389792660 (TRR 248), BMBF grant ITEA-01IS21084 (InnoSale), BMBF and DAAD grant 57616814 (SECAI), and the Center for Advancing Electronics Dresden (cfaed). David is funded by the ANR project CQFD (ANR-18-CE23-0003).

[1]https://www.cs.ox.ac.uk/isg/ontologies/ https://doi.org/10.5281/zenodo.18578 https://doi.org/10.5281/zenodo.16708
[2]https://doi.org/10.5281/zenodo.7375461

References

[1] P. Bourhis, M. Manna, M. Morak, A. Pieris, Guarded-based disjunctive tuple-generating dependencies, ACM Transactions On Database Systems (TODS) 41 (2016) 27:1–27:45.

[2] B. Marnette, Generalized schema-mappings: from termination to tractability, in: J. Paredaens, J. Su (Eds.), Proceedings of the 28th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS 2009), USA, ACM, 2009, pp. 13–22.

[3] C. Beeri, M. Y. Vardi, The implication problem for data dependencies, in: S. Even, O. Kariv (Eds.), Proceedings of the 8th International Colloquium on Automata, Languages and Programming (ICALP 1981), Israel, volume 115 of *Lecture Notes in Computer Science*, Springer, 1981, pp. 73–85.

[4] T. Gogacz, J. Marcinkowski, All-instances termination of chase is undecidable, in: J. Esparza, P. Fraigniaud, T. Husfeldt, E. Koutsoupias (Eds.), Proceedings of the 41st International Colloquium on Automata, Languages, and Programming (ICALP 2014), Denmark, Part II, volume 8573 of *Lecture Notes in Computer Science*, Springer, 2014, pp. 293–304.

[5] G. Grahne, A. Onet, Anatomy of the chase, Fundamenta Informaticae 157 (2018) 221–270.

[6] L. Gerlach, D. Carral, General acyclicity and cyclicity notions for the disjunctive skolem chase, in: B. Williams, Y. Chen, J. Neville (Eds.), Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence, 2023, USA, AAAI Press, 2023, pp. 6372–6379. URL: https://ojs.aaai.org/index.php/AAAI/article/view/25784. doi:10.1609/aaai.v37i5.25784.

[7] B. Cuenca Grau, I. Horrocks, M. Krötzsch, C. Kupke, D. Magka, B. Motik, Z. Wang, Acyclicity notions for existential rules and their application to query answering in ontologies, Journal of Artificial Intelligence Resesearch (JAIR) 47 (2013) 741–808.

[8] D. Carral, I. Dragoste, M. Krötzsch, Restricted chase (non)termination for existential rules with disjunctions, in: C. Sierra (Ed.), Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017), Australia, ijcai.org, 2017, pp. 922–928.

[9] B. Parsia, N. Matentzoglu, R. S. Gonçalves, B. Glimm, A. Steigmiller, The OWL reasoner evaluation (ORE) 2015 resources, in: P. Groth, E. Simperl, A. J. G. Gray, M. Sabou, M. Krötzsch, F. Lécué, F. Flöck, Y. Gil (Eds.), The Semantic Web - ISWC 2016 - 15th International Semantic Web Conference, Japan, 2016, Proc, Part II, volume 9982 of *Lecture Notes in Computer Science*, 2016, pp. 159–167.

[10] N. Matentzoglu, D. Tang, B. Parsia, U. Sattler, The manchester OWL repository: System description, in: M. Horridge, M. Rospocher, J. van Ossenbruggen (Eds.), Proceedings of the 13th International Semantic Web Conference (ISWC 2014), Posters & Demonstrations Track, Italy, volume 1272 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2014, pp. 285–288.

[11] L. Gerlach, D. Carral, General acyclicity and cyclicity notions for the disjunctive skolem chase - evaluation material, 2022. doi:10.5281/zenodo.7375461.