Towards Smart Cache Management for Ontology Based, History-Aware Stream Reasoning

Rui Yan, Deborah L. McGuinness
Tetherless World Constellation
Department of Computer Science
Rensselaer Polytechnic Institute

Brenda Praggastis, William P. Smith
Pacific Northwest National Laboratory,
Richland, WA, USA

Presented on Oct 12, 2015 at Linked Science Workshop, International Semantic Web Conference (ISWC) 2015

Contents

1. Introduction
   a. stream reasoning
   b. examples of the existing stream reasoning systems
2. Approach
   a. motivated use case
   b. why cache
   c. cache v.s. window
   d. historical data management
3. Discussion
4. Future work

Introduction / stream reasoning

- RDF streams [1]
  - streaming data modeled in RDF
  - linked data principles
- Data stream processing systems
- Semantic reasoning
- Stream reasoning [2]

[1] Barbieri, Davide F., and E. D. Valle. "A proposal for publishing data streams as linked data." *Linked Data on the Web Workshop*. 2010.
[2] Della Valle, Emanuele, et al. *A first step towards stream reasoning*. Springer Berlin Heidelberg, 2009.

Introduction / examples of the existing systems

- Existing stream reasoning systems
  - C-SPARQL [3]
    - continuous SPARQL, an extension of the standard SPARQL
    - window-based system
    - RDF data are stamped with timepoints
    - process RDF streams
  - EP-SPARQL [4]
    - event processing SPARQL, an extension of the standard SPARQL
    - window-based system
    - RDF data are stamped with time intervals
    - detect complex event patterns

[3] Barbieri, Davide Francesco, et al. "C-SPARQL: SPARQL for continuous querying." *Proceedings of the 18th international conference on World wide web*. ACM, 2009.

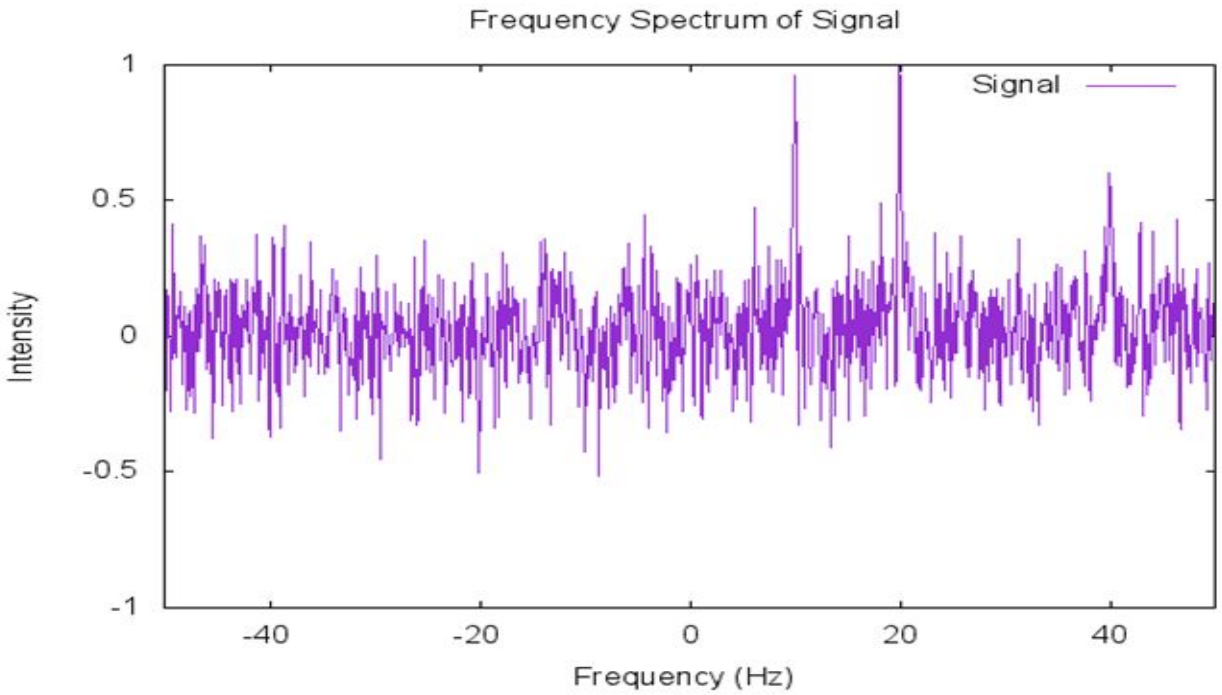
[4] Anicic, Darko, et al. "EP-SPARQL: a unified language for event processing and stream reasoning." *Proceedings of the 20th international conference on World wide web*. ACM, 20

Approach / motivated use case

Motivated Use Case:

- Nuclear Magnetic Resonance (NMR)

What is Nuclear Magnetic Resonance?



Approach / background ontology

Background ontology

- 30 different compounds are encoded with their unique frequency ranges
  - these compounds are sourced from Human Metabolome Database[1]
  - all metabolites (small molecules) that are found in human urine and/or blood plasma

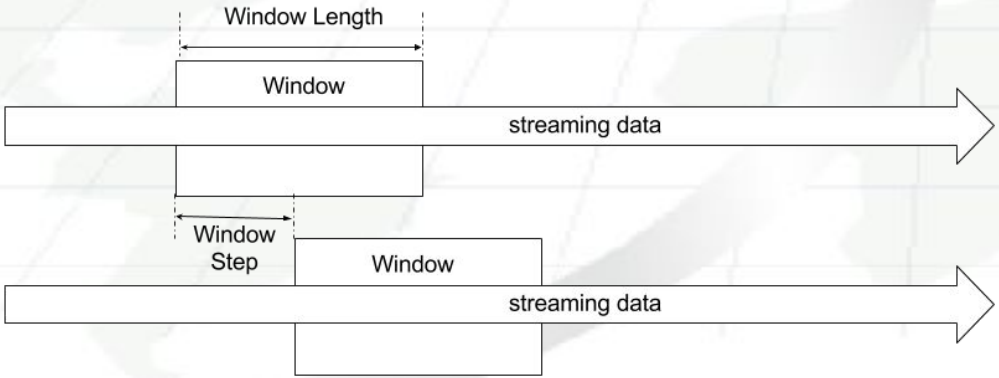
1. http://www.hmdb.ca/

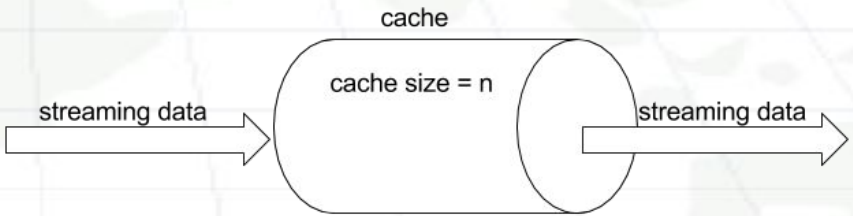
Approach / introducing the cache

What & Why cache ?

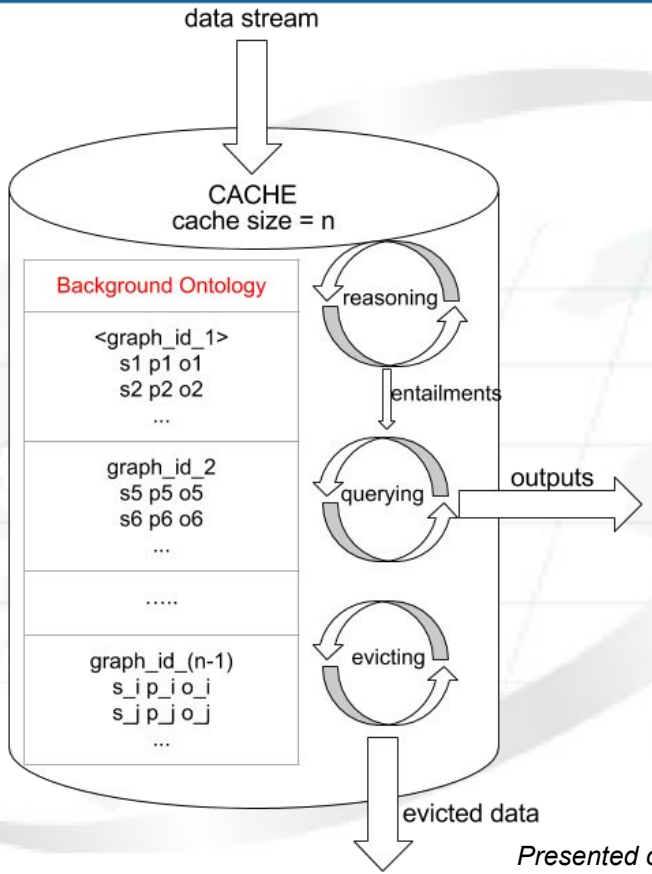
- memory-based or disk-based
- identify & store interesting portion of the streaming data
- cache management policy
- historical data management

a cache v.s. a window:





Approach / cache-enabling stream reasoning system architecture

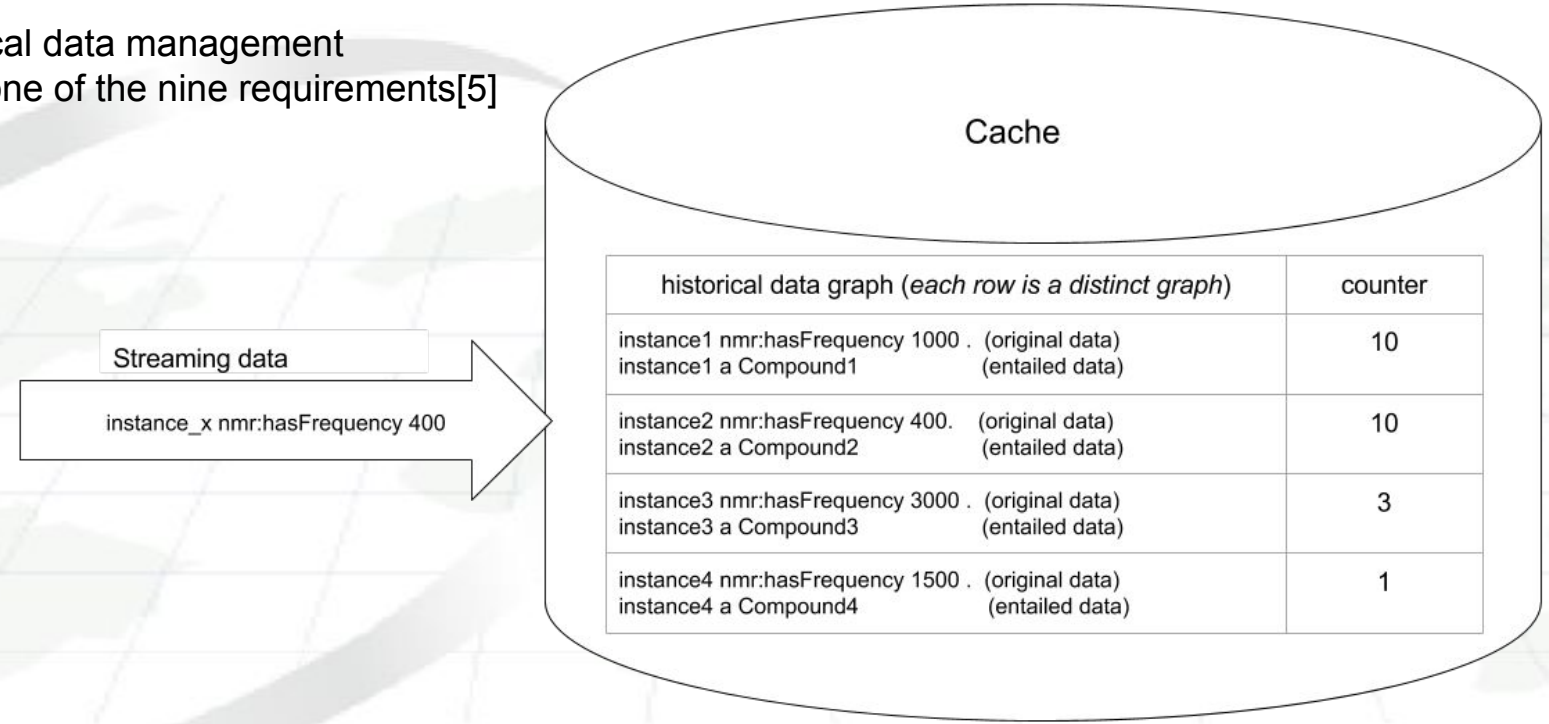


- cache size is limited
- background ontology is preloaded
- size can be in terms of triples/graph numbers
- reasoning and querying is constantly executed

- historical data: original data and entailments
- cache manages historical data with cache eviction policy

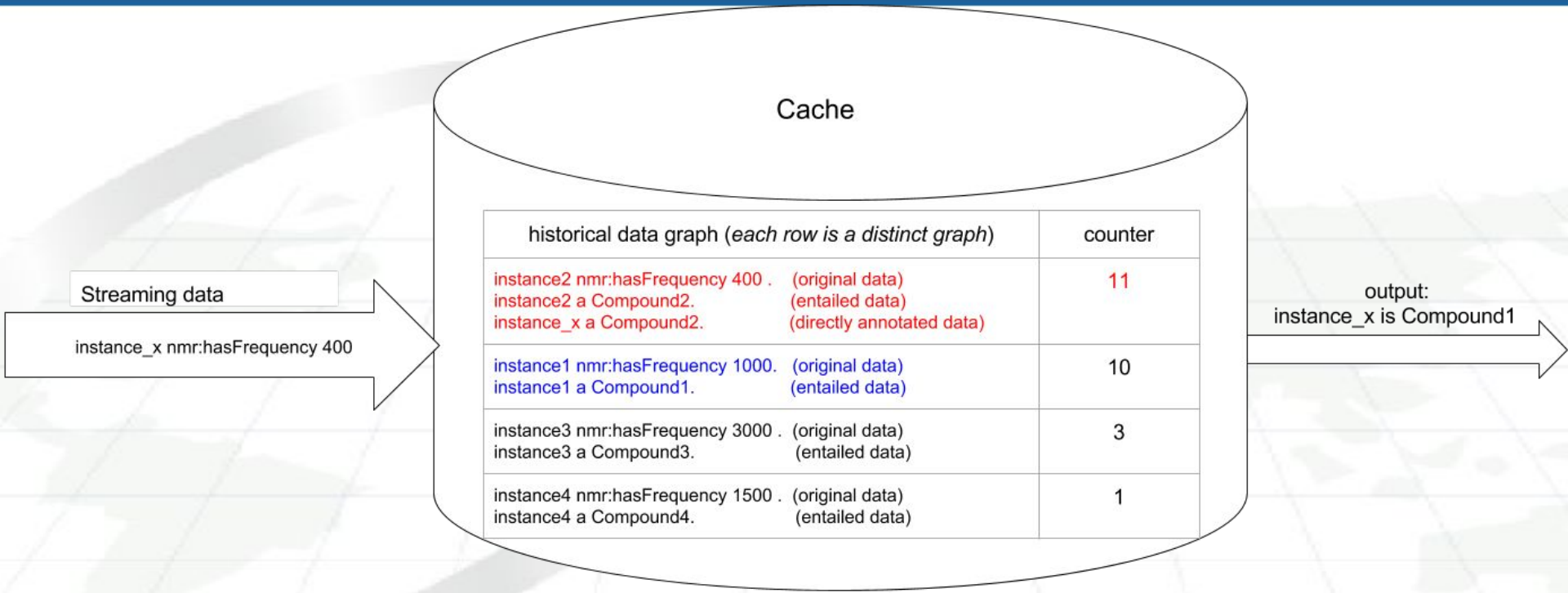
Approach / historical data management step 1

- historical data management
  - one of the nine requirements[5]



[5] Margara, Alessandro, et al. "Streaming the web: Reasoning over dynamic data." *Web Semantics: Science, Services and Agents on the World Wide Web* 25 (2014): 24-44.

Approach / historical data management step 2



Discussion

- scenarios where historical data are needed
  - anomaly detection
  - trend identification
  - historical data provides extra background
- multithreading can be leveraged
  - split different tasks to different threads make the system respond fast
  - but need to collaborate well: no eviction before query
  - easy to realize continuous querying with a thread
    - reduced the overhead of learning and applying other continuous sparql (like C-SPARQL, which has a different execution model and extra syntax)
- benefits of the semantics
  - background ontology
  - historical data management

Future Work & Next Steps

- explore different cache eviction policies' performances and effects on the system, such as least frequently used, least recently used, first in first out etc.
- the effects that expressiveness of the background ontology has on the system in terms of reasoning, querying and evicting.
- evaluation methods to benchmark the system

Acknowledgements

- The research described in the paper is part of the AIM Initiative at PNNL. It was conducted under the Laboratory Directed Research and Development (LDRD) program at PNNL, a multiprogram national laboratory operated by Battelle for the U.S. Department of Energy under contract DE-AC06-76RLO 1830.
- Project page: http://aim.pnnl.gov/projects/hypothesis_reasoning.stm

Q & A

Thank you!