



Personalized Stream Analysis with Preference SQL

Lena Rudenko and Markus Endres

University of Augsburg

Germany

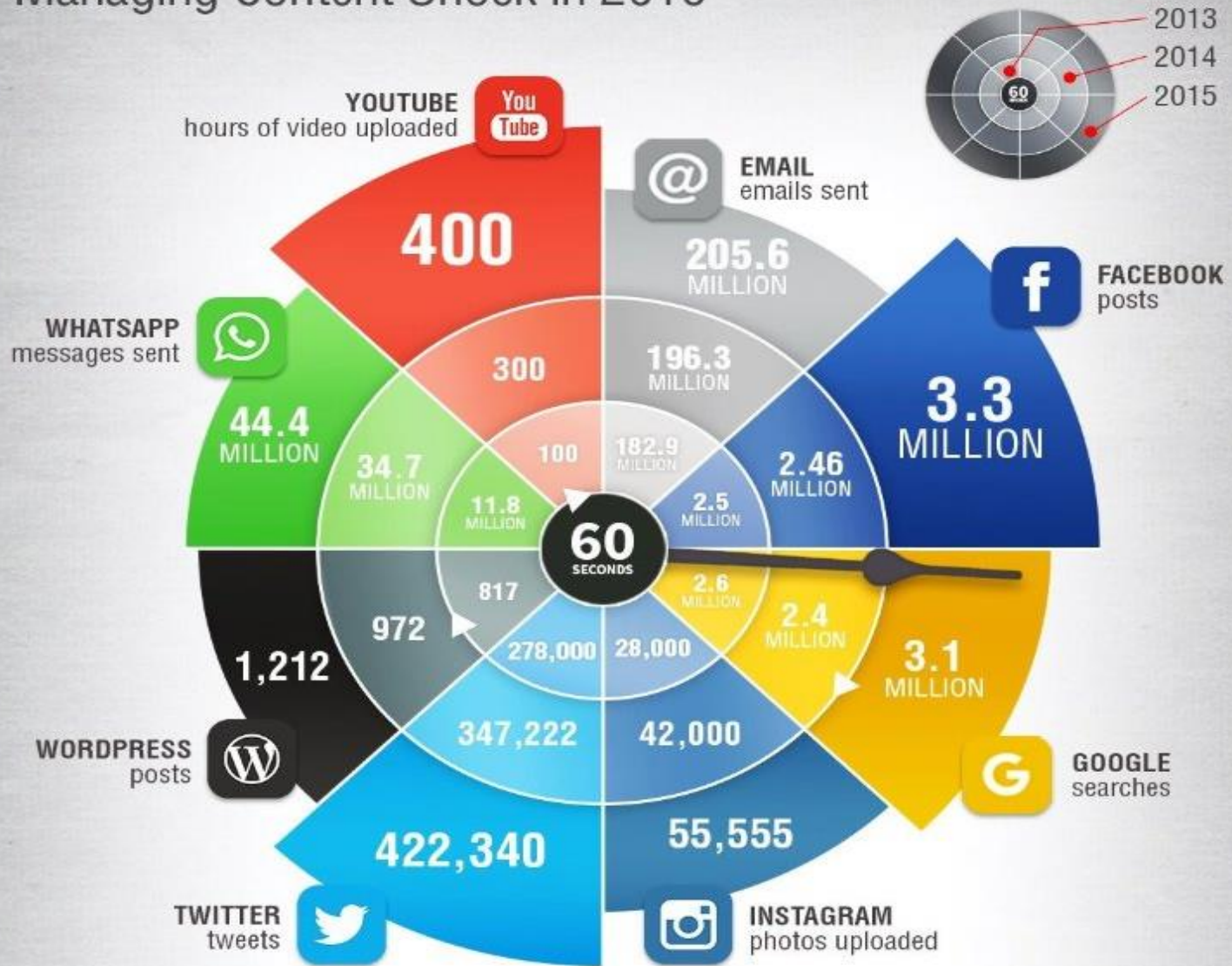
Workshop Präferenzen und Personalisierung in der
Informatik - PPI17 @ BTW 2017





What Happens Online in 60 Seconds?

Managing Content Shock in 2016






John Klee @_Hermeneutik · 1 Std.
Es ist sehr einfach in der Herde über **Trump** zu lästern. Zu analysieren, was Obama falsch gemacht hat, wäre allerdings wichtiger und richtig.

onyx @onyx0815 · 1 Std.
Sehr geile tweets unter #winieatwlv (warum ich nicht in einer afd/trump-welt leben will). Könnte alles durchfaven 🙌
Danke an @LetKiser dafür

Mario Sixtus 馬六 @sixtus · 1 Std.
Kann es nicht vielleicht sein, dass auch der Umschlag in dem "Donald Trump" stand, der falsche war?

The New York Times @nytimes
"This is not a joke. I'm afraid they read the wrong thing."
Video: nyti.ms/2muF8kS

VOGUE_Germany @VOGUE_Germany · 2 Std.
Ein Blick hinter die Kulissen der #Oscars 🏆 : bit.ly/2lMo3mF #video #BehindtheScenes @swarovski



Don Sebo @don_sebo · 2 Std.
Besteht noch die Möglichkeit, dass auch bei der US-Wahl die Umschläge vertauscht wurden?

Christoff @struki · 2 Std.
Personen, die eine wichtige Trophäe nur für ein paar Minuten gewonnen haben:
- @RyanGosling
- @GeraldAsamoah14
#Oscars 🏆 #bundesliga

Stream query processing is very important and on time today.

Examples:

- sensordata (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.



Our Goal

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

Preference SQL: declarative extension of SQL by preferences.

<pre>SELECT STREAM <attribute_list> FROM <stream_reference> WHERE <hard_conditions> PREFERRING <soft_conditions></pre>	<pre>SELECT STREAM * FROM TwitterStream PREFERRING tweet_language IN ('de') ELSE ('en') PARETO followers_count HIGHEST</pre>
--	--

(a) PreferenceSQL stream syntax

(b) PreferenceSQL stream example

User has the best possible results at any time, but never an empty set.



Twitter

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



Stream Processing Framework

```
SELECT STREAM tweet_text, source, location FROM TwitterStream
PREFERRING source IN 'Twitter for Android' PARETO acc_created_at LATEST;
```

SELECT STREAM - - +

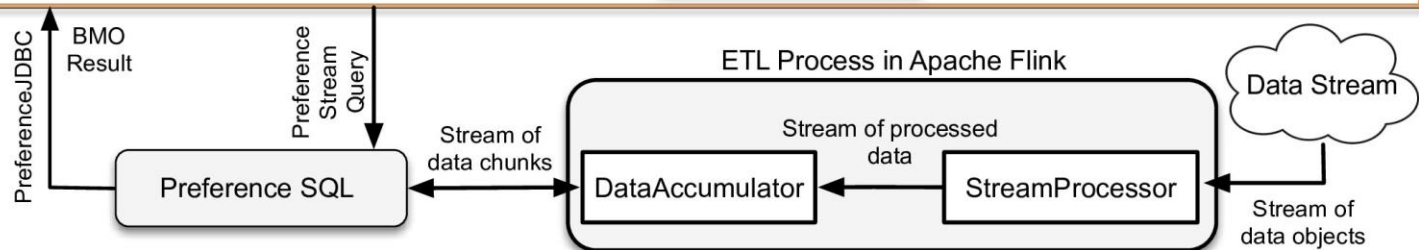
FROM

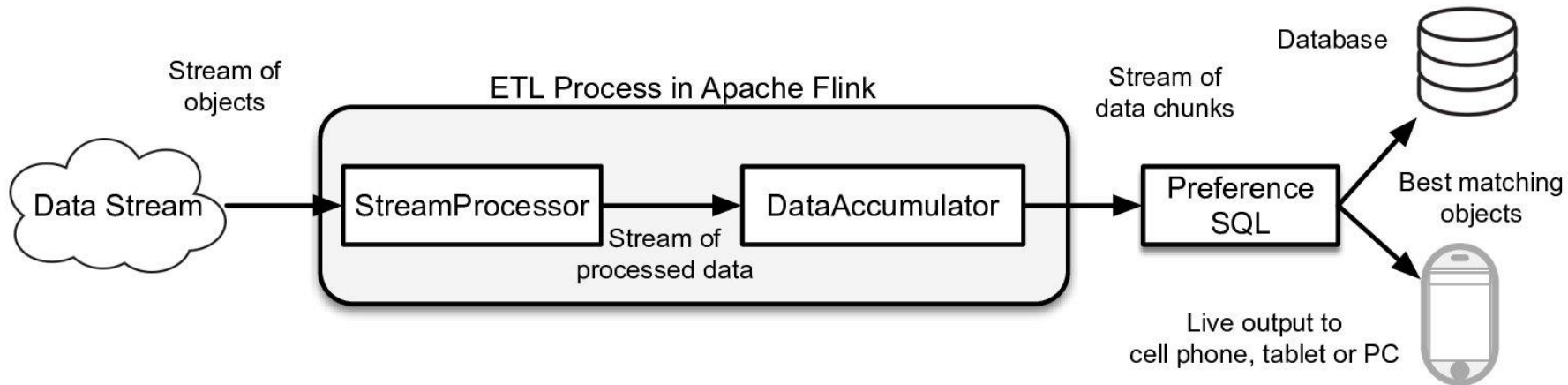
PREFERRING **IN** + -

PARETO **LATEST** + -

SEARCH

Results:
People aren't dolls, You can't just play with them until they
@DELTA my itinerary has disappeared from the system. my
ing inside, or all of the above,, Twitter for A
tickets dan jacobs GVXHQB, Twitter for iP





- **StreamProcessor** – 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