Personalized Stream Analysis with Preference SQL

Lena Rudenko and Markus Endres
University of Augsburg
Germany

Workshop Präferenzen und Personalisierung in der Informatik - PPI17 @ BTW 2017
What Happens Online in 60 Seconds?
Managing Content Shock in 2016

- YouTube: 400 hours of video uploaded
- Email: 205.6 million emails sent
- WhatsApp: 44.4 million messages sent
- Facebook: 3.3 million posts
- Wordpress: 1,212 posts
- Google searches: 347,222
- Twitter: 422,340 tweets
- Instagram: 55,555 photos uploaded

Yearly Growth:
- 2013: 60 seconds
- 2014: 100 seconds
- 2015: 118 seconds
- 2016: 128 seconds

Personalized Stream Analysis with Preference SQL – Lena Rudenko
Personalized Stream Analysis with Preference SQL – Lena Rudenko
Stream query processing is very important and on time today.

Examples:

- sensor data (weather data, positioning systems, vital signs tracking, etc.)
- exchanges (stocks, commodities, currency)
- social networks (Instagram, WhatsApp, Facebook, Twitter, etc.)

Stream – a flow of data objects. The stream data is:

- continuous
- endless
- available over time
- does not take the form of persistent database relation
Our Goal

We present an approach of data streams evaluation which takes user preferences into account to provide more relevant results for each user compared to approaches using hard constraints-evaluation.
User preferences are like soft constraints:

„If my favorite choice is available in the dataset, I will take it. Otherwise, instead of getting nothing, I am open to alternatives, but show me only the best ones available.“
Preference SQL: declarative extension of SQL by preferences.

SELECT STREAM * FROM TwitterStream
PREFERRING
WHERE tweet_language IN ('de') ELSE ('en')
PARETO followers_count HIGHEST

User has the best possible results at any time, but never an empty set.
As example stream source we use Twitter – online social networking service:

- very large number of tweets (500 million daily)
- important and interesting records together with spam and trash
- easy access by the public Twitter API
- huge amount of diverse attributes
Personalized Stream Analysis with Preference SQL – Lena Rudenko

Stream Processing Framework
Stream Processor – transformation of stream objects to a list of single attribute.

Data Accumulator – splitting of data stream into finite parts by grouping them into chunks.

Preference SQL – analysis of data chunks within Preference SQL.
DEMO
Summary:

- First preference-based stream analyzer
- Provides user personalized best-matches results

Outlook:

- Implementation of various stream connectors, e.g. Facebook, WhatsApp, Stock
- Developing of efficient evaluation algorithms
- Experiments
Thank you for the attention!

Lena Rudenko – lena.rudenko@informatik.uni-augsburg.de