TUM DI LAB – Celonis Project

NLP and Process Mining - Final presentation

The Team

Rachid Ellouze

Gayatri Kudchadker

Manisha Singhal

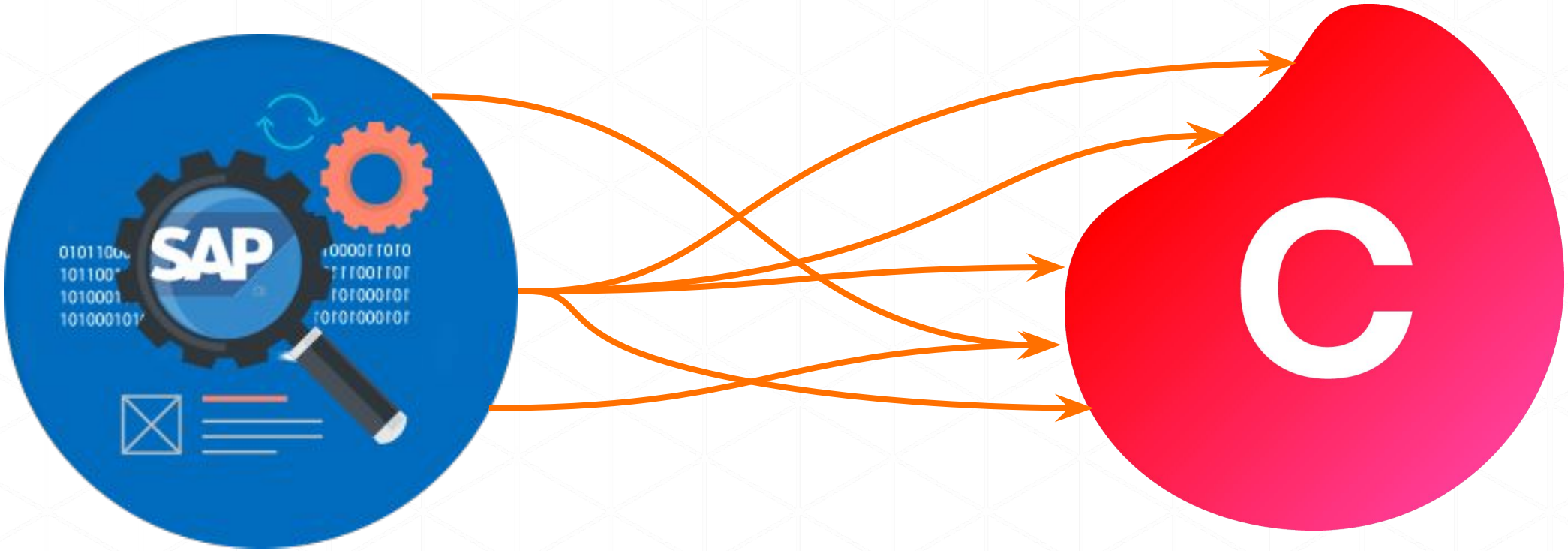
Claas Brüß

Eduardo Goulart

OUTLINE

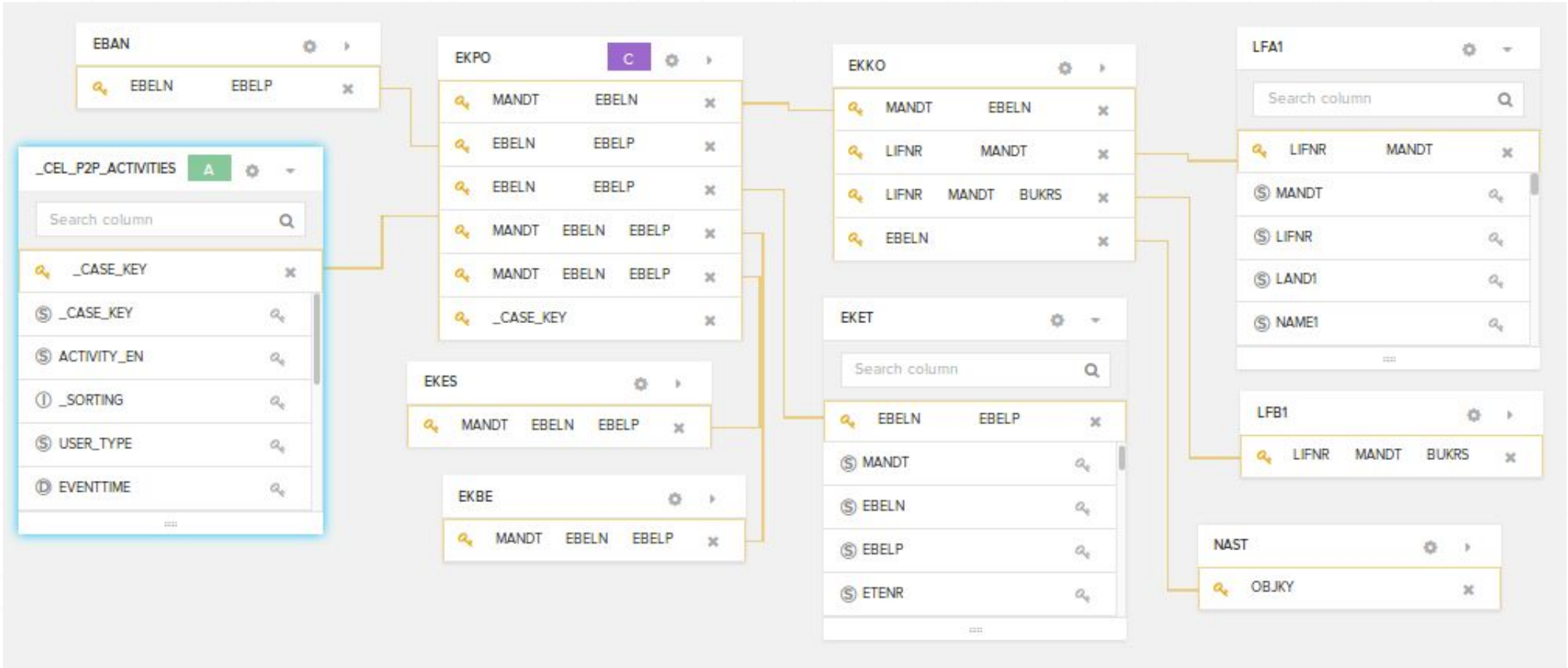
- Celonis Dashboard Interface
- Project Goals
- NLP Assistant
- NLP Demo
- Anomaly Detection
- Anomaly Demo
- Summary

Process Mining

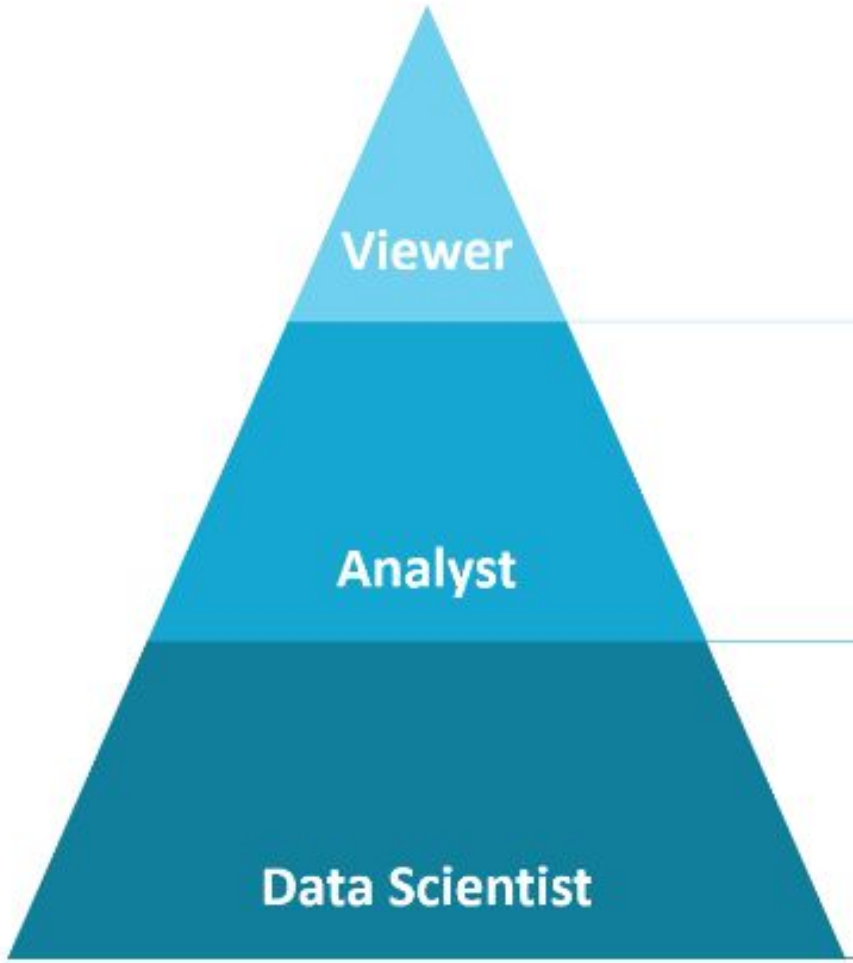


Celonis Dashboard Interface

Data Model


EBAN
EBELN EBELP
_CEL_P2P_ACTIVITIES A
Search column
_CASE_KEY
_CASE_KEY
ACTIVITY_EN
_SORTING
USER_TYPE
EVENTTIME
EKPO C
MANDT EBELN
EBELN EBELP
EBELN EBELP
MANDT EBELN EBELP
MANDT EBELN EBELP
_CASE_KEY
EKES
MANDT EBELN EBELP
EKBE
MANDT EBELN EBELP
EKKO
MANDT EBELN
LIFNR MANDT
LIFNR MANDT BUKRS
EBELN
EKET
Search column
EBELN EBELP
MANDT
EBELN
EBELP
ETENR
LFA1
Search column
LIFNR MANDT
MANDT
LIFNR
LAND1
NAME1
LFB1
LIFNR MANDT BUKRS
NAST
OBJKY

Celonis Role Model



Become familiar with the Celonis user interface. Learn how to navigate through the software and how to work on existing analyses in order to identify process inefficiencies/deviations and their root causes.

Learn every single aspect of the Celonis user interface and use the various functionalities to create appealing, target-oriented, and meaningful analyses.

Get to know Celonis from A to Z, covering installation & maintenance of the software, connection of processes from IT source systems, and analyses configuration.

Project Goals

Project goals

- Develop an intelligent assistant for strategic decision makers in the Viewer user category
- Enabling initial data exploration through simple queries without Analyst support
- Draw attention to problematic cases and areas with high improvement potential
- Motivate Analyst’s work in building dashboards with more specific requirements for root cause investigation

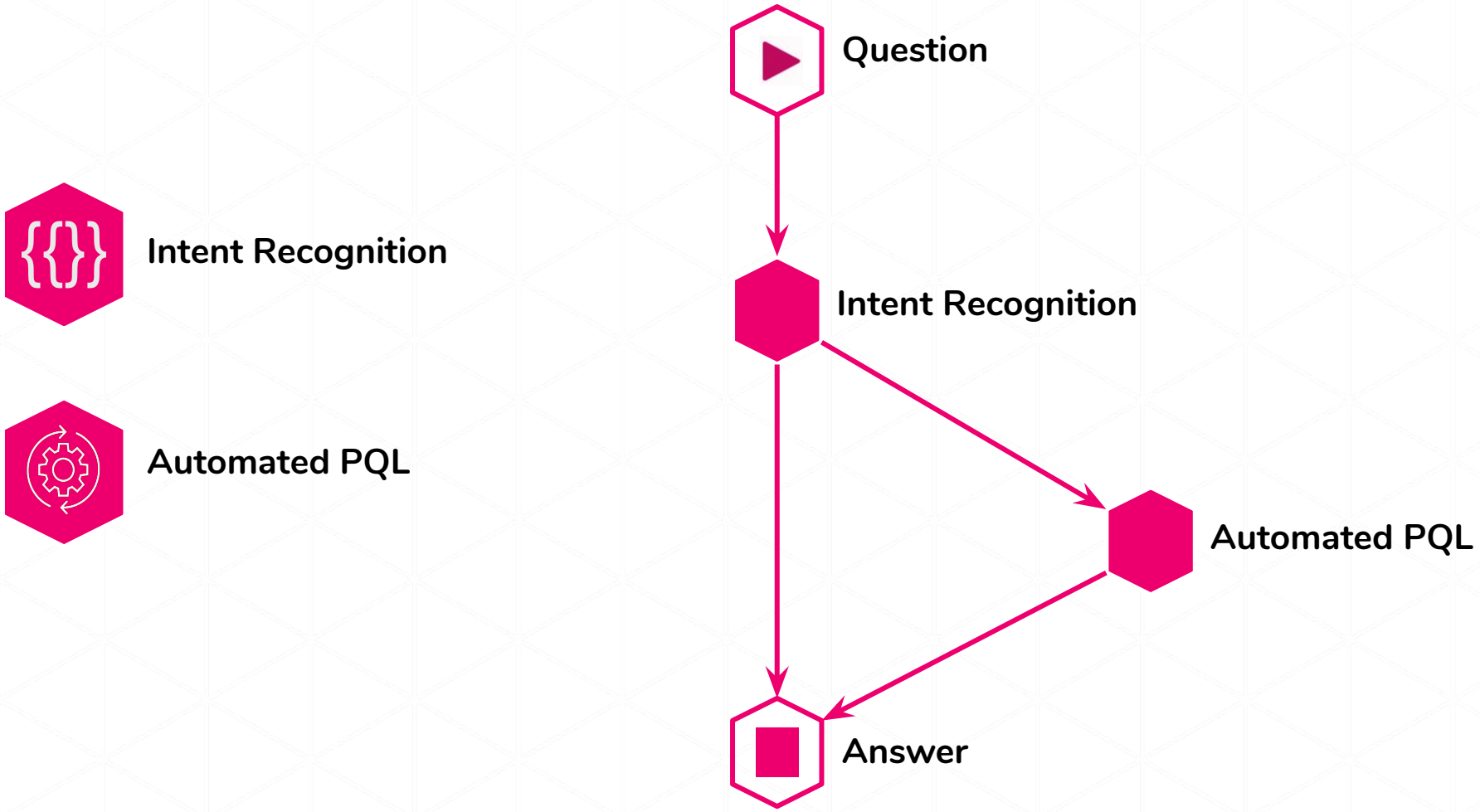
NLP Assistant

PQL Query

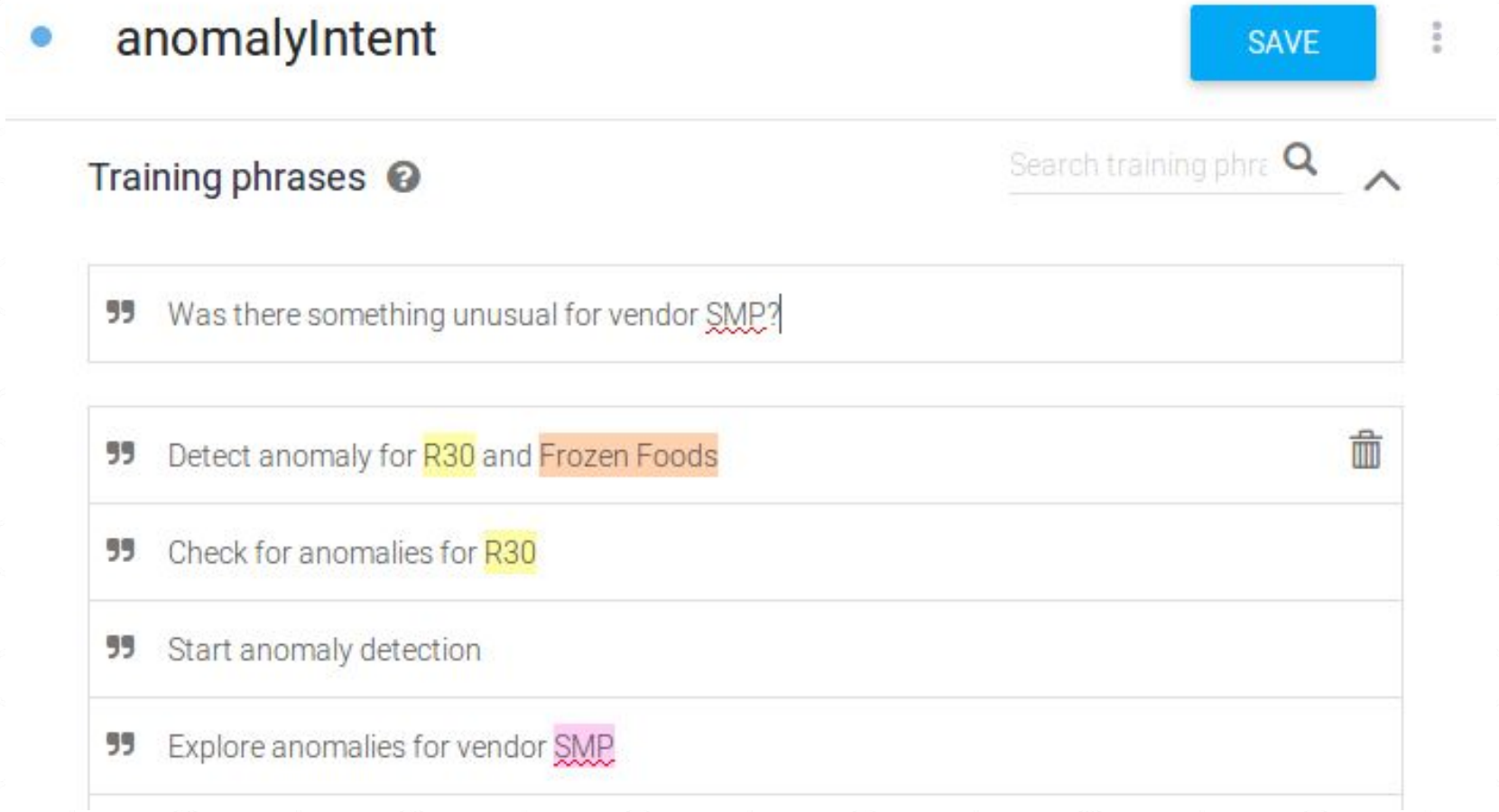
```
TABLE ( COUNT ( DISTINCT "LFA1"."LIFNR" ) )
```

```
TABLE (
    "EKPO"."EBELN", "EKPO"."EBELP", "EKPO"."MATNR",
    "EKPO"."MENGE", "EKPO"."MEINS", "LFA1"."LIFNR "LFA1"."NAME1",
    REMAP_TIMESTAMPS (
        {d '2018-04-27'}, DAYS, WEDNESDAY THURSDAY FRIDAY TUESDAY MONDAY
    ) - REMAP_TIMESTAMPS (
        PU_MAX (
            "EKPO", "_CEL_P2P_ACTIVITIES"."EVENTTIME",
        "_CEL_P2P_ACTIVITIES"."ACTIVITY_EN"
            IN ('Send Purchase Order','Send Purchase Order Update')
        ), DAYS, WEDNESDAY THURSDAY FRIDAY TUESDAY MONDAY
    )
)
```

Architecture- Hybrid Approach



Intent recognition



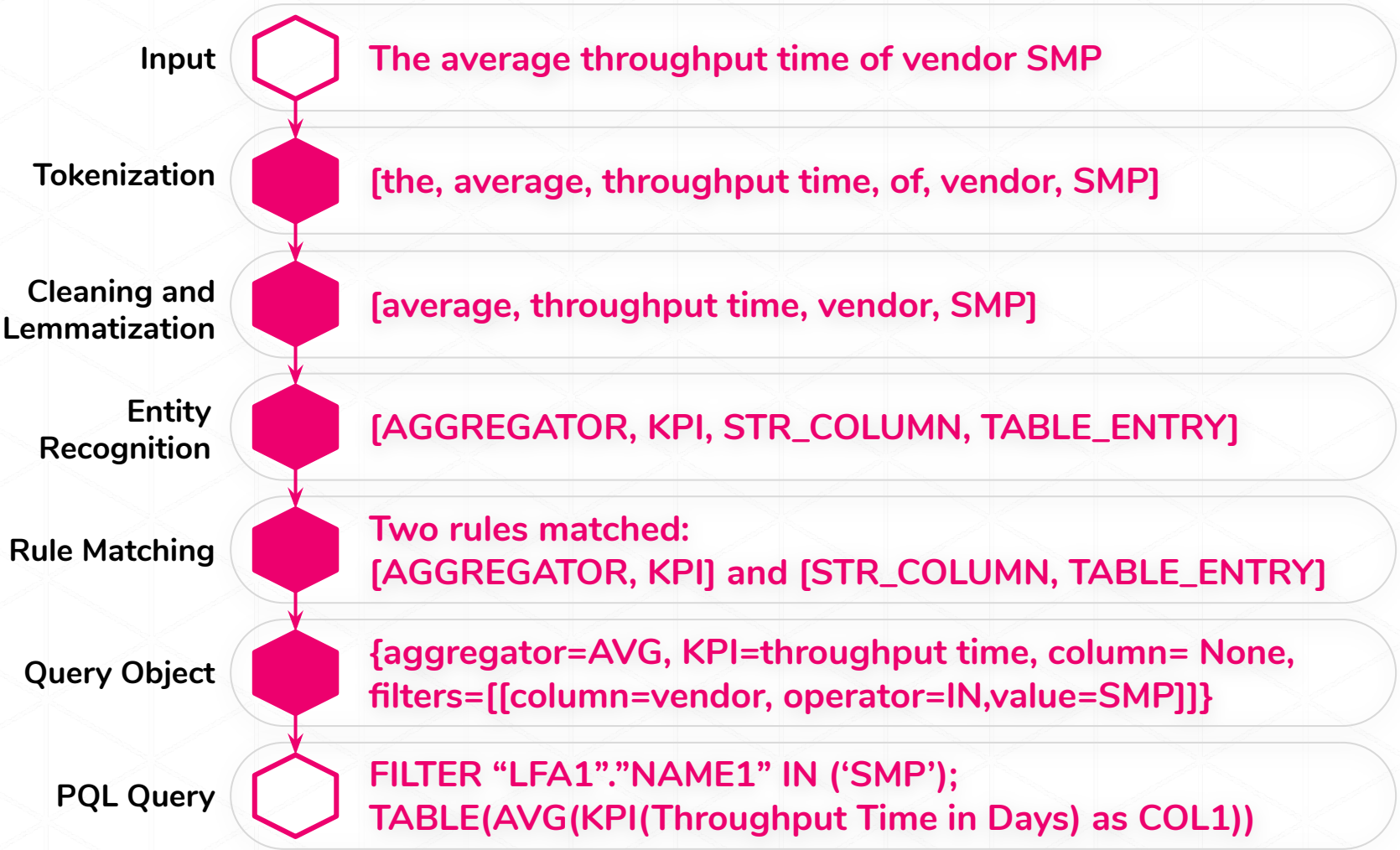
Automated PQL

Pre-processing

Columns

Table entries

KPI


Input
The average throughput time of vendor SMP
Tokenization
[the, average, throughput time, of, vendor, SMP]
Cleaning and Lemmatization
[average, throughput time, vendor, SMP]
Entity Recognition
[AGGREGATOR, KPI, STR_COLUMN, TABLE_ENTRY]
Rule Matching
Two rules matched:
[AGGREGATOR, KPI] and [STR_COLUMN, TABLE_ENTRY]
Query Object
{aggregator=AVG, KPI=throughput time, column= None,
filters=[[column=vendor, operator=IN,value=SMP]]}
PQL Query
FILTER “LFA1”.”NAME1” IN (‘SMP’);
TABLE(AVG(KPI(Throughput Time in Days) as COL1))

Features

- Detect Intents to trigger special functions
- Take action within Celonis' framework
- Generate complex PQL queries
- Parse PQL to natural language

Extra Features

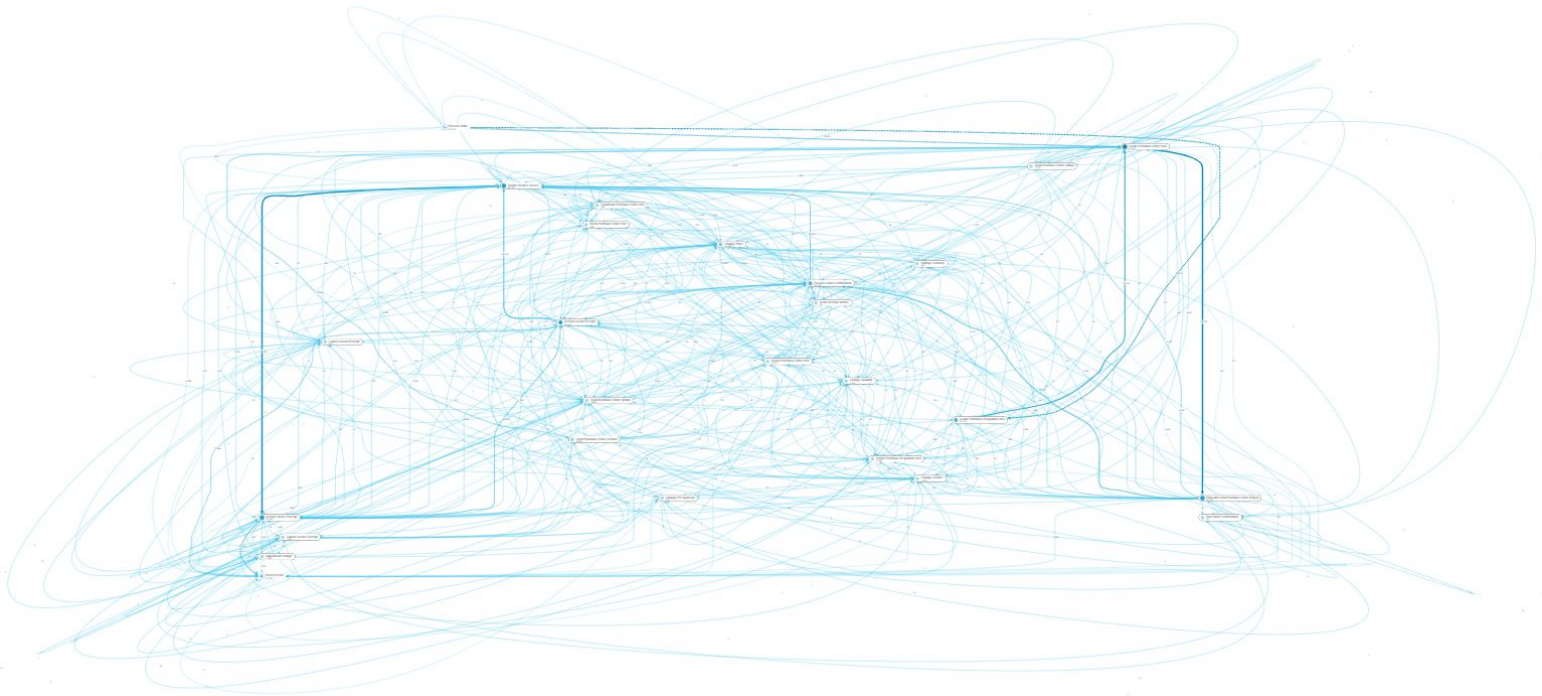
- Speech recognition
- Auto-complete
- Small talk
- Return answers in form of charts and tables

NLP Demo

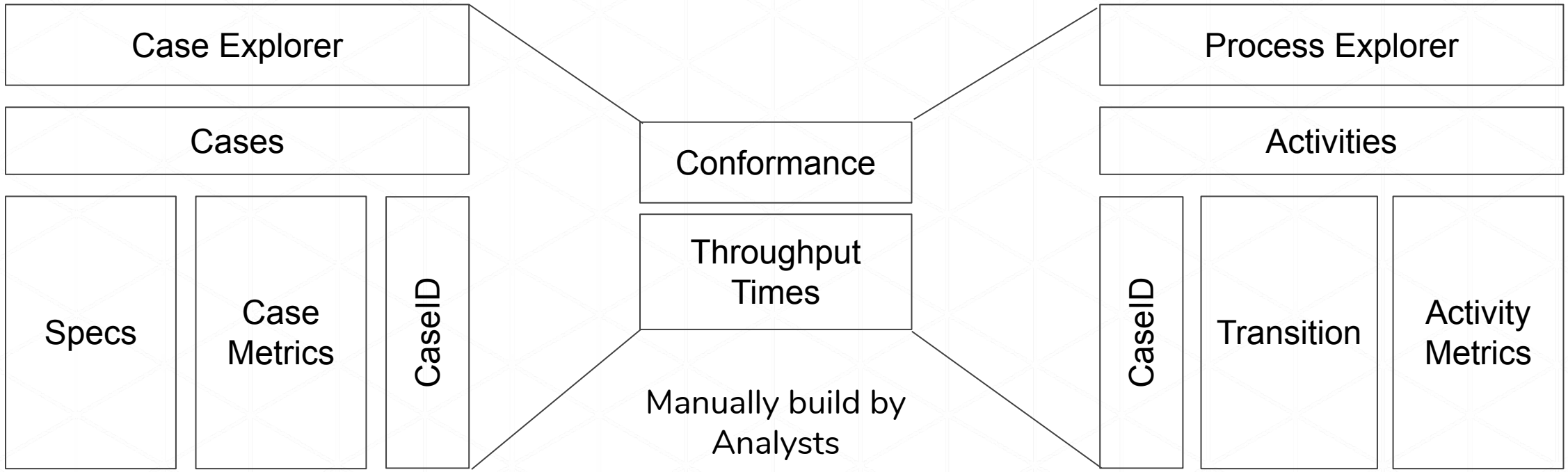
Anomaly Detection

The Dataset

- Purchase to Pay Process
- Ca. 67.000 Cases
- 24 Different Activities
- 347 Unique Transitions
- Covering 26 months from 2008 to early 2010
- 182 Vendors (10 to ca 20k cases)
- no clear indicators on data quality



Views on Data



Why and for Whom?

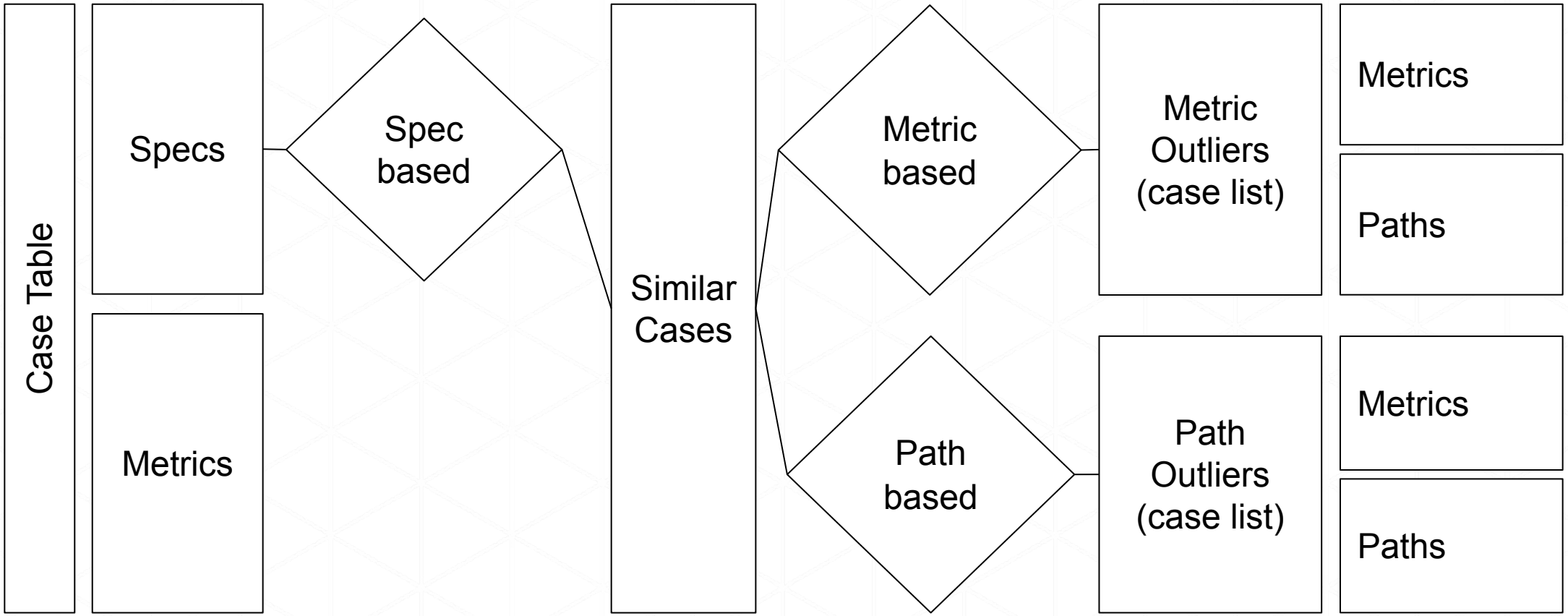
- Empower strategic decision making processes for users that Celonis describes with the Viewer group through
- Draw attention to problematic cases that nobody was looking for specifically in the form of predefined analysis
- Motivate specific Analysis being built for investigation of the root cause

Detection Approach

- Unfortunately we are lacking information about documentation context or the domain knowledge

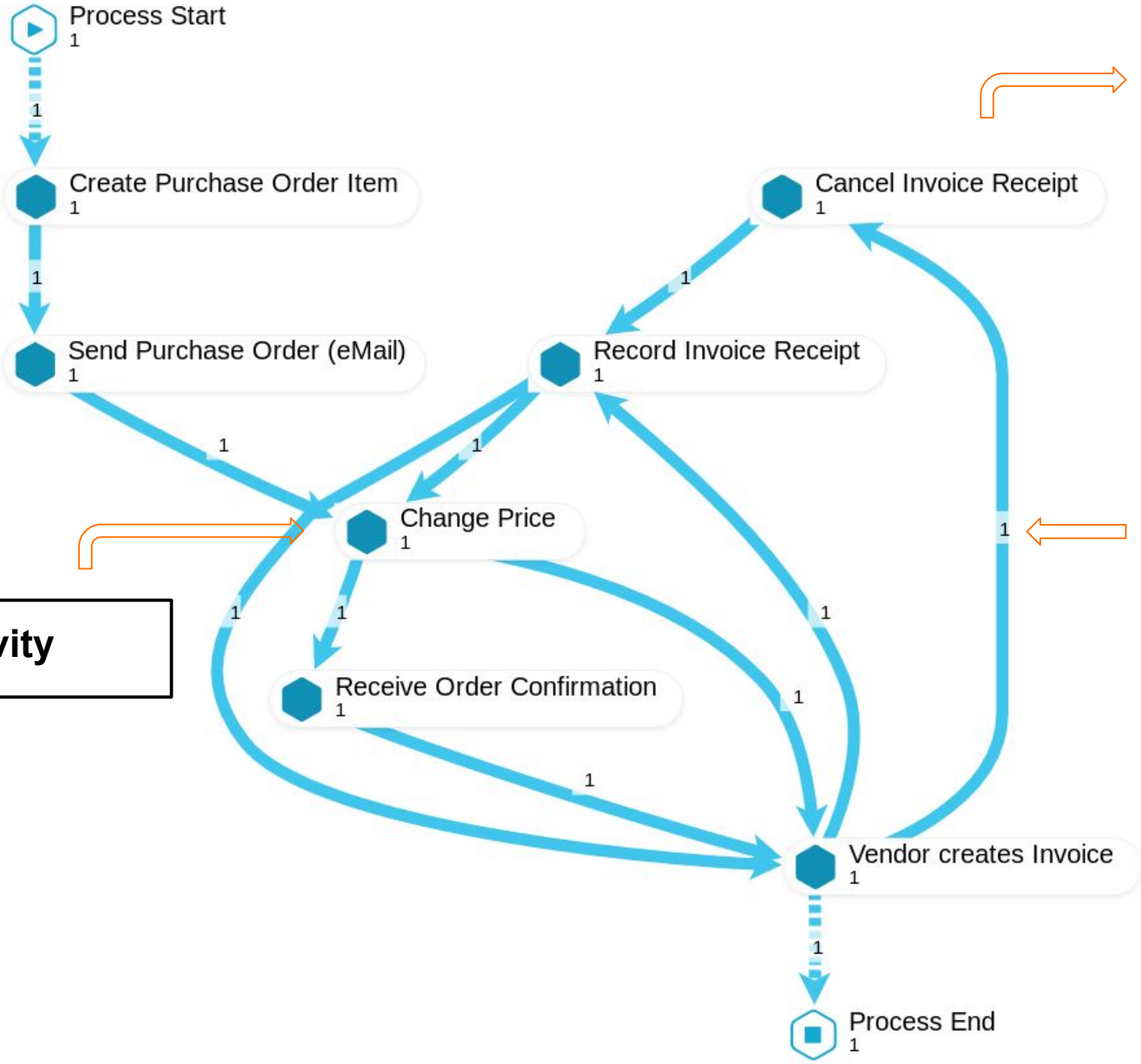
- What are anomalies then? And what is normal?
  - Are there multiples normals?
  - Can we bring the process information and case information together in a meaningful way?

Similarity Filter



Constructed KPIs

1	Case Count	Total number of Cases for a given Vendor and Time
2	Total Activities	Count of total Activities present in a Case
3	Manual Updates	Number of Activities executed manually
4	Throughput Time	Number of Days taken to finish an order
5	Deletion/Cancellations	Number of Deleted / Cancelled Activities
6	Repeated Activities	Total number of Repeated Activities in a Case
*7	Change Activities	Number of Activities with respect to Component Change
* Change Activities - { Change Price, Change Currency, Change Quantity, Change Vendor, Change PR Approval}		



Cancellation

Loop in the Process. Activities repeated multiple times.

Change Activity

Query: What was unusual for Vendor AluCast in August 2009?

KPI: Change Activities

PQL Query

```
FILTER "LFA1"."NAME1" ='AluCast';
```

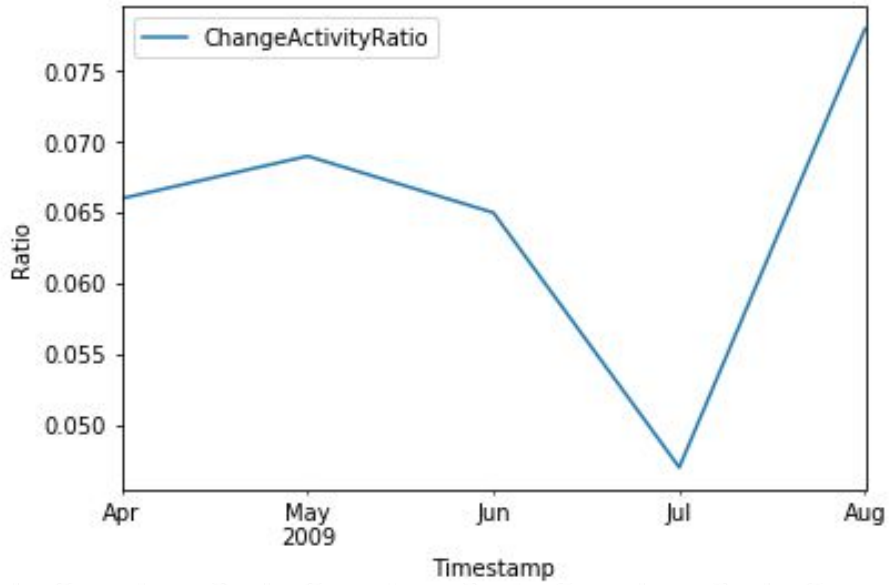
```
TABLE(ROUND_MONTH("_CEL_P2P_ACTIVITIES"."EVENTTIME") AS "Date",
      SUM(CASE WHEN "_CEL_P2P_ACTIVITIES"."ACTIVITY_EN" LIKE 'Change%'
          THEN 1
          ELSE 0
          END) AS "NumChangeActivity" ,
      COUNT("_CEL_P2P_ACTIVITIES"."ACTIVITY_EN") AS "TotalActv"
     )
```


28

Timestamp	NumChangeActivity	TotalActivity	ChangeActivityRatio
2009-07-01	14	299	0.047
2009-08-01	44	562	0.078

```
Vendor: AluCast
Queried Month: August 2009
Last Month for which the vendor was operational: July 2009
There was a increase in Change Activities by 65.96%
Please verify why there is more than 50% increase in Change Activities in the queried Month.

<Figure size 1080x720 with 0 Axes>
```

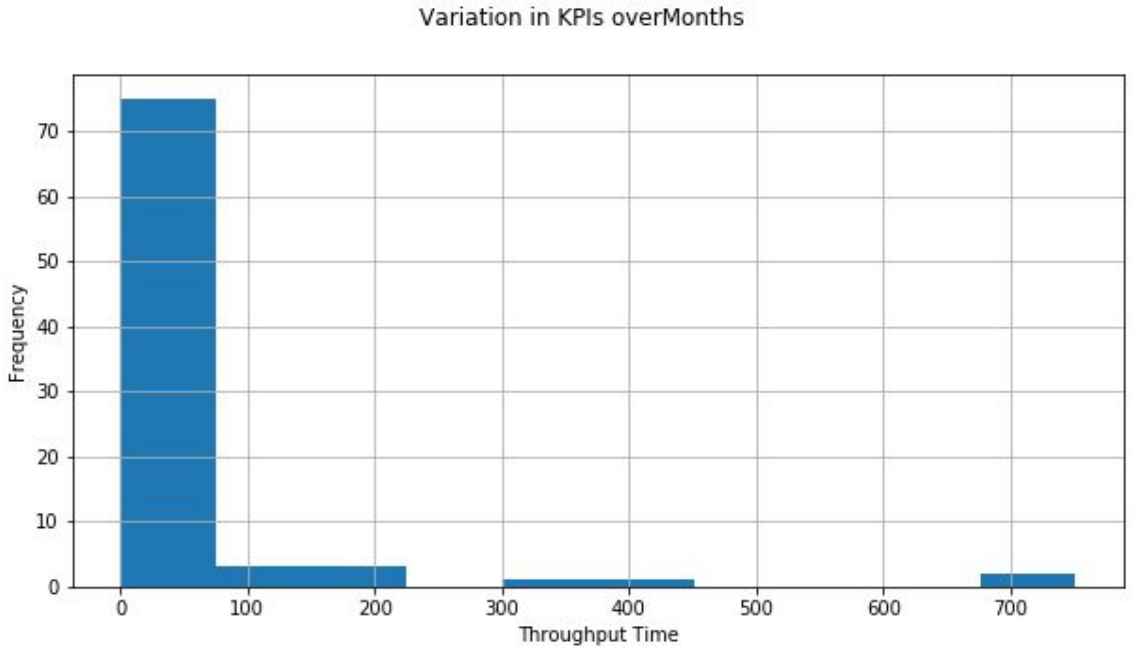


Query: What was unusual for Vendor lnes in Jan 2008 for the material Packaging and purchasing group 026?

KPI: Throughput Time

	CaseId	ThrDays	anomalyFactor	Variant
0	4376	751.0	0.500	Create Purchase Order Item, Send Purchase Orde...
1	4964	717.0	0.750	Create Purchase Requisition Item, Create Purch...
2	5440	427.0	1.000	Create Purchase Requisition Item, Change Vendo...
3	4228	365.0	0.667	Create Purchase Requisition Item, Create Purch...
4	5448	220.0	0.500	Create Purchase Order Item, Send Purchase Orde...

```
Vendor: Inés
Material: Packaging
Purchasing Group: 026
The Throughput Time plot is skewed to the right.
The 95th percentile value is 214.2 days which is 161.32 days higher than the average.
The table shows cases with throughput time above the 95th percentile having highest anomaly factor.
Anomaly Factor = Change Activity Ratio + Manual Update Ratio + Delete Cancel Ratio + Repeat Activities Ratio
```



Anomaly Detection Demo

Summary

Summary

A more intuitive interface to Celonis

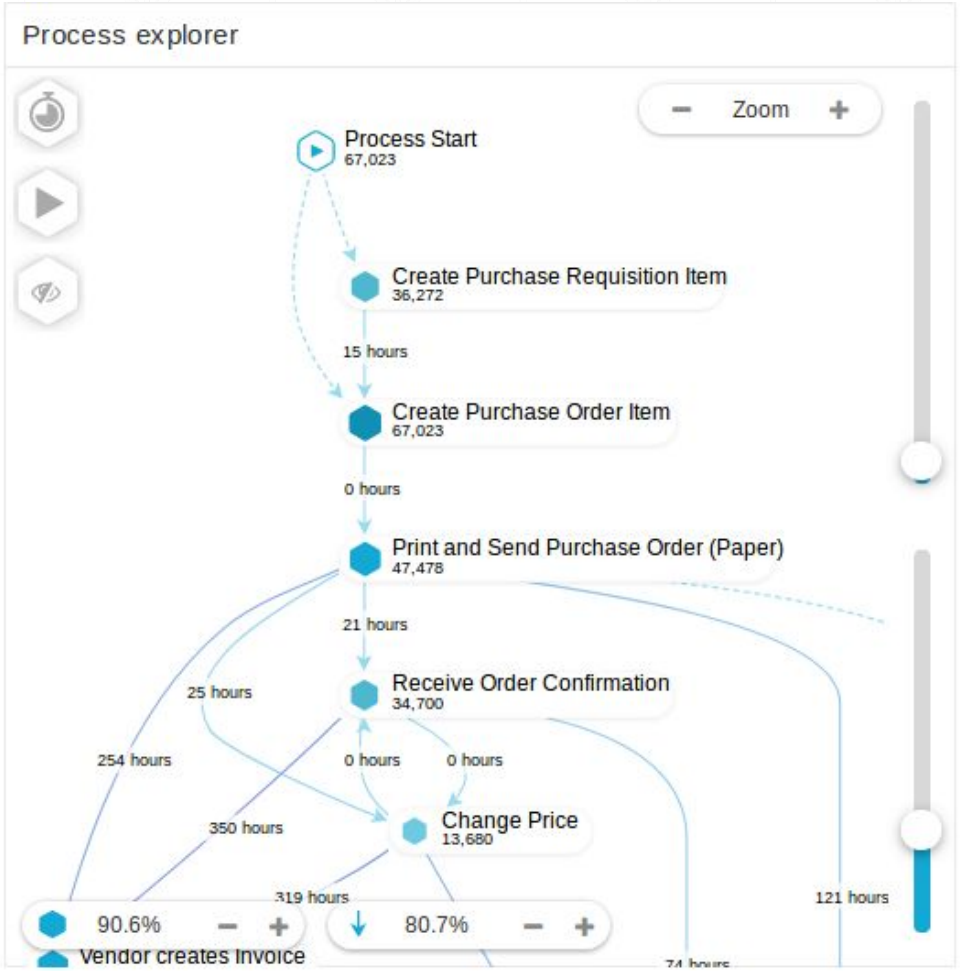
- Smart assistant for people who are not expert in PQL and Celonis framework
- It's also able to answer aggregation queries
- Furthermore, users can use the assistant to detect anomalies in the dataset
- Provide clearer requirements for analysis setup for root cause investigation

celonis

Thank you,

Any questions?

Celonis Framework



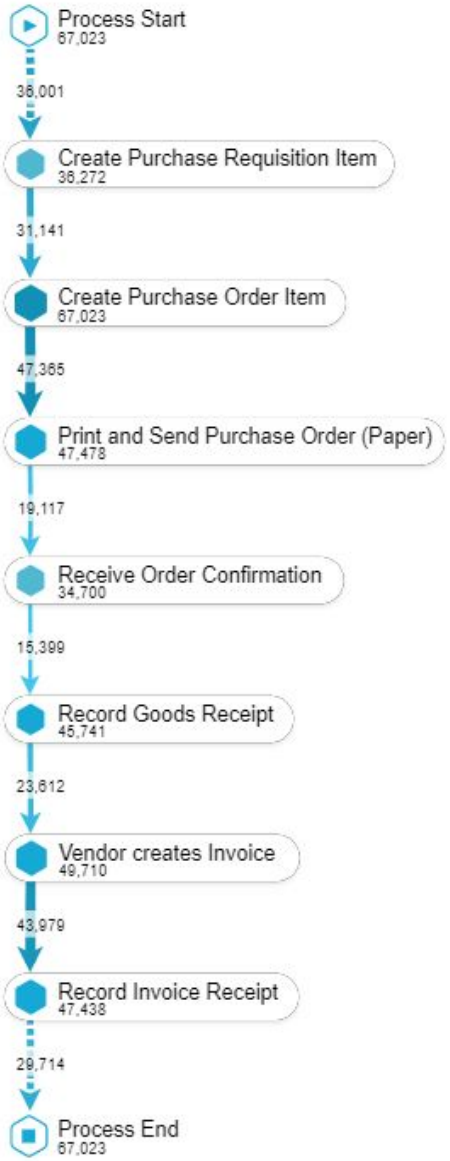
Purchase order items by Vendor — Vendor

Vendor	# PO items	Days between fir…	Throughput time i…
-	11,148	11	38.41
0000000002 - Ele…	19	17	20.75
0000000005 - Saf…	28	19	12.71
0000000008 - Jos…	9	46	8.67

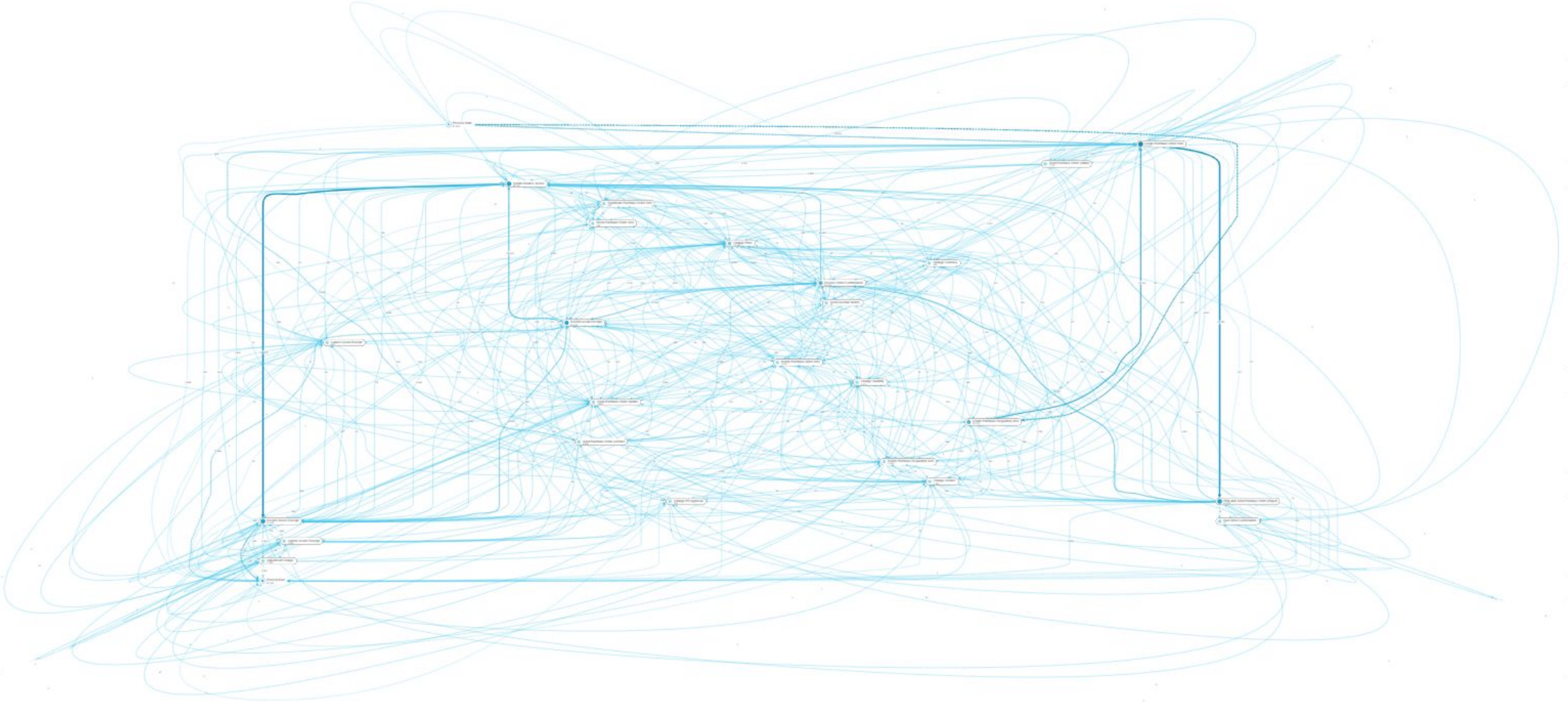
Purchase order items by Line

Document Number	Item Number	Material no.	Material description	N
0000233436	70010	M-05	Flatscreen LE 50 P	
0000023344	00030	100-510	Ball bearing	

Happy Path


Process Start
67,023
36,001
Create Purchase Requisition Item
36,272
31,141
Create Purchase Order Item
67,023
47,365
Print and Send Purchase Order (Paper)
47,478
19,117
Receive Order Confirmation
34,700
15,399
Record Goods Receipt
45,741
23,612
Vendor creates Invoice
49,710
43,979
Record Invoice Receipt
47,438
29,714
Process End
67,023

Real World Process



Approaches

1. Markov Chains

   Marks a sequence as anomalous if a most likely path is not being followed by the input example. With respect to the business process data, apart from sequences which do not follow an ideal path, sequences that follow a desired path but have an extreme value for a KPI e.g.a high throughput time, should also be reported to the user. Using Markov chains will not classify such sequences as anomalies since they follow a desired path.

2. Clustering - applied K-means

   The clusters obtained were showed
   similarities between a set of KPIs. But we could not identify a cluster which showed deviations with respect to all/maximum set of KPIs and clearly deviates from the group.

3. Path Sequence / Network Analysis - similar to genetic encoding and speciation used in NAS scenarios - inspired by speciation through a distance metric - this metric should take into account anomaly behavior such as loops, change-activities and cancelation

```
FILTER "LFA1"."NAME1" = 'Unisono AG';
FILTER "EKPO"."MATKL_TEXT" = 'Tools';
FILTER "EKKO"."EKGRP" = 'R30';
TABLE(  "EKPO"."MATKL_TEXT" AS "MaterialName" ,
        "EKKO"."EKGRP" AS "PurchasingGroup" ,
        ROUND_MONTH("_CEL_P2P_ACTIVITIES"."EVENTTIME") AS "Timestamp" ,
        "EKPO"."_CASE_KEY" AS "CaseId" ,
        SUM(CASE WHEN "_CEL_P2P_ACTIVITIES"."ACTIVITY_EN" LIKE 'Change%'
        THEN 1 ELSE 0 END) AS "NumChangeActivity" ,
        SUM(CASE WHEN ISNULL("_CEL_P2P_ACTIVITIES"."USER_TYPE") = 1 THEN NULL
                 WHEN "_CEL_P2P_ACTIVITIES"."USER_TYPE" IN ('B')
                 THEN 0.0 ELSE 1.0 END) AS "ManualCount" ,
        SUM(CASE WHEN "_CEL_P2P_ACTIVITIES"."ACTIVITY_EN" LIKE 'Delete%' THEN
1
                 WHEN "_CEL_P2P_ACTIVITIES"."ACTIVITY_EN" LIKE 'Cancel%'
                 THEN 1 ELSE 0 END) AS "DelCancelCount" ,
        COUNT("_CEL_P2P_ACTIVITIES"."ACTIVITY_EN") AS "TotalActivity" ,
        AVG(CALC_THROUGHPUT(ALL_OCCURRENCE['Process Start'] TO
            ALL_OCCURRENCE['Process End'],   REMAP_TIMESTAMPS
            ("_CEL_P2P_ACTIVITIES"."EVENTTIME", DAYS))) AS "ThrDays" )
```

```
FILTER "EKPO"."_CASE_KEY" in (cases');
TABLE(  "EKPO"."_CASE_KEY" AS "CaseId" ,
        "_CEL_P2P_ACTIVITIES"."ACTIVITY_EN" AS "Activities" ,
        ROUND_MONTH("_CEL_P2P_ACTIVITIES"."EVENTTIME") AS "Timestamp"
,
        CASE WHEN COUNT("_CEL_P2P_ACTIVITIES"."ACTIVITY_EN") > 1
             THEN COUNT("_CEL_P2P_ACTIVITIES"."ACTIVITY_EN")
             ELSE 0 END AS "RepeatedCount" )

FILTER "EKPO"."_CASE_KEY" in ('anomalous_cases');
TABLE(  "EKPO"."_CASE_KEY" AS "CaseId" ,
        VARIANT("_CEL_P2P_ACTIVITIES"."ACTIVITY_EN") AS "Variant" )
```

Query: What was unusual for Vendor SMP in the last week of Dec 2008?

KPI: Manual Updates

	Timestamp	TotalActivity
0	2007-12-31	1
1	2008-06-16	2
2	2008-12-29	2
3	2009-05-11	2
4	2009-10-19	4
5	2009-11-02	3
6	2009-11-16	3
7	2009-12-14	1
8	2009-12-28	1
9	2010-01-11	1
10	2010-01-18	1

```
Vendor: SMP
Queried Week: 2008-12-29
The queried Week contains less than 5 total activities which are below the threshold required for further processing.
The table displays the Weeks with number of activities less than 5.
Please examine why there are few activities executed by the Vendor in these Weeks.
```

Combinatory Categorial Grammar

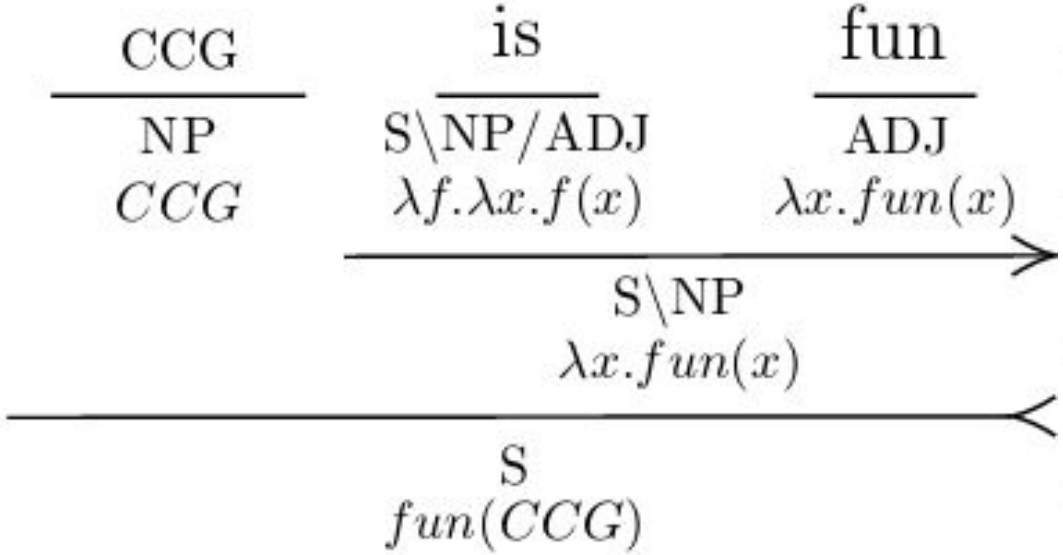
Lexicons

$$\text{CCG} \vdash NP : CCG$$

$$\text{is} \vdash (S\backslash NP)/ADJ : \ \lambda f.\lambda x.f(x)$$

$$\text{fun} \vdash ADJ : \lambda x.fun(x)$$

Parsing



Reproduced from Artzi et al. 2013

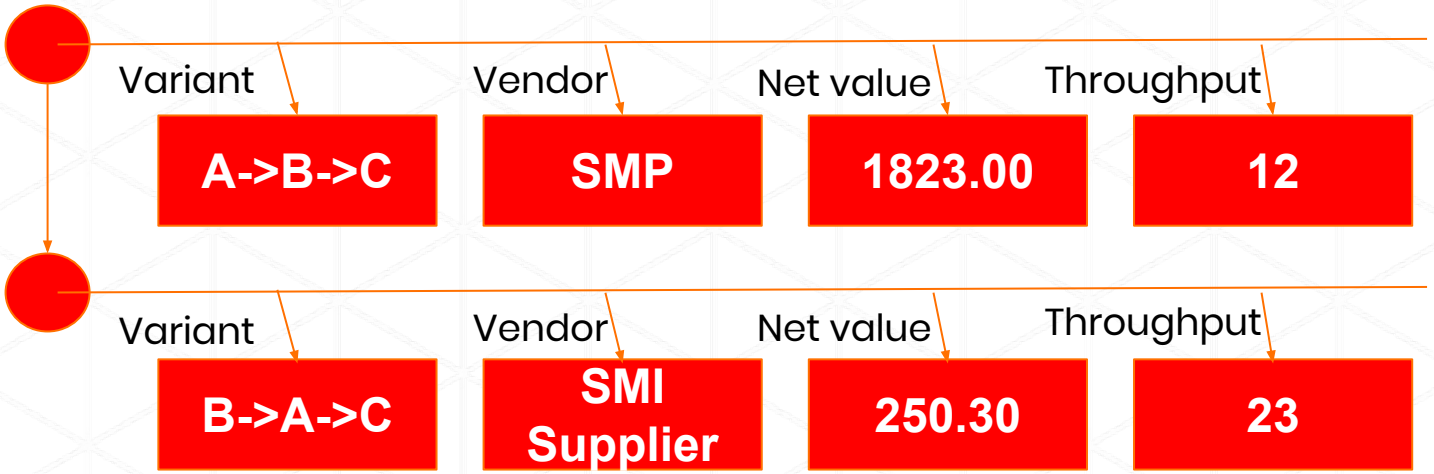
WIKI-Table

Compositional Semantic Parsing on Semi-Structured Tables

De-normalize table

Variant	Vendor	Net order value	Throughput-Time
A->B->C	SMP	1823.00	12
B->A->C	SMI Supplier	250.30	23

Generate graph



Adapted from Pasupat, Liang 2015

Libraries

Google APIs

Dialogflow

spaCy