

Semantics-Aware Caching for Concept Learning

1

Abstract. Concept learning is a form of supervised machine learning that operates on knowledge bases in description logics. State-of-the-art concept learners often rely on an iterative search through a countably infinite concept space. In each iteration, they retrieve instances of candidate solutions to select the best concept for the next iteration. While simple learning problems might require a few dozen instance retrieval calls to find a fitting solution, complex learning problems might necessitate thousands of calls. We alleviate the resulting runtime challenge by presenting a semantics-aware caching approach. Our cache is essentially a subsumption-aware map that links concepts to a set of instances via crisp set operations. Our experiments on 5 datasets with 4 symbolic reasoners, a neuro-symbolic reasoner, and 5 popular pagination policies demonstrate that our cache can reduce the runtime of concept retrieval and concept learning by an order of magnitude while being effective for both symbolic and neuro-symbolic reasoners.

2
3
4
5
6
7
8
9
10
11
12
13
14

1 Introduction

15

Knowledge bases are now first-class citizens of the Web, with roughly 50% of websites hosting assertions from RDF knowledge bases.¹ Given the advent of machine learning across disciplines and on corpora of increasing size, it stands to reason that the over 10^{11} assertions available on the Web need to be made amenable to machine learning. Embedding approaches and geometric deep learning are used to map the constituents of knowledge bases to vectorial representations [32]. However, these approaches do not make explicit of the semantics of RDF knowledge bases. In contrast, concept learning (also called class expression learning, short: CEL) [10, 25, 14] is designed to exploit the explicit semantics of description logics to derive ante-hoc explainable models.

16
17
18
19
20
21
22
23
24

The goal of CEL is, given a set of positive and negative examples, to describe the set of positive examples and not the set of negative examples. This task is equivalent to learn a class expression that describes individuals provided as positive examples [20, 19]. This problem is inherently challenging, as most CEL algorithms operate over a quasi-ordered infinite hypothesis space defined by the expressivity of the underlying description logic (short: DL). The majority of state-of-the-art CEL methods adopt a top-down refinement strategy, which initiates the search from the most general concept (e.g., \top) and incrementally specializes it through a sequence of refinement steps [20]. This iterative process continues until a hypothesis that best approximates the target concept in terms of covering positive examples while avoiding negatives is reached.

25
26
27
28
29
30
31
32
33
34

CEL has demonstrated its effectiveness across various application domains. In software engineering, it has been employed to support the development of software information systems by facilitating semantic integration and information retrieval [22]. In ontology engineering, CEL contributes to the (semi-)automatic construction, refinement, and validation of ontologies by learning new concept definitions from instance

35
36
37
38
39

¹ <https://webdatacommons.org/structureddata/>

40 data [20]. In the medical field, CEL has been applied to tasks such as the prediction of
 41 protein functions, where logical definitions of protein classes enhance interpretability
 42 and biological insight [18]. Additionally, it has been used in content recommendation
 43 systems, where learned class expressions help model user preferences and content cat-
 44 egories in a transparent and semantically meaningful manner [24].

45 A central bottleneck in CEL is the repeated use of DL reasoners to evaluate can-
 46 didate class expressions. Each candidate must be passed to a reasoner to retrieve its
 47 instances, an operation that is both computationally expensive and invoked thousands
 48 of times during the search. Accelerating this evaluation step, therefore, has a direct and
 49 substantial impact on the overall runtime of CEL algorithms. This work targets this
 50 bottleneck by accelerating the reasoner performance as a foundational component of
 51 efficient CEL.

52 We introduce a semantic caching mechanism designed to accelerate instance re-
 53 trieval in Description Logic reasoners. Our contribution can be summarized as follows:
 54 regardless of the underlying reasoner our approach can substantially improve perfor-
 55 mance by up to 80% for slower reasoners and by up to 20% for faster ones, provided
 56 that sufficient cache capacity² is available.

57 Traditional approaches, such as ELK [15], demonstrate how materialization-based
 58 techniques can drastically improve reasoning efficiency by precomputing subsumptions
 59 and class memberships. While our work shares with ELK the overarching goal of scala-
 60 bility, the two approaches are fundamentally different. ELK is itself a highly optimized
 61 reasoner restricted to the \mathcal{EL}^{++} description logic. In contrast, our method is not a new
 62 reasoner but rather a complementary caching layer that accelerates existing reasoning
 63 tasks. For this reason, a direct experimental comparison is not meaningful. Instead,
 64 our approach leverages the semantics of \mathcal{ALC} to systematically compose the set of in-
 65 stances of complex concepts. By reusing previously computed results through caching,
 66 our method avoids redundant reasoning calls and significantly reduces the computa-
 67 tional cost of instance retrieval of a given reasoner.

68 We demonstrate that this mechanism can be integrated with little effort into existing
 69 reasoning pipelines by evaluating the performance of 4 symbolic and 1 neuro-symbolic
 70 reasoners with and without our cache within the setting of concept retrieval and concept
 71 learning. Our results suggest that our caching approach can also reduce concept learning
 72 runtimes by up to three orders of magnitude. Our codes, datasets, and experimental
 73 setup can be found in the supplementary material³.

74 **2 Background**

75 **2.1 Concept learning algorithms**

76 CEL algorithms have significantly evolved during the last decade. Among the early
 77 approaches, CELOE [19], OCEL, and ELTL [6] are foundational examples of search-

² In our experiments, cache capacity refers to the number of distinct class expressions and their retrieved instances that can be stored in the cache dictionary, rather than physical memory usage.

³ <https://anonymous.4open.science/r/Cache-algorithm-for-concept-learners-F772/>

based CEL methods. These algorithms rely on refinement operators to iteratively generate and evaluate candidate class expressions and are all implemented into the DL-Learner framework [20]. OCEL uses heuristic rules to guide the search process, reducing the exploration of irrelevant or overly complex expressions. CELOE extends upon OCEL by improving the heuristic function, prioritizing syntactically shorter and semantically relevant expressions, which has made it one of the most effective search-based CEL algorithms within the DL-Learner framework [16].

EvoLearner [12], a more recent development, takes an evolutionary approach to CEL. Instead of relying only on refinement operators, EvoLearner initializes its population of candidate expressions through random walks on the knowledge graph, which are then converted into description logic concepts represented as abstract syntax trees. These candidate concepts are refined further using mutation and crossover operations, enabling EvoLearner to outperform traditional refinement operator-based methods like CELOE and OCEL in both efficiency and accuracy. Another example is ECII [26], a search-based algorithm that does not employ refinement operators, invoking a reasoner only once during execution. This design choice allows ECII to bypass the iterative computational overhead which is common in traditional methods.

As ontologies grow increasingly complex and data sets expand, the runtime limitations of these traditional approaches have become more apparent. Therefore, accelerating CEL has been crucial in the field of machine learning, with diverse methods explored to overcome the computational costs inherent to this domain. Recent efforts have focused on addressing these challenges through neural and neuro-symbolic methods. Neural Class Expression Synthesis (NCES) [16], represents a significant departure from traditional search-based methods. NCES employs neural networks to synthesize class expressions directly, avoiding the exhaustive refinement process and reducing computational overhead. NCES2 has further extended this approach [17], which broadens its applicability to more expressive description logics, such as \mathcal{ALCHIQ}^D , enabling CEL for more complex and expressive ontologies.

Another promising approach is Neuro-Symbolic Class Expression Learning (DRILL) [8], which addresses the limitations of heuristic-based methods like CELOE and OCEL. Traditional methods rely heavily on myopic heuristics and fixed rules for redundancy elimination and expression simplification, which can lead to suboptimal runtimes and memory inefficiencies. DRILL leverages deep Q-learning to dynamically adapt the search process, reducing runtimes and memory requirements while enhancing scalability. This reinforcement learning paradigm overcomes the drawbacks of fixed heuristic rules, providing a more flexible and efficient framework for CEL.

Other noteworthy approaches to CEL include CLIP [13], a pruning method that extends CELOE that integrates neural networks to predict an approximate solution length, thereby streamlining the concept learning process. Similarly, DL-FoCl [25], a variant of DL-Foil [9], introduces omission rates to prune the search space and further accelerate the learning process.

Despite recent progress, scaling CEL methods remains a major challenge due to persistent memory and runtime constraints [8, 16]. Caching mechanisms offer a promising solution by optimizing instance retrieval—the primary computational bottleneck

122 in CEL. The following subsection explores caching algorithms and eviction strategies
123 aimed at improving the scalability and efficiency of CEL systems.

124 **2.2 Cache algorithms with replacement policies**

125 Caching has long been recognized as a powerful technique to avoid redundant computa-
126 tions by reusing precomputed objects. By storing frequently accessed or computationally
127 expensive data for future use, caching systems significantly reduce computation
128 time and enhance overall efficiency. This concept finds widespread applications in various
129 domains, including storage systems, databases, and web servers [21].

130 Every caching mechanism requires dedicated memory space to store objects. How-
131 ever, memory resources are inherently limited, necessitating a strategy to manage the
132 space effectively when the cache becomes full. Objects already present in the cache
133 must be removed to accommodate new objects. This eviction process is governed by a
134 replacement policy, which selects the objects to evict from the cache. This policy plays
135 a crucial role in the efficiency of caching systems. The choice of a replacement policy
136 significantly impacts the effectiveness of a caching system. An optimal replacement
137 policy minimizes computation time and resource usage, leading to improved overall
138 system performance [27]. The theoretical ideal, known as Belady’s optimal replace-
139 ment policy [3], evicts the object that will not be accessed for the longest period in the
140 future. While Belady’s approach offers optimal performance, its practical implementa-
141 tion is infeasible because predicting future access is impossible in real-world scenarios.

142 Instead, practical cache replacement policies have been developed, each leverag-
143 ing specific heuristics to make eviction decisions. Among these are policies based on
144 frequency, where eviction decisions are determined by how often an object has been ac-
145 cessed. Notable examples include Least Frequently Used (LFU) and Most Frequently
146 Used (MFU) strategies, along with their variants. LFU evicts objects with the lowest ac-
147 cess frequency, while MFU does the opposite, removing the most frequently accessed
148 objects. Another widely adopted category of policies is based on recency, which con-
149 sider the timing of an object’s last access. For instance, the Least Recently Used (LRU)
150 strategy evicts objects that have not been accessed for the longest time, while the Most
151 Recently Used (MRU) strategy prioritizes removing the most recently accessed objects.
152 Some other methods, like ARC (Adaptive Replacement Cache) [21] have been proposed
153 and it combines both recency and frequency policies.

154 Policies based on the order of entrance into the cache form yet another category.
155 These include First In, First Out (FIFO) and Last In, First Out (LIFO) strategies. FIFO
156 removes the oldest object in the cache, whereas LIFO eliminates the most recently
157 added object. Following the same idea as the ensemble learning technique in machine
158 learning, [1] used LRU, LFU and FIFO together to decide which objects should be
159 evicted from the cache.

160 In our experiments, we aimed to evaluate the impact of policies based on recency
161 and order of entry on the performance of reasoners, and we observed that policies
162 based on frequency yielded similar results to recency based—specifically, MRU be-
163 haved similarly to MFU, and LRU to LFU. Additionally, we investigated the Random
164 Placement (RP) strategy, which involves randomly evicting an object from the cache to
165 make space for a new one. While modern approaches, such as those leveraging machine

learning, deep learning, and reinforcement learning [27, 33], have been explored to enhance cache replacement policies, these methods fall outside the scope of our study.

2.3 Metrics

The performance of a caching system is typically evaluated using the hit ratio and the total runtime with and without cache. The hit ratio is given by

$$\text{hit ratio} = \frac{H}{H + M}, \quad (1)$$

where H and M represent the number of cache hits (data found in the cache) and cache misses (data not found in the cache), respectively. The hit ratio measures the proportion of requests successfully served from the cache, indicating the efficiency of the caching strategy in reducing computational overhead. Conversely, the miss ratio ($1 - \text{hit ratio}$) quantifies the proportion of requests that could not be served from the cache, reflecting the instances where the system had to compute or retrieve the required data anew. Cache algorithms aim to maximize hit ratios and consequently minimize the miss ratios, as they signify reduced reliance on expensive computations and improved system efficiency.

Our work builds on these foundations by introducing a simple yet powerful caching mechanism tailored specifically for semantic reasoning tasks. By focusing on the subsumption relationships between concepts, we provide a novel perspective on integrating caching into CEL pipelines. The next section delves deeper into our proposed algorithms with a focus on how they can be used to accelerate instance retrieval in CEL.

3 Preliminaries

3.1 Description Logics and Knowledge Bases

Description Logics (DLs), initially introduced by Brachman [5], are a family of knowledge representation languages that can be seen as decidable fragments of first-order logic [20]. They serve as a formalism for representing structured and well-understood knowledge [22, 2]. DLs provide the foundation for modeling ontologies by enabling the expression of concepts (classes), roles (relationships), and individuals in a formal and logically grounded manner.

These logics range in expressivity, from simpler ones such as \mathcal{EL} and \mathcal{ALC} to more expressive variants like \mathcal{SHOIQ} and \mathcal{SROIQ} , which underlie standards such as OWL DL.

A Description Logic-based knowledge base (KB) is typically defined as a pair $(\mathcal{T}, \mathcal{A})$, where \mathcal{T} is the TBox (terminological box) and \mathcal{A} is the ABox (assertional box). The TBox contains general concept inclusions, i.e., axioms of the form $C \sqsubseteq D$, capturing hierarchical relationships between concepts. The ABox, on the other hand, contains instance-level assertions such as $C(a)$ (individual a is an instance of concept C) and $r(a, b)$ (individual a is related to b via role r), where C, D are concepts, r is a role, and a, b are individuals.

203 3.2 Class Expression Learning

204 Let N_I be a set of individuals. Given a KB \mathcal{K} , a description logic \mathcal{L} (e.g., \mathcal{ALC} ,
 205 \mathcal{SHOIN} , \mathcal{SROIQ}), and sets $E^+ \subseteq N_I$ of positive and $E^- \subseteq N_I$ of negative ex-
 206 amples, the goal of CEL is to construct a concept H in \mathcal{L} such that $\forall e^+ \in E^+ : \mathcal{K} \models$
 207 $H(e^+)$ and $\forall e^- \in E^- : \mathcal{K} \not\models H(e^-)$ [16, 8, 10]. A large proportion of the approaches
 208 to CEL [20, 19, 10, 25, 14] rely on iterative search through the infinite space $\mathcal{C}(\mathcal{L})$ of all
 209 concepts that can be built in \mathcal{L} . In each iteration, they compute candidate concepts C
 210 and their instances C^I by issuing an instance retrieval query to a reasoner. As pointed out
 211 by [14], this instance retrieval step is the main runtime bottleneck for concept learners.
 212 In particular, concept learners may need to retrieve the instances of thousands of concepts
 213 during each iteration, further exacerbating the computational load due to instance
 214 retrieval.

215 *Notation* Unless stated otherwise, we adopt the following notations. Given a concept C ,
 216 $Ret(C)$ denotes the retrieved instances of C . Given a set S , $|S|$ is the cardinal of S . N_I ,
 217 N_C , and N_R represent the set of named individuals, concepts, and roles, respectively.

218 4 Approach

219 While our approach is inherently adaptable to more expressive DL, this work specifi-
 220 cally concentrates on the \mathcal{ALC} DL. The syntax and semantics of \mathcal{ALC} are presented in
 221 Table 1.

Table 1: Syntax & semantics for \mathcal{ALC} concepts. \mathcal{I} stands for an interpretation with domain $\Delta^{\mathcal{I}}$.

Construct	Syntax	Semantics
Atomic concept	A	$A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
Atomic role	r	$r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$
Top concept	\top	$\Delta^{\mathcal{I}}$
Bottom concept	\perp	\emptyset
Negation	$\neg C$	$\Delta^{\mathcal{I}} \setminus C^{\mathcal{I}}$
Conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
Disjunction	$C \sqcup D$	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$
Existential restriction	$\exists r.C$	$\{x \mid \exists y.(x, y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
Universal restriction	$\forall r.C$	$\{x \mid \forall y.(x, y) \in r^{\mathcal{I}} \implies y \in C^{\mathcal{I}}\}$

222 4.1 Overview

223 Our approach is based on the definition of subsumption: Let C and D be two concepts.
 224 If $C \sqsubseteq D$, then $Ret(C) \subseteq Ret(D)$. While a naïve instance retrieval would check
 225 whether $C(a)$ holds for all $a \in N_I$ to compute $Ret(C)$, we are guaranteed to complete

the same computation by checking solely $|Ret(D)|$ individuals. Our intuition is that if $|Ret(D)| \ll |N_I|$, then the computation of $Ret(C)$ should be more time-efficient using our cache.

The formulation of our intuition elucidates the main challenge behind our approach, i.e., finding the most specific D with $C \sqsubseteq D$. Determining whether $C \sqsubseteq D$ is known to be NP-hard in \mathcal{ALC} [7] hence, we adopt the following set of heuristics:

- $C \equiv D \sqcap E \rightarrow C \sqsubseteq D$
- $C \equiv \exists r.C \wedge D \equiv \exists r.\top \rightarrow C \sqsubseteq D$
- $C \sqsubseteq D_1 \wedge \dots \wedge C \sqsubseteq D_n \rightarrow C \sqsubseteq \prod_{i=1}^n D_i$

The correctness of these heuristics is a direct consequence of the definition of concept subsumption. Based on these heuristics, we can now describe how our cache, which we denote \mathcal{A} , works. Given a concept C , we begin by searching for all D_i in \mathcal{A} 's key set with $C \sqsubseteq D_i$ using the heuristics aforementioned. Then, we retrieve $\bigcap_i D_i$ and check whether $C(a)$ holds for every individual $a \in \bigcap_i D_i$. In particular, if we can deduce $C \equiv D$ for any D in \mathcal{A} 's key set, then we simply fetch $Ret(D)$ from \mathcal{A} . The details of our approach are described in Algorithm 3 and the subsequent sections.

Although the semantic cache algorithm can operate without an initialization phase, we argue that performing this step is almost always worthwhile. When time permits, especially for datasets that are frequently used—it is beneficial to perform this step to improve effectiveness. We add the following precomputed instance retrieval results to \mathcal{A} :

- $Ret(A)$ for all $A \in N_C$
- $Ret(\neg A)$ for all $A \in N_C$
- $Ret(\exists r.C)$ for all $r \in N_R$ and $C \in N_C \cup \{\top\}$

The details of this implementation is shown in Algorithm 1.

Algorithm 1 initializeCache()

```

1: for all  $r \in N_R$  do
2:   store( $\exists r.\top$ ,  $Ret(\exists r.\top)$ )
3: end for
4: for all  $A \in N_C$  do
5:   store( $A$ ,  $Ret(A)$ )
6:   store( $\neg A$ ,  $N_I \setminus Ret(A)$ )
7:   for all  $r \in N_R$  do
8:     store( $\exists r.A$ ,  $Ret(\exists r.A)$ )
9:   end for
10: end for

```

251 4.2 Fetching from Cache and Space Management

252 The overall space management of the cache is handled by Algorithm 2 while the cache
 253 fetch is governed by Algorithm 3, which allows for the retrieval of instances of con-
 254 cepts. The algorithm begins by checking whether the queried concept C is a named
 255 concept. In this case, its instances are already precomputed and can be returned imme-
 256 diately without further reasoning. If C is not a named concept, the algorithm considers
 257 trivial cases (lines 2–5).

258 If the concept is neither named nor trivial, the algorithm proceeds recursively by de-
 259 composing the concept according to its semantics. For negations (line 6), the instances
 260 of the sub-concept are retrieved, and the complement with respect to N_I is taken. For
 261 conjunctions (line 8), the intersection of the instances of D and E is computed, while
 262 for disjunctions (line 8), their union is returned. Existential restrictions (lines 12–23)
 263 are handled by first computing the instances of the filler concept D and then iterating
 264 over all individuals $a \in N_I$ to check, for each candidate $b \in \text{instances}(D)$, whether the
 265 reasoner confirms that $r(a, b)$ holds. If such a witness is found, a is added to the result
 266 set. Universal restrictions (line 24) are reduced to a negated existential restriction, i.e.,
 267 $\neg(\exists r. \neg D)$, which is then handled by the previous case.

268 Once the recursive decomposition has been exhausted, the algorithm checks whether
 269 the result for C is already stored in the cache. If it is not, the reasoner is invoked to com-
 270 pute the instances of C , and the result is stored in the cache as a new entry (lines 28-30).
 271 This ensures that subsequent queries for the same concept can be answered without re-
 272 peating the reasoning process. Finally, the result is returned from the cache (line 32).

273 Cache space management is handled by the Algorithm 2. When a new result R is
 274 to be stored, the algorithm first checks whether the cache can accommodate it (line 1).
 275 If the maximum size is exceeded, a purge operation is triggered to free space (line 2).
 276 Afterwards, the pair (C, R) is added to the cache (line 4).

Algorithm 2 store (Concept C , Set R)

```

1: if  $R.size() + this.size() > maxSize$  then
2:   purge( $maxSize - R.size()$ )
3: end if
4: add( $C, R$ )

```

277 5 Experiments

278 Our experiments are designed to evaluate the performance gains provided by our caching
 279 system in both instance retrieval and concept learning tasks across multiple reason-
 280 ers. To better understand the contribution of semantic awareness in our approach, we
 281 compare three configurations: (i) no cache, (ii) our semantic cache, and (iii) the non-
 282 semantic-aware variant of the cache (Non-semantic Cache). This cache follows the
 283 same storage and retrieval interface as our semantic cache, but does not analyze the

Algorithm 3 Semantics-based implementation of fetchInstances(Concept C)

```

1: if  $C \notin N_C$  then
2:   if  $C \equiv \top$  then
3:     return  $N_I$ 
4:   else if  $C \equiv \perp$  then
5:     return  $\emptyset$ 
6:   else if  $C \equiv \neg D$  then
7:     return  $N_I \setminus \text{fetchInstances}(D)$ 
8:   else if  $C \equiv D \sqcap E$  then
9:     return  $\text{fetchInstances}(D) \cap \text{fetchInstances}(E)$ 
10:  else if  $C \equiv D \sqcup E$  then
11:    return  $\text{fetchInstances}(D) \cup \text{fetchInstances}(E)$ 
12:  else if  $C \equiv \exists r.D$  then
13:    Set instances =  $\text{fetchInstances}(D)$ 
14:    Set  $R = \emptyset$ 
15:    for  $a \in N_I$  do
16:      for  $b \in \text{instances}$  do
17:        if  $\text{reasoner.check}(r(a, b))$  then
18:           $R.\text{add}(a)$ 
19:          break
20:        end if
21:      end for
22:    end for
23:    return  $R$ 
24:  else if  $C \equiv \forall r.D$  then
25:    return  $\text{fetchInstances}(\neg(\exists r.\neg D))$ 
26:  end if
27: end if
28: if  $\neg \text{keys.contains}(C)$  then
29:   Set  $R = \text{reasoner.getInstances}(C)$ 
30:    $\text{store}(C, R)$ 
31: end if
32: return  $\text{get}(C)$ 

```

syntactic or semantic structure of the concept expression. Unlike our semantic imple- 284
 mentation, this baseline performs no decomposition into DL constructors, therefore 285
 corresponds to a naïve memoization mechanism and allows us to isolate the effect of 286
 semantic awareness from the effect of caching alone. 287

Since, to the best of our knowledge, no existing caching mechanisms have been 288
 specifically developed for class expression learning, nor even for instance retrieval, we 289
 use these three settings as our baselines to assess the impact of caching and, more 290
 importantly, the benefits of semantic-awareness in accelerating reasoning tasks. 291

The reasoners represent the state-of-the-art methods commonly used in the concept 292
 learning framework. In our evaluation, we selected the symbolic reasoners JFact [31], 293
 HermiT [11], Pellet [29], and its extended version Openllet [28]. Complementing these, 294

Table 2: Datasets statistics. $|\mathbf{NC}|$ is the total number of concepts generated on each dataset. $|\mathbf{Ind.}|$ is the total number of individuals and $|\mathbf{Prop.}|$ is the total number of properties (relations) in each dataset.

Dataset	$ \mathbf{Ind.} $	$ \mathbf{Classes} $	$ \mathbf{Prop.} $	$ \mathbf{TBox} $	$ \mathbf{ABox} $	$ \mathbf{NC} $
Vicodi	33,238	194	10	204	116,181	2,105
Carcinogenesis	22,372	142	4	144	74,223	1,684
Mutagenesis	14,145	86	5	82	47,722	1,572
Family	202	18	4	26	472	1,236
Father	6	4	1	3	4	90

we included an embedding-based reasoner (EBR) [30], to explore a neural reasoning paradigm.

The evaluation involves two stages. Initially, we measure the performance of the reasoners without using our caching mechanism. Following this, we integrate our caching system into the experiments, varying the cache sizes to examine its impact. The cache size is relative to the total number of concepts generated for each dataset; for example, a cache size of 10% implies that the cache can accommodate up to 10% of the number of generated concepts on the corresponding dataset (see Table 2 to see the number of generated concepts on each dataset).

5.1 Experimental Setup

Concepts generation Even though this algorithm can be easily extended to more expressive DL (e.g. \mathcal{ALCQ}), we restrict our experiments to generate concepts of length up to 5 in the \mathcal{ALC} DL (see Table 1 for syntax). The refinement operator described in [16] is used for this purpose. The total number of concepts generated for each dataset is summarized in Table 2.

Datasets We evaluate our caching system on five benchmark datasets: Father, Family, Carcinogenesis, Mutagenesis, and Vicodi [4]. The Father and Family datasets are relatively small (based on the number of individuals) and are commonly used for testing algorithms involving basic family and hierarchical relationships. To assess the scalability of our approach, we used larger and more complex datasets: Vicodi, Carcinogenesis, and Mutagenesis. These datasets are well-known in bioinformatics and cheminformatics, containing detailed information about chemical compounds and their biological effects, such as carcinogenicity and mutagenicity [17]. The key statistics of these datasets are presented in Table 2.

6 Results and Discussion

The results aim to demonstrate the impact of integrating our caching system on the performance of reasoners across various datasets. The evaluation focuses on two critical

metrics: runtime and hit ratio as functions of cache size. The results section is structured into three parts. First, we integrate our caching system with five distinct eviction strategies—FIFO, LIFO, LRU, MRU, and RP—on small datasets, highlighting the differences in reasoner speed with and without caching. This analysis illustrates how the choice of cache policy interacts with our caching algorithm. Next, we identify the best-performing eviction strategy and demonstrate its effectiveness in enhancing reasoner performance on larger datasets when combined with our caching algorithm. Finally, we show how the cache can be easily applied to speed up concept learning algorithms.

6.1 Application to Concept Retrieval

Runtime analysis In Figure 2, the gray dotted line represents the baseline performance of each reasoner without caching. Since we reran the reasoner independently for each cache size setting, the gray line is not perfectly flat but shows slight variations. Across all reasoners, the integration of caching significantly reduces the run time, particularly at larger cache sizes. Notably, the LRU eviction strategy consistently delivers the best performance in minimizing run time. For example, when the cache has enough space to contain 80% of the concepts, LRU enables dramatic run time reductions, outperforming other strategies. On the other hand, the performance of eviction strategies such as RP and MRU fluctuates, indicating less consistency across reasoners. LIFO also exhibits reasonable run time reductions but is outperformed by LRU in most cases.

Hit ratio analysis Figure 3 illustrates the hit ratio performance on the Family dataset (all the other datasets showed similar hit ratio behaviour). As expected, larger cache sizes lead to higher hit ratios across all eviction strategies. LRU emerges as the most efficient strategy, although all reasoners achieve near-perfect hit ratios at higher cache allocations, LRU already shows high hit ratios at lower cache sizes (20%, 40%). In contrast, FIFO and MRU exhibit slower increases in hit ratios, suggesting less efficient utilization of cache space.

These results clearly indicate that our caching system enhances the performance of all reasoners, with LRU consistently outperforming other eviction strategies. This is evident in both reduced run times and increased hit ratios. This confirms the findings of O’neil et al. [23] who proved the superiority of LRU among different replacement policies. The scalability of our caching algorithm is demonstrated effectively on small datasets, setting the stage for its evaluation on larger datasets.

Figure 1 presents the runtime performance of all reasoners on the Carcinogenesis and Mutagenesis using the LRU eviction strategy. The results are shown as a function of cache size, ranging from 10% to 100%. The dotted horizontal lines in each plot indicate the baseline runtime of the reasoners without caching. Across both datasets, the runtime consistently decreases as the cache size increases, highlighting the effectiveness of our caching mechanism in improving performance.

HermiT, the slowest reasoner without caching, benefits the most, with performance improvements reaching up to 60% at larger cache sizes. Similarly, EBR, while faster than HermiT, shows runtime reductions of approximately 30% to 50%, also at higher cache sizes. Pellet and Openllet show almost similar performance, with Openllet consistently maintaining its position as the fastest reasoner, both with and without caching.

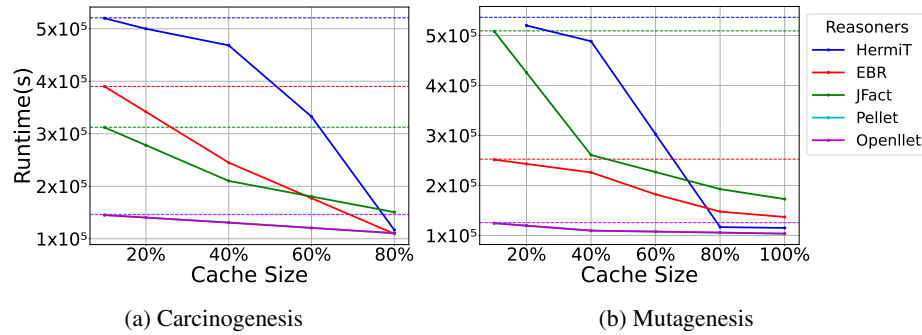


Fig. 1: Run time performance vs cache size for five reasoners on large datasets. The dotted horizontal lines represent the baseline performance of each reasoner without caching.

365 They achieve runtime reductions of approximately 45% – 50% at higher cache allocations.
 366 They achieve runtime reductions of approximately 45% – 50% at higher cache allocations.
 367 They achieve runtime reductions of approximately 45% – 50% at higher cache allocations. Notably,
 368 the Carcinogenesis dataset shows slightly greater performance gains for most reasoners
 369 compared to the Mutagenesis dataset, reflecting the higher reasoning complexity of
 370 Carcinogenesis and the cache’s ability to handle it effectively. For instance, the initial
 371 runtime of EBR was approximately 700,000 seconds, corresponding to more than eight
 372 days. However, with the integration of our cache, that runtime was drastically reduced
 373 to around 100,000 seconds, approximately equivalent to just a single day. The dotted
 374 baseline lines emphasize the advantage of integrating our cache, showcasing its ability
 375 to accelerate reasoning processes even for computationally demanding datasets.

376 6.2 Application to Class Expression Learning

377 To assess the effectiveness of our caching algorithm in CEL, we evaluated its impact on
 378 runtime performance across four datasets and several state-of-the-art CEL algorithms.
 379 Specifically, we included three top-down refinement-based learners—OCEL, CELOE,
 380 and CLIP—as well as EvoLearner, which follows a different strategy (see Section 2 for
 381 details). All learning problems and corresponding sets of positive and negative exam-
 382 ples were generated following the protocol introduced by [8], and are available in the
 383 supplementary material. In this experiment, the maximum storage capacity of the cache
 384 has been set to 1024 concepts.

385 The results presented in Figure 4 demonstrate the substantial impact our cache algo-
 386 rithm has on accelerating the concept learning process, particularly for top-down
 387 refinement-based learners such as OCEL, CELOE, and CLIP. Among these, OCEL
 388 benefits the most. For instance, on the Carcinogenesis dataset, the average runtime of
 389 OCEL drops dramatically from over 100 seconds without the cache to under 20 seconds
 390 with it. Similarly, on the Mutagenesis dataset, the runtime decreases from more than 25
 391 seconds to approximately 12 seconds. These improvements are consistent with OCEL’s

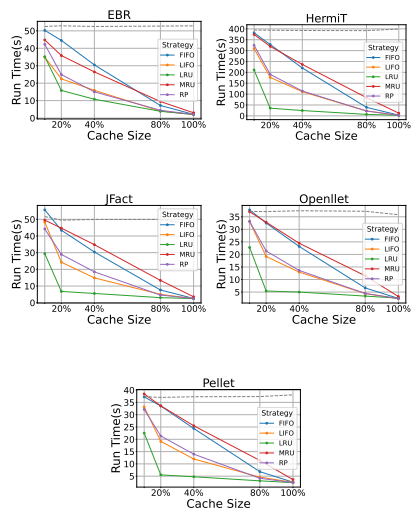


Fig. 2: Run time performance vs. cache size for each reasoner on the Family dataset. The gray dotted line indicates the runtime of the reasoner without our cache.

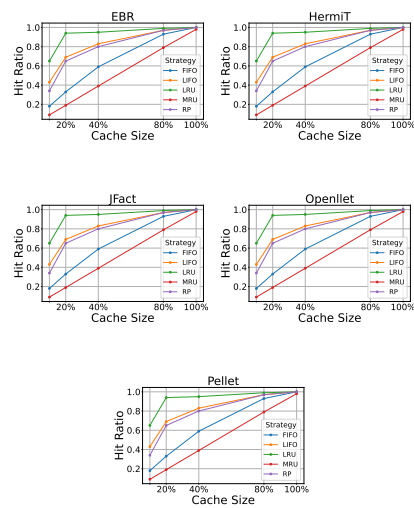


Fig. 3: Hit score performance vs. cache size for each reasoner on the Family dataset.

reliance on expensive instance retrieval operations, which our cache significantly optimizes by avoiding redundant reasoning through semantic reuse. 392

By contrast, the non-semantic cache shows little to no improvement for most learners and, in some cases, even leads to worse performance. This degradation stems from the cache being quickly filled with syntactically distinct but semantically irrelevant concepts, overwhelming the memory with uninformative entries. As a result, many cache lookups fail or return unhelpful data, causing additional overhead rather than reducing it. These findings highlight the importance of semantic awareness: without it, caching can hinder rather than help. 393 394 395 396 397 398 399 400

CELOE and CLIP also show measurable runtime reductions across datasets, albeit to a lesser extent. On the Vicodi dataset, CELOE’s average runtime drops from roughly 135 seconds to about 100 seconds. These learners, while still based on top-down refinement strategies, tend to produce fewer costly retrievals or shallower refinement chains, which may explain why the performance gains, though present, are not as dramatic as with OCEL. Nevertheless, the consistent improvements observed with CELOE and CLIP confirm that our caching approach generalizes well across multiple concept-learning systems that employ refinement operators. 401 402 403 404 405 406 407 408

However, EvoLearner shows no benefit from our caching mechanism. This result aligns with EvoLearner’s nature, which does not follow a top-down refinement strategy. Instead, EvoLearner begins the learning process by sampling a large population of candidate concepts and their associated individuals upfront. This sampling-based design inherently reduces its reliance on repeated instance retrieval during the search process. 409 410 411 412 413

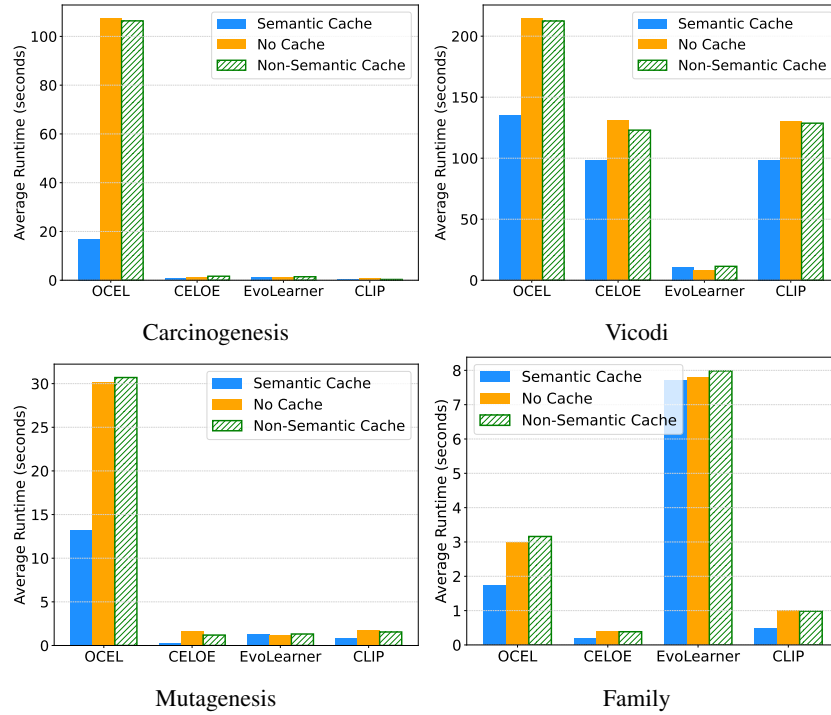


Fig. 4: Average runtime of concept learners using no cache, the semantic cache, and the non-semantic cache baseline across different datasets.

414 Furthermore, the large initial population leads to a wide set of unique candidate concepts
 415 whose instance sets must be computed and stored early on. Attempting to cache
 416 all this information is not only memory-intensive but also largely inefficient, as the like-
 417 lihood of reusing these exact expressions is low in an evolutionary setting. As a result,
 418 our cache provides limited benefit in EvoLearner’s workflow, explaining the difference
 419 in runtime between the cached and non-cached variants.

420 7 Conclusion

421 In this work, we proposed a simple yet effective caching algorithm to optimize the re-
 422 trieval process, which is the most computationally intensive task in concept learning. By
 423 integrating our caching mechanism with various cache eviction strategies, we demon-
 424 strated its ability to enhance reasoners’ performance significantly. Our results showed
 425 that the LRU eviction strategy works best with our cache, improving reasoning speed by
 426 at least 20% and up to 60% in some cases. This work highlights the importance of lever-
 427 aging semantic caching to alleviate computational bottlenecks in ontology-based sys-
 428 tems, providing a powerful yet straightforward solution for accelerating concept learn-
 429 ing workflows.

References

1. Ari, I., Amer, A., Gramacy, R.B., Miller, E.L., Brandt, S.A., Long, D.D.: Acme: Adaptive caching using multiple experts. In: WDAS. vol. 2, pp. 143–158 (2002)
2. Baader, F., Horrocks, I., Sattler, U.: Description logics. In: Handbook on ontologies, pp. 21–43. Springer (2009)
3. Belady, L.A.: A study of replacement algorithms for a virtual-storage computer. IBM Systems journal **5**(2), 78–101 (1966)
4. Bin, S., Bühmann, L., Lehmann, J., Ngonga Ngomo, A.C.: Towards sparql-based induction for large-scale rdf data sets. In: ECAI 2016, pp. 1551–1552. IOS Press (2016)
5. Brachman, R.J.: A structural paradigm for representing knowledge. Bolt, Beranek and Newman (1978)
6. Bühmann, L., Lehmann, J., Westphal, P.: Dl-learner—a framework for inductive learning on the semantic web. Journal of Web Semantics **39**, 15–24 (2016)
7. Ceylan, I.I., Penaloza, R.: The bayesian ontology language bel. Journal of Automated Reasoning **58**(1), 67–95 (2017)
8. Demir, C., Ngomo, A.C.N.: Neuro-symbolic class expression learning. In: IJCAI. pp. 3624–3632 (2023)
9. Fanizzi, N., d’Amato, C., Esposito, F.: Dl-foil concept learning in description logics. In: Inductive Logic Programming: 18th International Conference, ILP 2008 Prague, Czech Republic, September 10–12, 2008 Proceedings 18. pp. 107–121. Springer (2008)
10. Fanizzi, N., Rizzo, G., d’Amato, C., Esposito, F.: Dlfoil: Class expression learning revisited. In: Knowledge Engineering and Knowledge Management: 21st International Conference, EKAW 2018, Nancy, France, November 12–16, 2018, Proceedings 21. pp. 98–113. Springer (2018)
11. Glimm, B., Horrocks, I., Motik, B., Stoilos, G., Wang, Z.: Hermit: an owl 2 reasoner. Journal of automated reasoning **53**, 245–269 (2014)
12. Heindorf, S., Blübaum, L., Düsterhus, N., Werner, T., Golani, V.N., Demir, C., Ngonga Ngomo, A.C.: Evolearner: Learning description logics with evolutionary algorithms. In: Proceedings of the ACM Web Conference 2022. pp. 818–828 (2022)
13. Heist, N., Paulheim, H.: Uncovering the semantics of wikipedia categories. In: The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part I 18. pp. 219–236. Springer (2019)
14. Karalis, N., Bigerl, A., Demir, C., Heidrich, L., Ngomo, A.N.: Evaluating negation with multi-way joins accelerates class expression learning. In: Bifet, A., Davis, J., Krilavicius, T., Kull, M., Ntoutsi, E., Zliobaite, I. (eds.) Machine Learning and Knowledge Discovery in Databases. Research Track - European Conference, ECML PKDD 2024, Vilnius, Lithuania, September 9–13, 2024, Proceedings, Part VI. Lecture Notes in Computer Science, vol. 14946, pp. 199–216. Springer (2024). https://doi.org/10.1007/978-3-031-70365-2_12, https://doi.org/10.1007/978-3-031-70365-2_12
15. Kazakov, Y., Krötzsch, M., Simančík, F.: The incredible elk: From polynomial procedures to efficient reasoning with *el* ontologies. Journal of automated reasoning **53**(1), 1–61 (2014)
16. Kouagou, N.J., Heindorf, S., Demir, C., Ngonga Ngomo, A.C.: Neural class expression synthesis. In: European Semantic Web Conference. pp. 209–226. Springer (2023)
17. Kouagou, N.J., Heindorf, S., Demir, C., Ngonga Ngomo, A.C.: Neural class expression synthesis in alchIQ (d). In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 196–212. Springer (2023)
18. Kulmanov, M., Khan, M.A., Hoehndorf, R.: Deepgo: predicting protein functions from sequence and interactions using a deep ontology-aware classifier. Bioinformatics **34**(4), 660–668 (2018)

- 479 19. Lehmann, J., Auer, S., Bühmann, L., Tramp, S.: Class expression learning for ontology en-
480 gineering. *Journal of Web Semantics* **9**(1), 71–81 (2011)
- 481 20. Lehmann, J., Hitzler, P.: Concept learning in description logics using refinement operators.
482 *Machine Learning* **78**, 203–250 (2010)
- 483 21. Megiddo, N., Modha, D.S.: Outperforming lru with an adaptive replacement cache algo-
484 rithm. *Computer* **37**(4), 58–65 (2004)
- 485 22. Nardi, D., Brachman, R.J., et al.: An introduction to description logics. *Description logic*
486 *handbook* **1**, 40 (2003)
- 487 23. O’neil, E.J., O’Neil, P.E., Weikum, G.: An optimality proof of the lru-k page replacement
488 algorithm. *Journal of the ACM (JACM)* **46**(1), 92–112 (1999)
- 489 24. Oramas, S., Ostuni, V.C., Noia, T.D., Serra, X., Sciascio, E.D.: Sound and music recommen-
490 dation with knowledge graphs. *ACM Transactions on Intelligent Systems and Technology*
491 *(TIST)* **8**(2), 1–21 (2016)
- 492 25. Rizzo, G., Fanizzi, N., d’Amato, C.: Class expression induction as concept space exploration:
493 From dl-foil to dl-focl. *Future Generation Computer Systems* **108**, 256–272 (2020)
- 494 26. Sarker, M.K., Hitzler, P.: Efficient concept induction for description logics. In: *Proceedings*
495 *of the AAAI Conference on Artificial Intelligence*. vol. 33, pp. 3036–3043 (2019)
- 496 27. Sethumurugan, S., Yin, J., Sartori, J.: Designing a cost-effective cache replacement policy us-
497 ing machine learning. In: *2021 IEEE International Symposium on High-Performance Com-*
498 *puter Architecture (HPCA)*. pp. 291–303. IEEE (2021)
- 499 28. Singh, G., Bhatia, S., Mutharaju, R.: Owl2bench: a benchmark for owl 2 reasoners. In: *The*
500 *Semantic Web–ISWC 2020: 19th International Semantic Web Conference, Athens, Greece,*
501 *November 2–6, 2020, Proceedings, Part II* 19. pp. 81–96. Springer (2020)
- 502 29. Sirin, E., Parsia, B., Grau, B.C., Kalyanpur, A., Katz, Y.: Pellet: A practical owl-dl reasoner.
503 *Journal of Web Semantics* **5**(2), 51–53 (2007)
- 504 30. Teyou, L.M.K., Friedrichs, L., Kouagou, N., Demir, C., Mahmood, Y., Heindorf, S.,
505 Ngomo, A.C.N.: Neural reasoning for robust instance retrieval in shoiq. *arXiv preprint*
506 *arXiv:2510.20457* (2025)
- 507 31. Tsarkov, D., Horrocks, I.: Fact++ description logic reasoner: System description. In: *Inter-*
508 *national joint conference on automated reasoning*. pp. 292–297. Springer (2006)
- 509 32. Wang, M., Qiu, L., Wang, X.: A survey on knowledge graph embeddings for link predic-
510 tion. *Symmetry* **13**(3), 485 (2021). <https://doi.org/10.3390/SYM13030485>, [https://doi.org/](https://doi.org/10.3390/sym13030485)
511 [10.3390/sym13030485](https://doi.org/10.3390/sym13030485)
- 512 33. Zhou, Y., Wang, F., Shi, Z., Feng, D.: An end-to-end automatic cache replacement policy
513 using deep reinforcement learning. In: *Proceedings of the International Conference on Au-*
514 *tomated Planning and Scheduling*. vol. 32, pp. 537–545 (2022)