Abstract

More and more applications require real-time processing of data streams in oil&gas operations, in weather monitoring, in customer relationship management, in Smart Cities and in Social Media Analytics. For instance, in the last two domains, is public transportation where the people are? Who is driving the discussion about the top 10 emerging topics across all the social networks?

A system able to answer those queries must:

- handle massive datasets,
- process data streams on the fly,
- cope with heterogeneous incomplete and noisy data,
- provide reactive answers,
- support fine-grained information access, and
- integrate complex domain models.

Indeed, systems capable of scalable stream processing exist. They can provide reactive fine-grained information access even in the presence of noisy data. Similarly, recent research on Semantic Web, an in particular to scalable Ontology Based Data Access (OBDA), showed that complex domain models can be used to offer fine grained information access to heterogenous and incomplete datasets. However, none of those solutions covers the entire spectrum of requirements.

Therefore, the research question of this thesis is:

Q.1 is it possible to make sense in real time of multiple, heterogeneous, gigantic and inevitably noisy and incomplete data streams in order to support the decision process of extremely large numbers of concurrent users?
In the spirit of a PhD thesis as collection of papers, this thesis presents a number papers that envision, elaborate, evaluate and discuss seven years of research in Stream Reasoning. The papers document that a) the Semantic Web stack can be extended so to incorporate streaming data as a first class citizen, b) the Stream Reasoning task is feasible, c) the very nature of streaming data offers opportunities to optimise reasoning tasks where data are ordered by recency and can be forgotten after a while, d) a combination of deductive and inductive stream reasoning techniques can cope with incomplete and noisy data, and e) there are application domains where Stream Reasoning offers an adequate solution. Moreover, from the lessons learnt in these years stems the research question: can orders (beyond the specific order in time) be harnessed to optimise the reasoning tasks? Answering this question is the main future work pointed out in this thesis.