On Teaching Web Stream Processing
Lessons Learned
ICWE 2020 - Virtual Conference

Riccardo Tommasini - 10/06/2020
Who We Are
Who I Am

Riccardo Tommasini, PhD
Assistant Professor
at University of Tartu, Estonia
Web Stream Processing
What is?
Data Velocity is the speed at which data are processed and insights are obtained.

— Doug Laney, Gartner, 2001
— Philip Russom, TDWI, 2011.
Web Stream Processing

What it is?

Sort out all the colours in the streams
Web Stream Processing

What it is?

Which are the top-2 most frequent colours in the last minute?

(time, 13), (red, 8), (pink, 8)
Web Stream Processing

What it is?

• The Web is a decentralised environment built around the notion of resources, i.e., anything that an has a URI.

• Web agents are software components that creates, deletes, and interact with resources in behalf of a user.

• Web users are typically interested in complex analyses that span across data sources, e.g., search engines.
Data Variety is the diversity of data in a problem space.

– Doug Laney, Gartner, 2001
– Philip Russom, TDWI, 2011.
Can we tame Variety? on the Web

Semantic technologies, e.g., ML, RDF, OWL, and SPARQL, foster data sharing and interoperability across applications.
Web Stream Processing

What is it?

[Diagram showing various activities with statistics within a 60-second timeframe.]

Created By:
@LoriLewis
@OfficiallyChadd
Semantic Technologies

- Process Heterogeneous Resources
  - Incomplete & Noisy Data
- High Complex Domains
- Expressive Query Languages

Stream Processing

- Process Streams & Events
- Continuous Semantics & Reactive Answers
- Simple Data Models (Tuples)
- Simple Query Languages
Which are the top-2 most frequent cool colours in the last minute?
Web Stream Processing
Adding Context to Stream Processing

Which are the most frequent sentiments in the last minute?

Enthusiasm!

1 minute wide window
Almost Done!!

Almost!!

Oh Yes!

Oh Man!

Semantic Technologies

Semantic Technologies

Web Stream Processing

Vice-Versa is also valid
Learning/Teaching Web Stream Processing

Challenges

• Students must master two technological stacks
  • Semantic Web Stack to tame variety
  • Stream Processing stack to tame velocity
• They need to understand the stream processing paradigm-shifts
  • From Data at rest to data in motion
  • From post-hoc analysis to real-time insights
It is hard to estimate the effort for teaching/learning WSP
From Problem Statement to Research Questions

• (Q1) Can the participants formulate a Continuous Information Need?

• (Q2) Can the participants design a Web stream processing solution to answer a Continuous Information Need?

• (Q3) Can the participants implement a Web stream processing solution to answer a Continuous Information Need?
ACTION RESEARCH CYCLE
VISUAL BY JOHN SPENCER

PLANNING

ACTION

ANALYSIS

CONCLUSION
# Action Research

**For our Investigation**

<table>
<thead>
<tr>
<th>What</th>
<th>Action Research</th>
<th>Interoperability &amp; Semantic Technology Course</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Who</strong></td>
<td>Conducted by educators on students.</td>
<td>Riccardo Tommasini</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Emanuele Della Valle</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Marco Balduini</td>
</tr>
<tr>
<td><strong>Where</strong></td>
<td>In schools and Classrooms</td>
<td>During the IST Course</td>
</tr>
<tr>
<td><strong>Why</strong></td>
<td>To take action and possibly induce changes in the studied environment.</td>
<td>To estimate the effort required to learn WSP</td>
</tr>
<tr>
<td><strong>How</strong></td>
<td>Explore the reality and effects of interventions on it using qualitative and</td>
<td>Establishing an educators/students feedback loop by Using</td>
</tr>
<tr>
<td></td>
<td>quantitative methods</td>
<td>Webforms</td>
</tr>
</tbody>
</table>
Concept Generation

Description and Aims

- This phase aims at assessing the students' ability to formulate a continuous information need starting from sample data.

- The students have to formulate an information need on (historical) data streams from the Telecom Italia Open Big Data challenge.

- Students are also required to understand the meaning of the data consulting the available documentation. Notably, the students were not domain experts.
### Results

<table>
<thead>
<tr>
<th></th>
<th>How many calls were made to “Milan's Dom” from anywhere in Milan during rainy weekends?</th>
</tr>
</thead>
<tbody>
<tr>
<td>II</td>
<td>Report the calls number made to “Milan’s Dom” from anywhere in Milan, every weekend if it rained for 24 hours.</td>
</tr>
<tr>
<td>III</td>
<td>Report every hour the number of calls that were made to “Milan’s Dom” from in Milan, if was raining within 24h.</td>
</tr>
</tbody>
</table>
Concept Generation

Complexity

![Bar chart showing complexity over different datasets and formulations.]

- **Dataset Comprehension**
- **Formulation**
Design Phase

Description and Aims

• This phase aims at assessing the students’ ability to design a solution that solves the continuous information need.

• As in the previous phase, we collected design proposals using a Web-forms.

• We asked the student to defend their design choices.
Design Phase

Results

• we accepted two alternative approaches:
  
  • A pure Semantic Web approach that used a SPARQL query to evaluate periodically,
  
  • A Web Stream Processing approach that used a C-SPARQL query.
  
• It emerged that the two tasks had similar complexities.
Design Phase

Complexity

![Bar Chart]

- SPARQL Design
- C-SPARQL Design
Design Phase

Complexity

![Bar Chart]

- **Dataset Comprehension**
- **Ontology Design**

Values on the y-axis range from 0.00 to 1.00.
Design Phase

Complexity

![Bar Graph showing complexity of Dataset Comprehension and Mapping Design over different values.]
Implementation Phase

Description and Aims

- This phase aims at assessing the students' ability to implement the designed solution.

- The task required to realise two alternative architectures of increasing complexity:
  - one based on Semantic Technologies
  - one based on RDF Stream Processing Engines.

- We reserved 5 hours of class-work where the student worked under our supervision.
Implementation Phase

Results

- We collected the projects with web form.
- We interviewed those who evaluated the project as complete,
- we asked the other to write down a retrospective during the final exam.
Conclusion

and future work

• Form our three-phase evaluation analysis, it emerges that the overall complexity of the project was quite high

• The WSP community needs
  • more advanced teaching material and
  • more mature tools for designing and implementing applications.
Thank you for listening!

Questions???

In Memoriam
Sherif Sakr
1979-2020