

# Syndication on the Web using Incremental Description Logic Reasoning

Christian Halaschek-Wiener

University of Maryland, College Park, MD, USA

Department of Computer Science

halasche@cs.umd.edu

## Abstract

In this work, we present a syndication framework in which Description Logic (DL) reasoning is used for matching subscription requests with newly published information. One of the main limitations of previous efforts is related to the poor scalability of DL reasoners, especially when handling updates. This has made a DL based approach impractical for syndication of time sensitive information. To overcome this, the presented framework leverages recent work on incremental reasoning through changes of ontologies, thereby making such a system more appropriate for such applications.

## 1 Introduction

Over the past few years there has been substantial interest in syndication on the Web; this is evident by the wide spread use of syndication standards such as RSS, Atom, etc. There has also been research on efficiently filtering published contents based on subscription requests and preferences. Until recently this has focused on the utilization of XML as the primary representation format for syndication.

With the emergence of the Web Ontology Language (OWL), there has been recent interest in using Description Logics (DLs) to represent the contents of feeds. This in turn provides increased expressivity over an XML representation, potentially providing more accurate matching of subscription requests with published documents. The main limitation with recent approaches [Haarslev and Möller, 2003; Uschold *et al.*, 2003] has been related to the complexities and cost/overhead of DL reasoning, in particular with reasoning through updates. Until recently, DL reasoning algorithms have been static in nature. That is after an update reasoning on the updated knowledge base (KB) must be performed from scratch. It is clear that in a syndication context this is a substantial limitation as new knowledge is continuously published. Further, in domains such as the syndication of financial news feeds, response times must be minimal as critical information must be delivered in near real time (e.g., for stock trading purposes).

In this paper, we present a framework for syndication on the Web using ontologies expressed in the Description

Logic *SHI* (which corresponds to a large portion of OWL-DL) and conjunctive ABox queries for representing published information and subscription requests respectively; additionally, DL reasoning services are utilized for publication/subscription matching. Our framework mainly differs from recent approaches [Haarslev and Möller, 2003; Uschold *et al.*, 2003] in that we leverage our recent work incremental DL reasoning through updates. This provides a syndication framework positioned to meet the time critical demands of various use cases which had been previously unachievable using a DL based approach.

## 2 Syndication Framework

In this section we define a syndication architecture in which brokers deliver published information to the appropriate subscription requests. The contents of published documents/information is treated as ABox facts, denoted  $P_f$ , and subscribers define their interests via subscription requests which are represented as conjunctive ABox queries, denoted  $S_q$ . Additionally, brokers contain domain knowledge represented as both ABox and TBox assertions, denoted  $B_{KB}$ ; for example, in the financial domain this knowledge would model concepts and relations related to stocks, companies, etc., as well as contain ABox facts about specific companies. Newly published information,  $P_f$ , is then pushed to a subscriber with interests  $S_q$  if  $P_f \cup B_{KB} \models S_q$ . We note that subscription queries can be viewed as continuous queries; that is they are issued once over a changing KB and their results sets are incrementally updated as the underlying KB is updated (by newly published data).

This framework is clearly different from previous approaches which have modeled published data as DL concepts [Uschold *et al.*, 2003]. We note here that our approach is similar to the approach described in [Haarslev and Möller, 2003]. However, [Haarslev and Möller, 2003] suffers from performance issues due to the previously discussed limitation of generally having to reason from scratch when new information is published.

## 3 Practical Reasoning for Syndication

In order to disseminate newly published information, we must reason through updates. This is evident as newly published information must be integrated into the brokers local KB so

that subscriptions can be matched. For the purpose of integrating new information into the brokers KB we adopt syntactic changes/updates of KB assertions, which we refer to as *Syntactic Updates*. Intuitively *Syntactic Updates* can be described as an update in which all new assertions are directly added (or removed) to the asserted (base) axioms. Formally, we describe this update as follows:

**Definition 1** (*Syntactic Updates*) Let  $S$  be the set of assertions in an initial knowledge base  $K$ . Then under Syntactic Updates, updating  $K$  with an addition (resp. deletion)  $\alpha$ , written as  $K + \alpha$  (resp.  $K - \alpha$ ), results in an updated set of axioms  $S'$  such that  $S' = S \cup \{\alpha\}$  (resp.  $S' = S \setminus \{\alpha\}$ ).

This type of update semantics is adopted as we feel it is fitting for the task syndication on the Web.

### Incremental Consistency Checking

After a brokers KB is updated, consistency must be rechecked; if the KB becomes inconsistent then everything is trivially entailed. Due to the overhead of consistency checking of large KBs, it is clear that when disseminating time sensitive information we cannot simply recheck consistency from scratch. For this purpose, we leverage our recent work on incremental consistency checking of the Description Logic *SHIQ* [Halaschek-Wiener *et al.*, 2006a]. In [Halaschek-Wiener *et al.*, 2006a] an approach is presented for incrementally updating tableau completion graphs under ABox updates; this provides a mechanism to incrementally determine if the updated KB is consistent. The approach has shown orders of magnitude performance improvement and performs in real time for a variety of realistic ontologies. This is demonstrated for varying sized datasets and updates from the LUBM benchmark (1, 2 and 3 Universities) in Figure 1 (incremental consistency checking is denoted by 'Opt', where as checking consistency from scratch is denoted by 'Reg'). Fur-

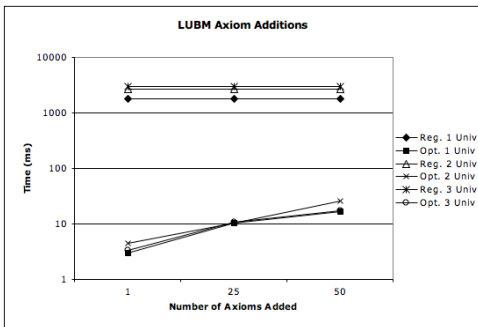


Figure 1: Addition Updates of LUBM datasets

ther details are omitted here, however they can be found in [Halaschek-Wiener *et al.*, 2006a].

### Continuous Conjunctive Query Answering

Once consistency is guaranteed, knowledge brokers must match newly published information with registered subscriptions (i.e., queries). In order for a DL-based approach to be practical, registered queries cannot simply be re-evaluated from scratch. In [Halaschek-Wiener *et al.*, 2006b], we have formally defined continuous conjunctive queries over OWL-DL knowledge bases; additionally we investigated various

techniques for continuous query answering focused on reducing candidate variable bindings when the knowledge base is updated. For purpose of this work, we adopt such an approach for matching subscription requests, as it demonstrates orders of magnitude performance improvement over simply re-evaluating queries from scratch. This is demonstrated in Figure 2, which compares the continuous query answering approach (denoted 'Pellet-C') with evaluating queries from scratch (denoted 'Pellet'). Figure 2 also compares the approach with the DL reasoner KAON2, which reduces reasoning to disjunctive datalog and provides some incremental query answering support. Further details are omitted here, however

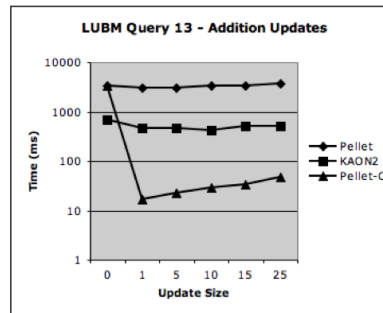


Figure 2: Continuous Query Answering for LUBM Query 13 they can be found in [Halaschek-Wiener *et al.*, 2006b].

## 4 Conclusion and Future Work

In this paper, we have presented a DL-based syndication framework which leverages recent work on incremental reasoning. This provides a practical DL-based approach for syndication of time sensitive information. Future work includes expanding the framework to support more expressive Description Logics, providing additional incremental reasoning techniques and providing further evaluations of the approach.

## References

- [Haarslev and Möller, 2003] V. Haarslev and R. Möller. Incremental query answering for implementing document retrieval services. In *Proc. of the Int. Workshop on Description Logics*, pages 85–94, 2003.
- [Halaschek-Wiener *et al.*, 2006a] C. Halaschek-Wiener, B. Parsia, and E. Sirin. Description logic reasoning with syntactic updates. In *In Proc. of the 5th Int. Conference on Ontologies, DataBases, and Applications of Semantics*, 2006.
- [Halaschek-Wiener *et al.*, 2006b] C. Halaschek-Wiener, B. Parsia, and E. Sirin. Towards continuous query answering on the semantic web. In *UMIACS Technical Report*, 2006. <http://www.mindswap.org/papers/2006/ContQueryTR2006.pdf>.
- [Uschold *et al.*, 2003] M. Uschold, P. Clark, F. Dickey, C. Fung, S. Smith, S. Uzcakaj, M. Wilke, S. Bechhofer, and I. Horrocks. A semantic infosphere. In *Proc. of the Int. Semantic Web Conference*, 2003.