H.F.R.I.
Hellenic Foundation for Research & Innovation

Description of the funded research project

2nd Call for H.F.R.I. Research Projects to Support Post-Doctoral Researchers

Title of the research project: **iQARuS - Efficient Query Answering over RDF streams using Summaries**

Principal Investigator: **Charidimos Kondylakis**

Reader-friendly title: **iQARus**

Scientific Area: **Mathematics and Information Sciences**

Institution and Country: **FORTH, Greece**

Host Institution: **FORTH**

Collaborating Institution(s):-

Project webpage: **http://www.ics.forth.gr/~kondylak/iQARuS**

Budget: **130.000 €**

Duration: **3 years**

H.F.R.I.
Hellenic Foundation for Research & Innovation

Research Project Synopsis

Streams of data are now massively, constantly arrive from Internet of Things Devices (IoT), sensors and social media, and require rapid processing, querying and integration with background knowledge in order to support further data analysis. To this direction, RDF Stream processing platforms are valuable tools, enabling query answering over RDF streams. However, so far, the current state-of-the-art in RDF Stream processing has provided either centralized engines that cannot deal with massive RDF data streams or distributed engines that offer limited reasoning capabilities. Summarization techniques on the other hand have already proved their value for indexing data, query answering, reasoning, source selection, graph visualization, and schema discovery. However, to the best of our knowledge they have not yet been exploited for stream data, which remains a completely unexplored area. iQARuS objective is to enable effective and efficient query answering over RDF stream data using summaries. Generated summaries will be smaller than the original graphs and as such they will reduce drastically the data space, enabling efficient query answering and reasoning. However, due to this reduced data space exact answers might not be always possible to be retrieved directly from summaries. We intend to explore approximate query answering and then to offer exploration operations that will allow expanding the summaries for exact query answering. In addition, incremental algorithms will enable summary updating to avoid the overhead of summary recomputation from scratch. The developed solution will cover and combine both recent window stream data and background, staged knowledge and will be evaluated extensively using both well-established RDF Stream Processing benchmarks and a new one to be generated during the lifetime of the project.

Project originality

So far, the current state-of-the-art in RDF Stream processing has focused on continuous query languages to cope with RDF data streams, and has provided either centralized engines that cannot deal with massive RDF data streams or distributed engines that offer limited reasoning capabilities.

Summarization techniques have already proved their value for indexing data, query answering, estimating the size of query results, source selection, graph visualization, and schema discovery. However, to the best of our knowledge they have not yet been exploited for stream data, which remains a completely unexplored area.

The main novelties of the iQARuS project are the following:

1. *Summarization techniques for stream data will be developed exploiting semantic summaries, investigating approaches of both super node creation and node selection.*
2. *The summarized graphs will be used for query answering, reducing drastically the data space, enabling efficient query answering and reasoning. According to the selected method for building the summary, and due to the rapidly changing information and query needs, both approximate and exact query answers will be exploited to deliver instant answers over the summaries.*
3. *Incremental update algorithms will be devised for the summaries generated over the stream data ensuring that the generated summaries have always the necessary information of interest.*
4. *Algorithms for answering queries that require the combination of both dynamic stream data with historic static or staged data. This will result in a unified platform that will enable reasoning on stream data, enabling further analysis tasks.*

To the best of our knowledge, the field of stream processing using summaries is yet completely unexplored and has the potential to offer tangible results.

Expected results & Research Project Impact

Effective and efficient stream processing has a high scientific and social impact. It has the potential to enable rapid detection of catastrophic events, weather prediction, to be used for opinion mining, detection of irregular health events, outbreaks of infectious diseases, fraud detection etc. based on rapidly changing streams of data.

As the platform will be domain independent will be able to benefit various scientific and social endeavors that require rapid processing of big volume of streams. Already the stream processing benchmarks include scenarios related to smart city applications, e-commerce and social networks and within our project, we intent also to look for interesting exploitation scenario for e-health streams data streams. To conclude, our proposal has to potential for a really high scientific and social impact.

The importance of this funding

A lot of studies have already analyzed the impact of funding and have already shown the great impact on the productivity and the performance of the funded researchers in terms of quantity and quality of their publications and on the potential scientific collaborations.

The research proposed in this project would not have progressed otherwise, as the proposed ideas are in the stage of conception and further research is required till they mature and create actual benefit.

H.F.R.I.
Hellenic Foundation for Research & Innovation

COMMUNICATION

185 Syggrou Ave. & 2 Sardeon St. 2
171 21, N. Smyrni, Greece
+30 210 64 12 410, 420
communication@elidek.gr
www.elidek.gr