

Review of Paper:

Towards Expressive Syndication on the Web
- *Christian Halaschek-Wiener, James Hendler*

March-03-2009

by Ankesh

Outline

	Score	
RELEVANCE	7	RELEVANT TO VERY SPECIALIZED SUBAREA OF SEMANTIC WEB
SIGNIFICANCE	8	SIGNIFICANT
NOVELTY	9	NOVEL
QUALITY OF EVALUATION	10	VERY CONVINCING
TECHNICAL SOUNDNESS	9	TECHNICALLY SOUND
CLARITY	10	EXCELLENT
PRESENTATION	-2-	THIS PAPER COULD BE PRESENTED REASONABLY WELL ORALLY OR IN POSTER SESSION
OVERALL SCORE	9	GOOD PAPER, ACCEPT
CONFIDENCE SCORE	4	THE REVIEWER IS FAIRLY CONFIDENT, POSSIBLE THAT HE NDID NOT UNDERSTAND CERTAIN PARTS

Relevance (7)

- Domain of research- adding semantics to the syndication process
 - To aid filtering and automated reasoning for discovering implicit subscription matches
- Sound extension of tableau based DL reasoning algorithms for syndication-kind of Abox updates (addition/ deletion)- well supported by proofs.

Syndication Framework

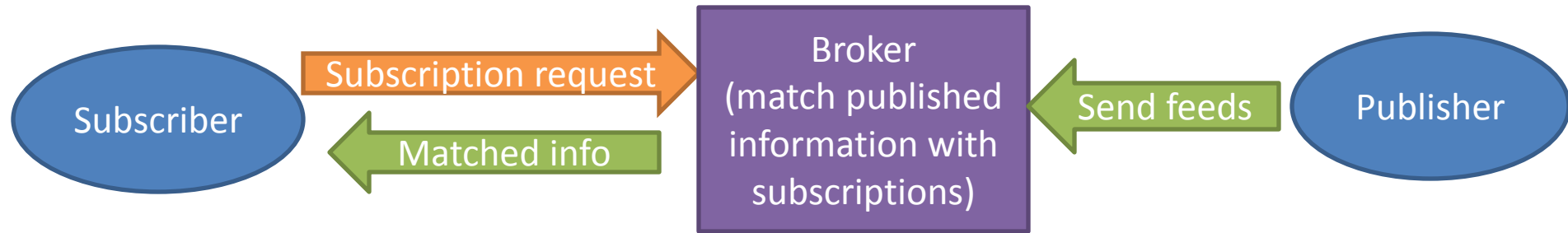


Fig1. Syndication Framework

```
:RiskyCompany a owl:Class;  
  owl:intersectionOf (  
    [ a owl:Restriction; owl:onProperty :onRecommendation;  
      owl:someValuesFrom :SellList ]  
    [ a owl:Restriction; owl:onProperty :downgradedBy;  
      owl:someValuesFrom [ owl:oneOf ( :SandP :Moody's ) ] ]  
    :Company  
  ).  
:movedToJunk a owl:ObjectProperty;  
  rdfs:subPropertyOf :downgradedBy .  
:onRecommendation a owl:ObjectProperty;  
  owl:InverseOf :hasRecommendation .
```

Fig2. Schema provided by Broker

1. RiskyCompany(x)
2. onRecommendation(x, y) \wedge Company(x) \wedge SellList(y).

Fig 3. Sample Subscriptions

Significance (8)

1. **Formally defines continuous queries** for OWL KBs & gives **novel algorithm** for continuous query answering for OWL KBs that are being updated
 - a) Primarily focusing on reducing the size of the KB that must be considered as candidate query bindings
2. **Formalizes** DL-based syndication framework
3. **Publication matching** perspective to Subscription matching
 - a) Attaching information matches (individuals bound to variables) with minimal sets of publications that caused this match to occur
 - b) Based on minimal justifications for an entailment in DLs ($K' \subseteq K$ is a min. justification for α if $K' \models \alpha$ and $K'' \not\models \alpha$ for every $K'' \subset K'$)

Formalization of Syndication Framework

- Subscription - $\langle Q_c, t \rangle$
 - Q_c : Subscribers interests are represented in form of conjunctive Abox Query ($\text{onRecommendation}(x, y) \wedge \text{Company}(x) \wedge \text{SellList}(y)$)
 - t : Number of time units subscription is valid (expiry time)
- Subscriber - $\langle s, i \rangle$, subscription, identifier
- Publication - $\langle \alpha, t, i \rangle$
 - α : set of assertions
 - t : time units publication is valid
 - i : identifier of publisher
- Syndication Broker – $\langle S, P, K_l \rangle$
 - S : set of subscribers
 - P : set of publishers
 - Local DL KB
- **Match** – a match for subscription can be a composition of information from multiple publications.

Novelty (9)

- **Syntactic Updates** (different from belief revision, for example)
 - Add all new assertions to asserted axioms s.t. it results in a consistent KB, o/w reject newly published information
- **Incremental consistency checking**: add new [edges, nodes, labels] introduced by update to completion graph from consistency check prior to the update, then fire standard tableau completion rules
 - Observation : Updates did not have large effect on existing completion graph
- **Continuous Conjunctive Abox Query**, Q_c : defined w.r.t K_t s.t. it produces results at time t $Q_c(t) = \{ \langle a_1, \dots, a_n \rangle \in \mathbf{I}_{K_t}^n \mid K_t \models Q_c[x_1/a_1, \dots, x_n/a_n] \}$
- **Continuous query answering** – next slide

Continuous Query Answering: Estimate affected individuals

- Guarantee consistency of updated KB, reevaluate the subscriptions:
 - Additions: Determine any new bindings that are entailed by the KB
 - Deletions: Determine that previous bindings are still entailed.
- **Explicitly Affected Individuals (EAI)** - intuitively are the individuals manipulated during the incremental update of all completions for a KB.
- **Affected Individuals** – $EAI + \{a \mid a \text{ is at distance } \leq \text{length of query, from some } b \in EAI\}$. (Remember: $\text{onRecommendation}(x, y) \wedge \text{Company}(x) \wedge \text{SellList}(y)$).
- S_G , **Summary Root Graph** of G – The completion graph built for SHI KB K by applying all tableau expansion rules with following modification:
 - when nondeterministic choice is encountered, add all labels in the disjunction to the node without creating new branch
 - If a clash is encountered ignore it
- S_G is used to localize an overestimate of the explicitly affected individuals.

Continuous Query Answering: Set of Candidate Bindings

Conjunctive query can be answered by **syntactically mapping the query into all completion graphs** for the KB.

If $Q \hookrightarrow G$ query can be mapped to all completions of the KB then KB satisfies query. (now, take structure of the query into account!)

Set of candidate bindings for distinguished variable x:

$$A_{QI}(x) = \{a \mid a \in AI_O(K, \alpha) \wedge Q \hookrightarrow_{\{x \leftarrow a\}} G\} \cup \{a \mid b \in AI_O(K, \alpha) \wedge Q \hookrightarrow_{\{x \leftarrow a, y \leftarrow b\}} G\}$$

- Additions: each individual in $AI_O(K, \alpha)$ is iteratively substituted into variables in the query; inspect neighbor nodes in the updated completion graph and see if they can be mapped into the remaining nodes (via roles whose labels match the query graph) in the query graph. (Remember: $Company(x) \wedge onRecommendation(x, y) \wedge SellList(y)$)
- If mapping does not exist- do not consider this mapping. because we have just found a completion graph (i.e., model) in which the query cannot be mapped.
- If a named individual can be mapped into a distinguished variable then consider this individual as a candidate binding.

Algorithm 1 *update_query_results*(K, S_G, Q_c, R, α)

Input:

K : initial KB
 S_G : summary root graph for K
 Q_c : continuous conjunctive query
 R : set of all current bindings (answer set)
 α : ABox update

Output:

K : updated KB
 S_G : updated summary root graph
 R : updated bindings (answer set)

```
1:  $K \leftarrow K \oplus \alpha$ 
2: if  $K$  is not consistent then
3:    $K \leftarrow$  Retract  $\alpha$  from  $K$ 
4:   return  $K, S_G, R$ 
5: end if
6:  $AI_O(K, \alpha) \leftarrow localize\_effects(S_G, \alpha)$ 
7:  $QI_S \leftarrow query\_impact(AI_O(K, \alpha), Q)$ 
8: if  $\alpha$  is an addition then
9:   for all  $a_1 \in A_{QI}(x_1), \dots, a_n \in A_{QI}(x_n)$  s.t.  $x \in DV(Q_c)$  do
10:    if  $K \models Q_c[x_1/a_1, \dots, x_n/a_n]$  then
11:       $R \leftarrow R \cup \{ \langle a_1, \dots, a_n \rangle \}$ 
12:    end if
13:  end for
14: else if  $\alpha$  is a deletion then
15:   for all  $\langle a_1, \dots, a_n \rangle \in R$  do
16:    if  $QI_S \cap \{a_1, \dots, a_n\} = \emptyset$  then
17:      continue
18:    else if  $K \not\models Q_c[x_1/a_1, \dots, x_n/a_n]$  then
19:       $R \leftarrow R \setminus \{ \langle a_1, \dots, a_n \rangle \}$ 
20:    end if
21:  end for
22: end if
23: return  $K, S_G, R$ 
```

Quality of Evaluation (10)

- Basic functionality of framework described in paper
 - Published information remains infinitely valid
 - Information match
- Data set- LUBM + queries 1,3, 13
 - Update sizes – 1, 5, 10, 15, 25

$\langle x \rangle \leftarrow \text{GraduateStudent}(x) \wedge \text{takesCourse}(x, \text{GraduateCourse}0)$

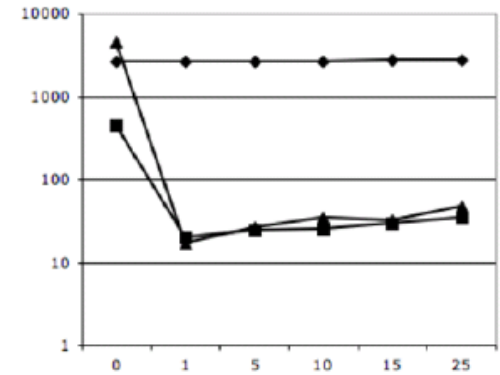
$\langle x \rangle \leftarrow \text{Publication}(x) \wedge \text{publicationAuthor}(x, \text{AssistantProfessor}0)$

$\langle x \rangle \leftarrow \text{Person}(x) \wedge \text{hasAlumnus}(\text{University}0, x)$

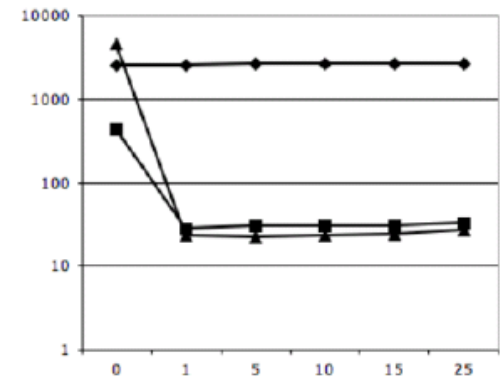
- Comparison with Pellet and KAON2
 - KAON2 optimized for query answering
 - Functionality to add & remove assertions

Empirical results: Relevant Statistics

- Overhead- summary root graph (SRG) generation
 - Time to build SRG – 2.7 secs.
 - Time to build SRG for original version- 26.1 secs
- 1 to 3 orders of improvements to Pellet
- Incremental consistency checking times 7 ms compared with 2200 ms in normal case
- Query answering time- 500 to 1000 ms compared with 33 ms with pruned candidate variable bindings
- **Number of affected individuals were proportional to number of individuals referenced in the update.**
- In update of size 1 it took 0.23 ms to localize the affected individuals.



(a) Query 1 Additions



(b) Query 1 Deletions

Technical Soundness (9)

- Concepts are correct and accurate.
- Well supported experimental results
- However, proofs for the propositions are missing (and no pointers as well):

PROPOSITION 2. Given a SHI KB \mathcal{K} , ABox update α and summary root graph S_G for \mathcal{K} , then the approach for finding $EI_{S_G}(\mathcal{K}, \alpha)$ is terminating and $EI(\mathcal{K}, \alpha) \subseteq EI_{S_G}(\mathcal{K}, \alpha)$.

PROPOSITION 3. Algorithm 1 is sound, complete, and terminating.

Clarity (10)

- Paper is clearly written.
- It is well-organized too.

- A (small) suggestion:

In section 4.3.1 ‘Localizing Effects of Updates’ authors provide the definitions of ‘Explicitly Affected Individuals’ (12), ‘Affected Individuals’ (14), ‘Overestimate of Explicitly Affected Individuals’ (16) and ‘Overestimate of Affected Individuals’ (17). Definitions 16 and 17 are used in the algorithm to localize effects. However, inclusion of definitions 12 and 14 in the manuscript is not supported well enough. Probably, 12 and 14 build up the ideas on normal completion graphs (components) which are applied to Summary Root Graph in 16 and 17.

CONCLUSION

- OVERALL SCORE: 9
- CONFIDENCE SCORE: 4
- PRESENTATION: paper could be presented reasonably well **orally** or in poster session
- Why OWA in such a setting?
 - Is absence of warning for some company, something positive?

Questions

- Can subscriber contribute Tbox?
- What happens after deletions- is the subscriber informed of the bindings that are not valid?