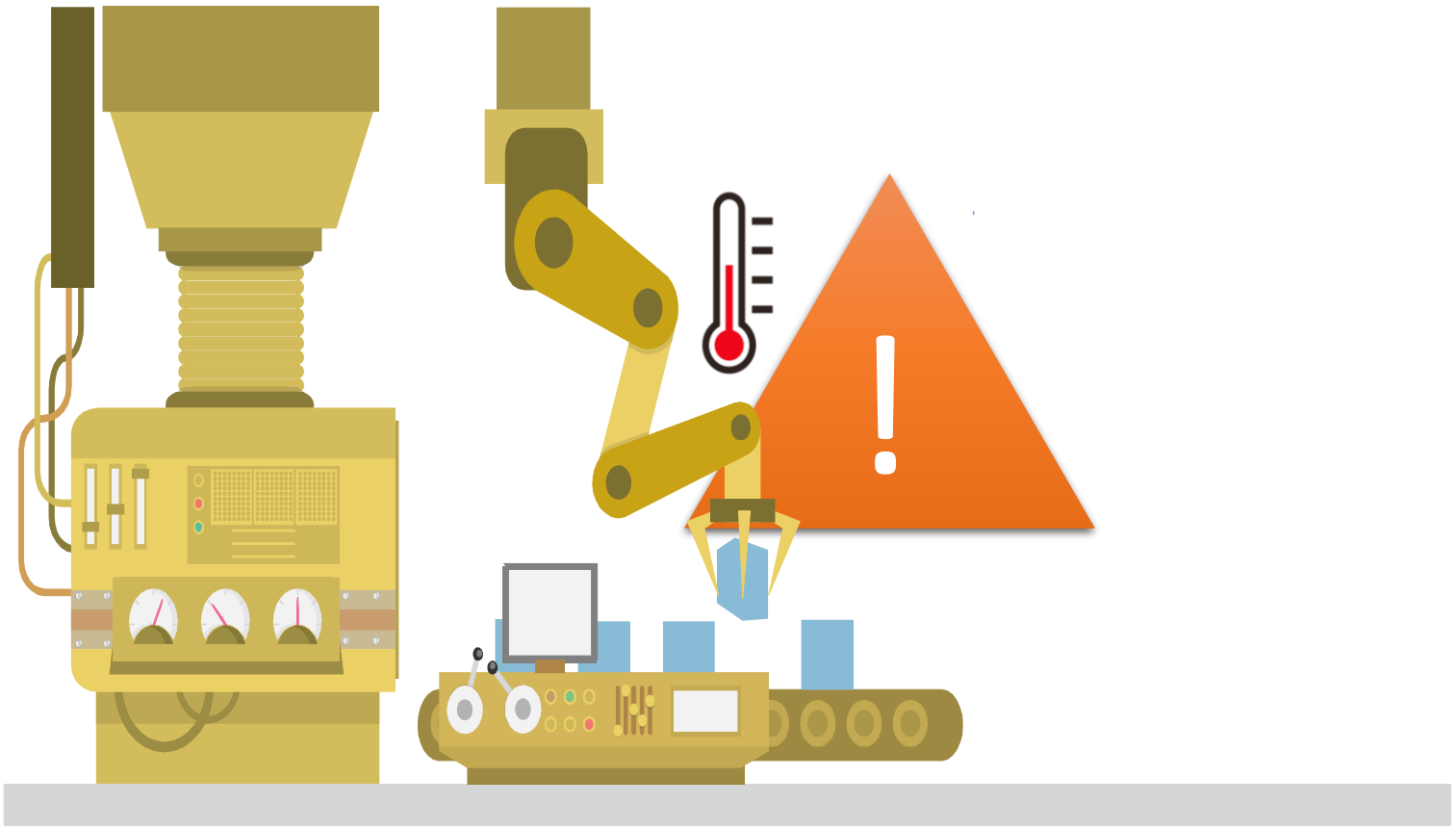
Augmented Business Intelligence

Matteo Francia, Matteo Golfarelli, Stefano Rizzi

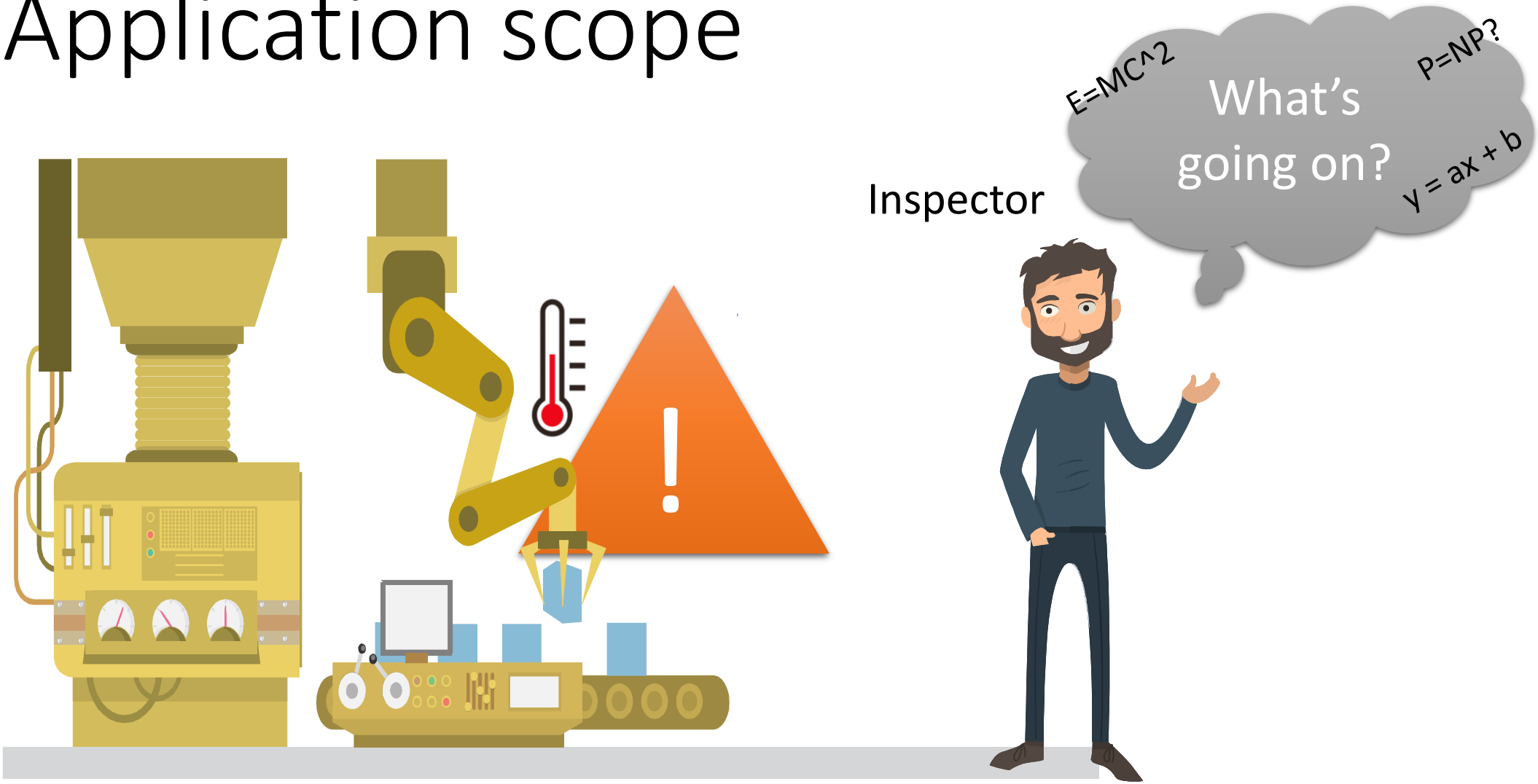
DISI – University of Bologna

{m.francia, matteo.golfarelli, stefano.rizzi} @unibo.it

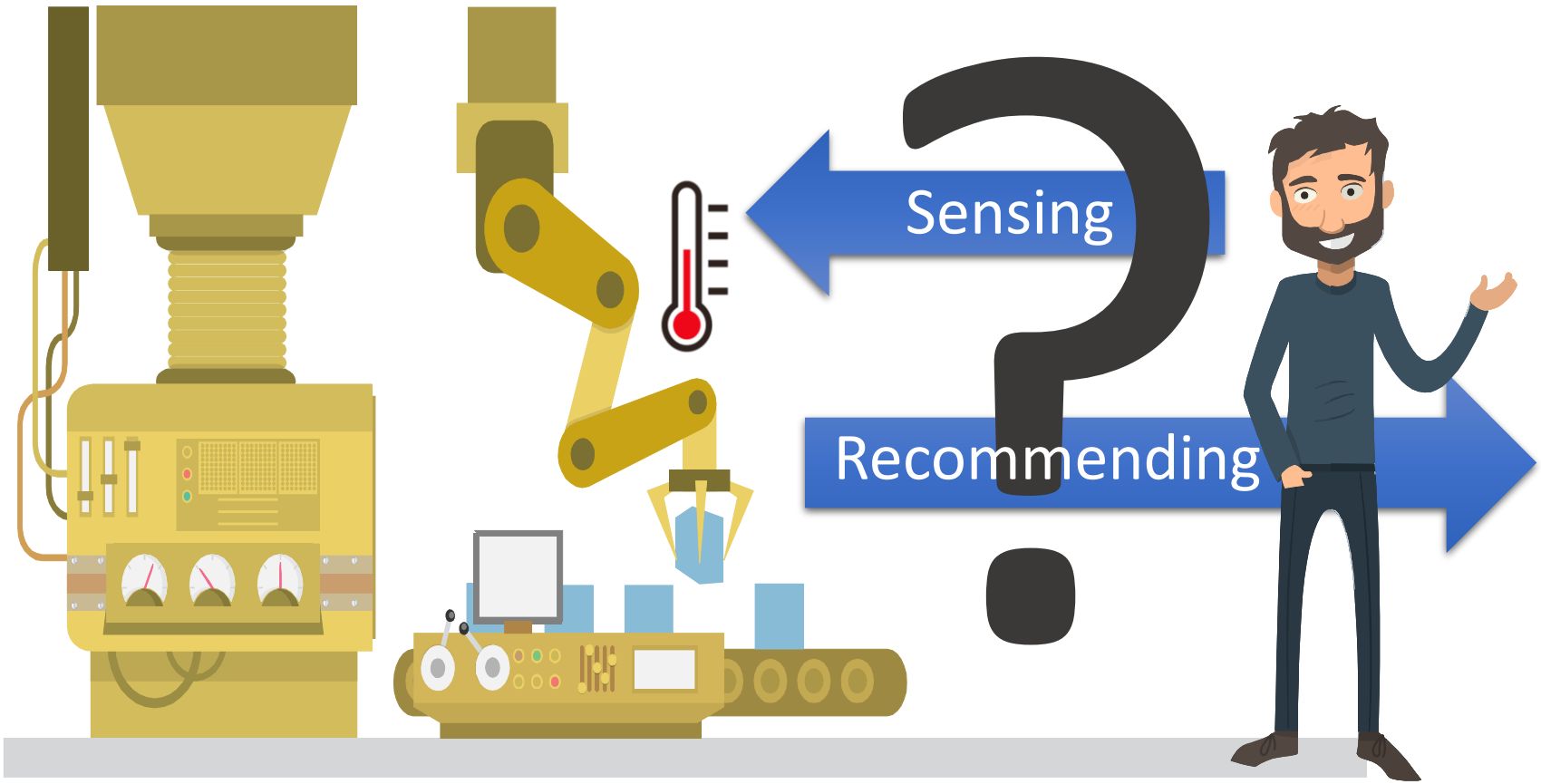
Application scope



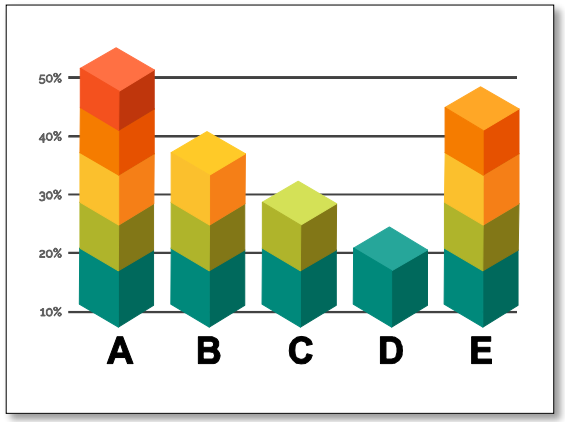
Application scope



Application scope



Analytical report



We have data!

- Internet of Things
- Digital twin [1]

[1] Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., & Sui, F. (2018). Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, *94*(9-12), 3563-3576.

Augmented Business Intelligence

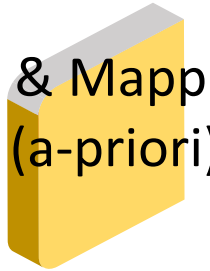
A-BI: a 3D-marriage

- Augmented Reality
- Business Intelligence
- Recommendation

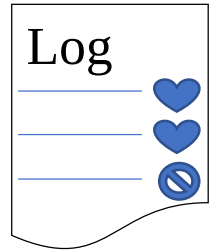
A-BI: Overview

Data Sources

Augmented Reality (real-time)



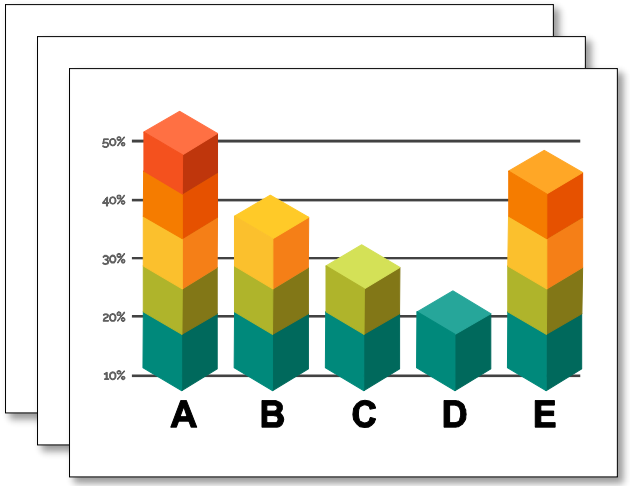
DM & Mappings (a-priori)



Query Log (user exp.)

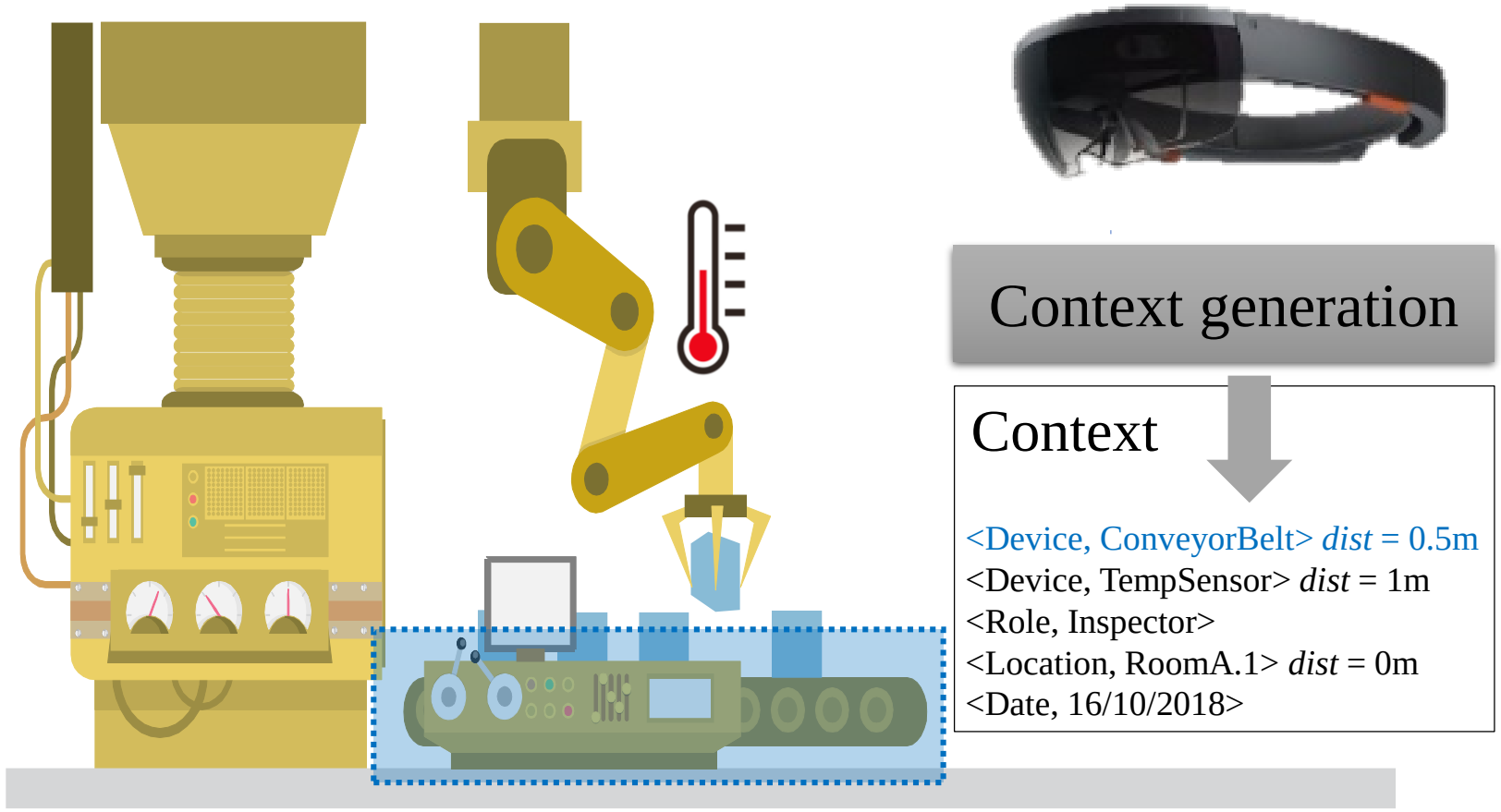
Augmented Business Intelligence

Outputs



OLAP reports

A-BI: Augmented Reality



Context generation

Context

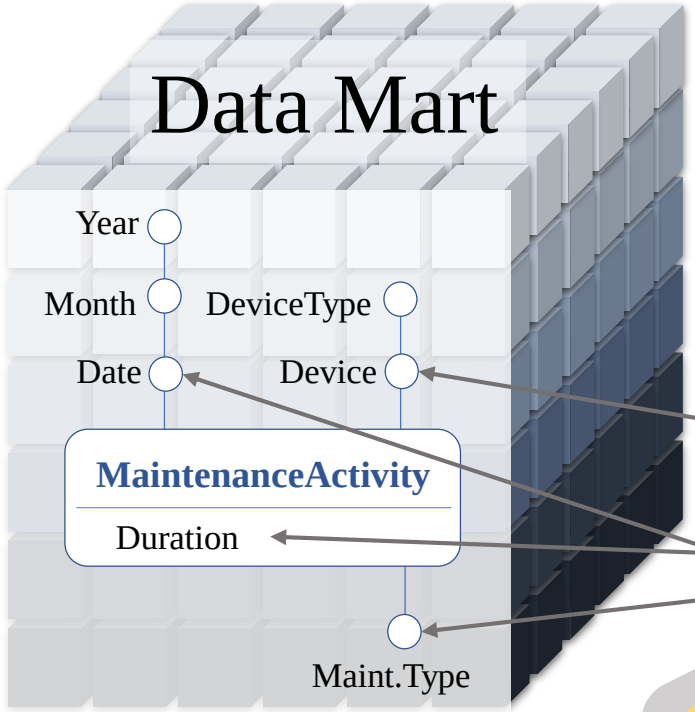
<Device, ConveyorBelt> *dist* = 0.5m
<Device, TempSensor> *dist* = 1m
<Role, Inspector>
<Location, RoomA.1> *dist* = 0m
<Date, 16/10/2018>

- Sensing augmented environments [2]
  - Real-time information
- Interaction
  - Engagement [3]
- Constrained visualization
  - i.e., cardinality constraint

[2] Croatti, A., & Ricci, A. (2017, April). Towards the web of augmented things. In 2017 IEEE International Conference on Software Architecture Workshops (ICSAW)
[3] Su, Y. C., & Grauman, K. (2016, October). Detecting engagement in egocentric video. In European Conference on Computer Vision

A-BI: Business Intelligence

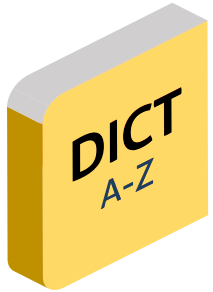
- Data dictionary
  - What do we recognize?
  - Context: subset of data dictionary entries
- Mappings to md-elements
  - A-priori interest
- OLAP
  - Report generation



Context generation

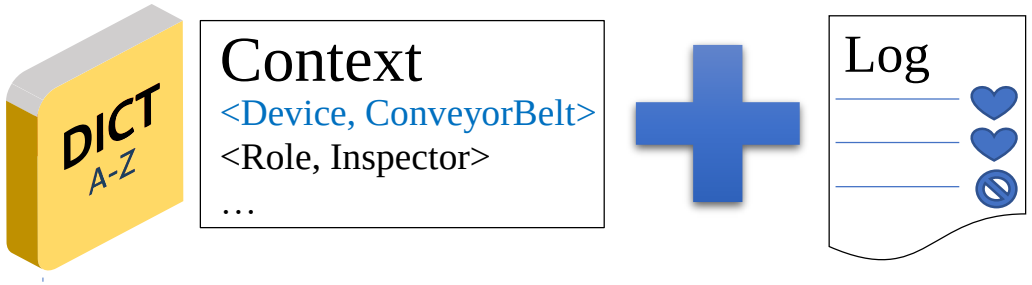
Context

<Device, ConveyorBelt> *dist* = 0.5m
<Device, TempSensor> *dist* = 1m
<Role, Inspector>
<Location, RoomA.1> *dist* = 0m
<Date, 16/10/2018>



A-BI: Recommendation

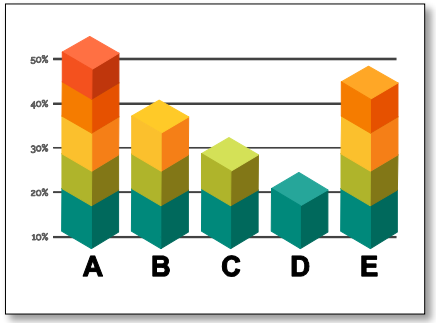
1. Get the context
   - Context T over data dictionary
   - Follow (a-priori) mappings…
   - … Project T to image I of md-elements



2. Add the log L
   - Get queries with positive feedback from similar contexts
   - Enrich I to I^* with «unperceived» elements from T

3. Get the queries

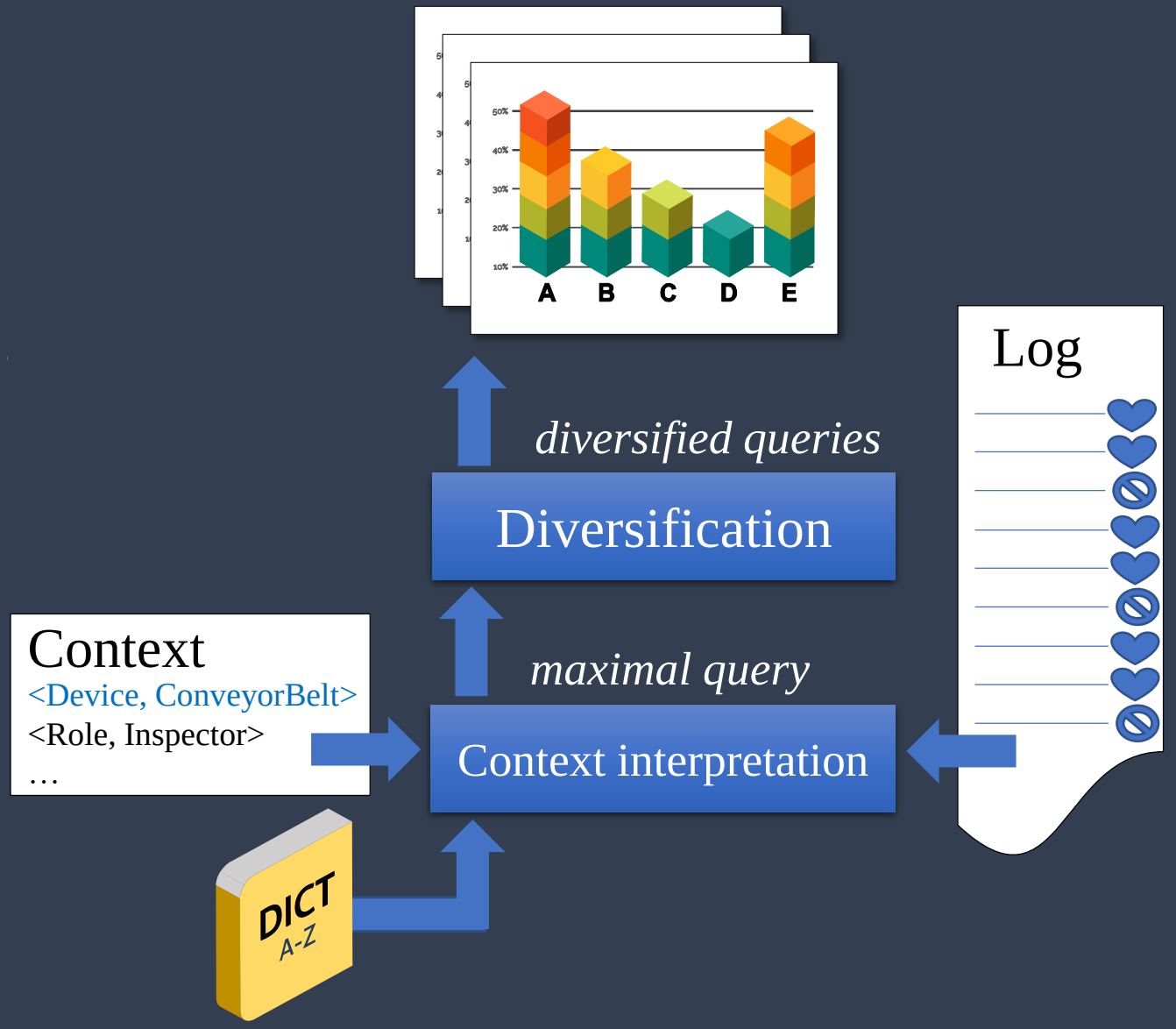
   Directly translate I^* into a well formed query
   - High cardinality I^* = hardly interpretable «monster query»
   - Single query, no diversification



A-BI

A two-step approach:

- Context interpretation
- Diversification



A-BI: Context Interpretation

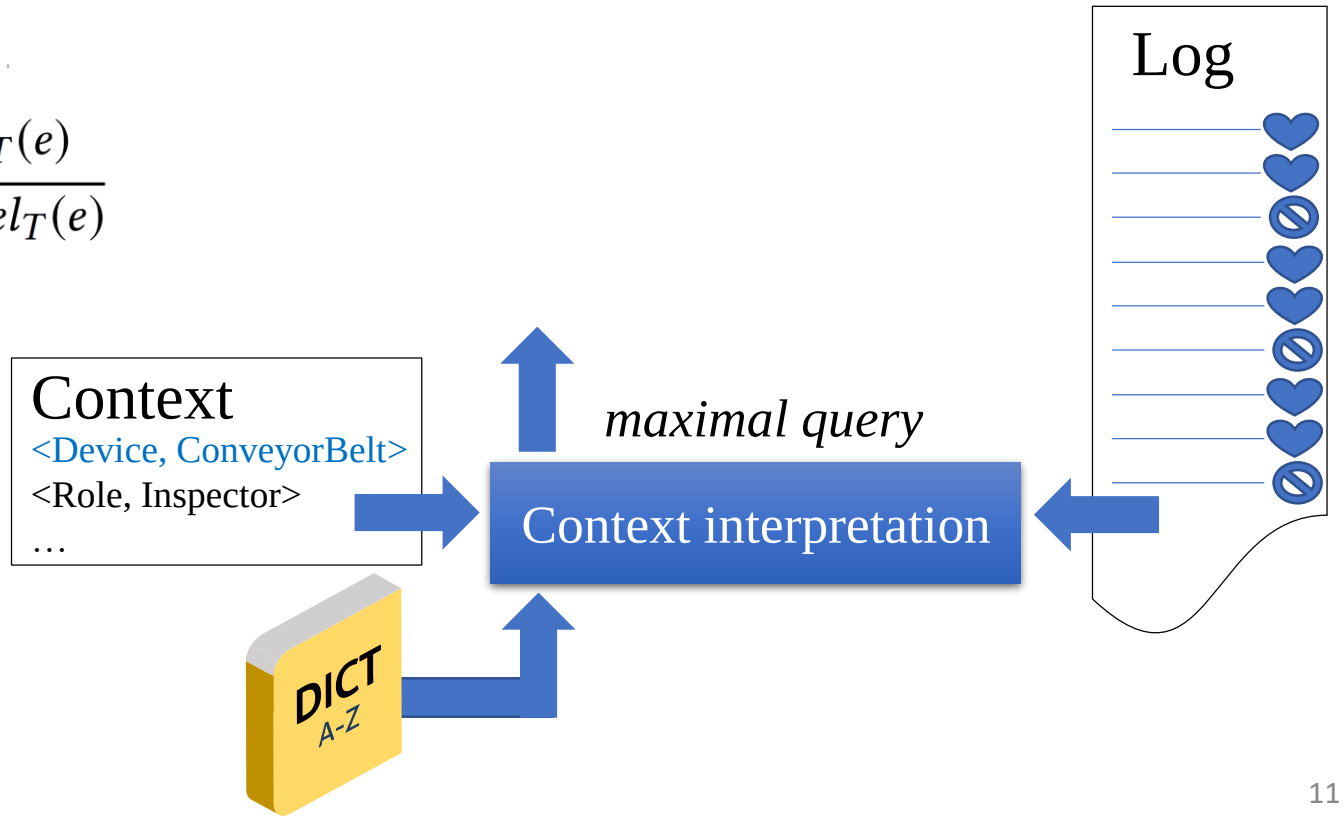
- md-element relevance
  - Context weight
  - Mapping weight
  - Relevance over log

$$rel_T(e) = \begin{cases} \sum_{d \in T \mid e \in \mu(d)} w(T,d) \cdot w(d,e), \text{ if } L_T(e) = \varnothing \\ \sum_{d \in T \mid e \in \mu(d)} w(T,d) \cdot w(d,e) \cdot \rho_T(L,e), \text{ otherwise} \end{cases}$$

- Query relevance

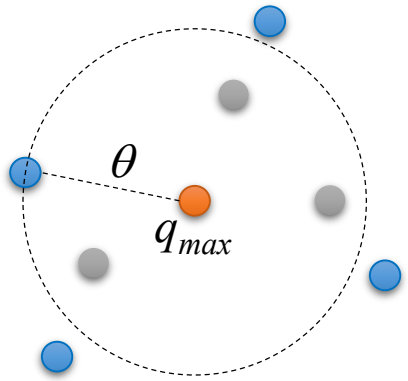
$$rel_T(q) = \frac{\sum_{e \in q} rel_T(e)}{\sum_{e \in I^*_\mu(T)} rel_T(e)}$$

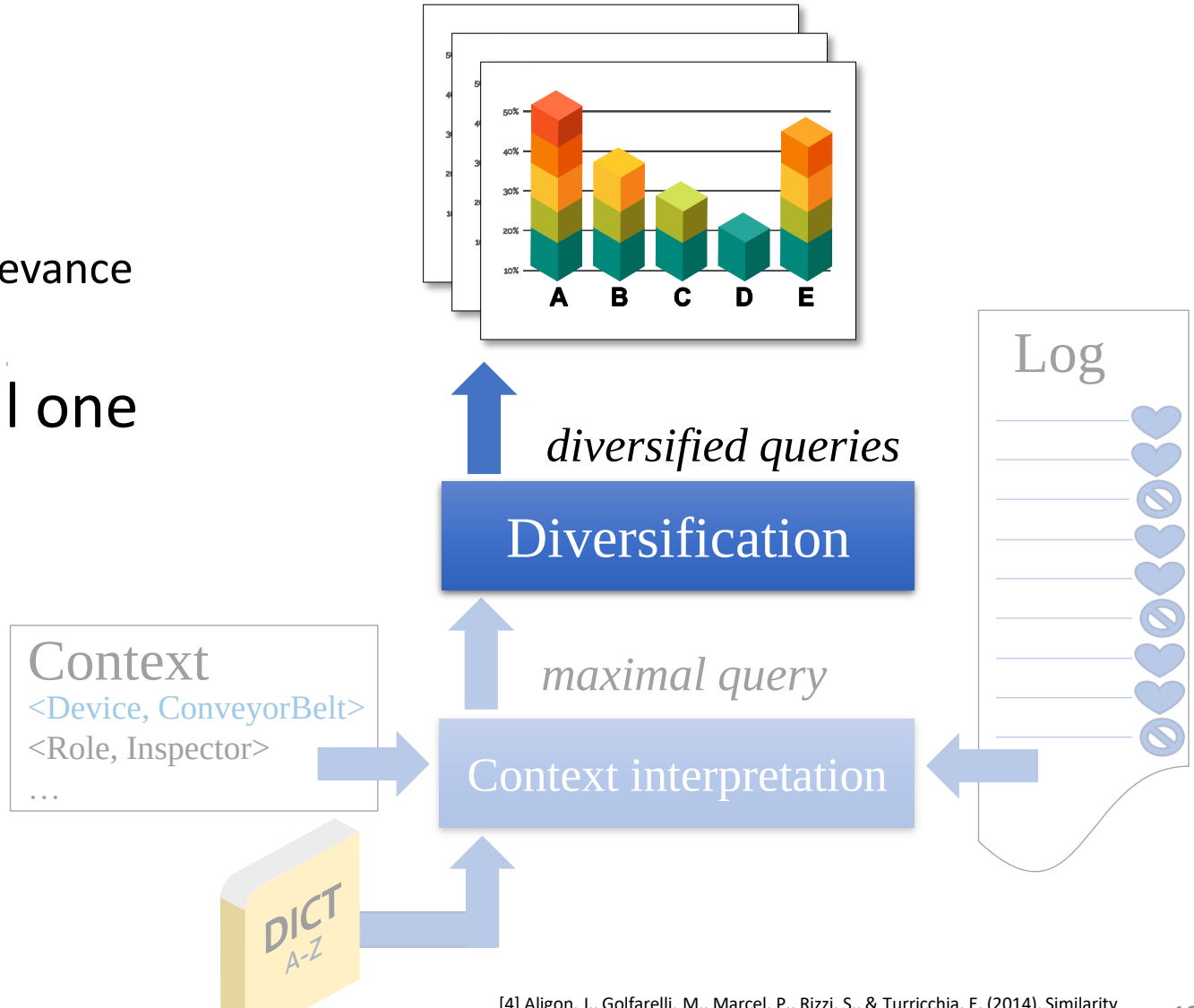
- Maximal query
  - Most relevant query enforcing cardinality constraint
  - = Knapsack Problem
    - Draw most relevant DM-elements
    - s.t. query cardinality is below threshold



A-BI: Diversification

- Diversification
  - Different flavors of same information
  - = Top-N queries maximizing diversity and relevance
- Generate queries from the maximal one
  - Operators: rollup/drill/slice
  - Query similarity sim (with $div = 1 - sim$) [4]
  - θ = amount of diversification

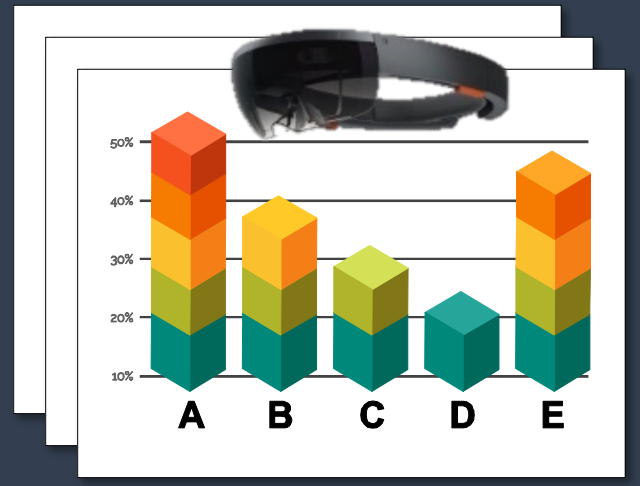




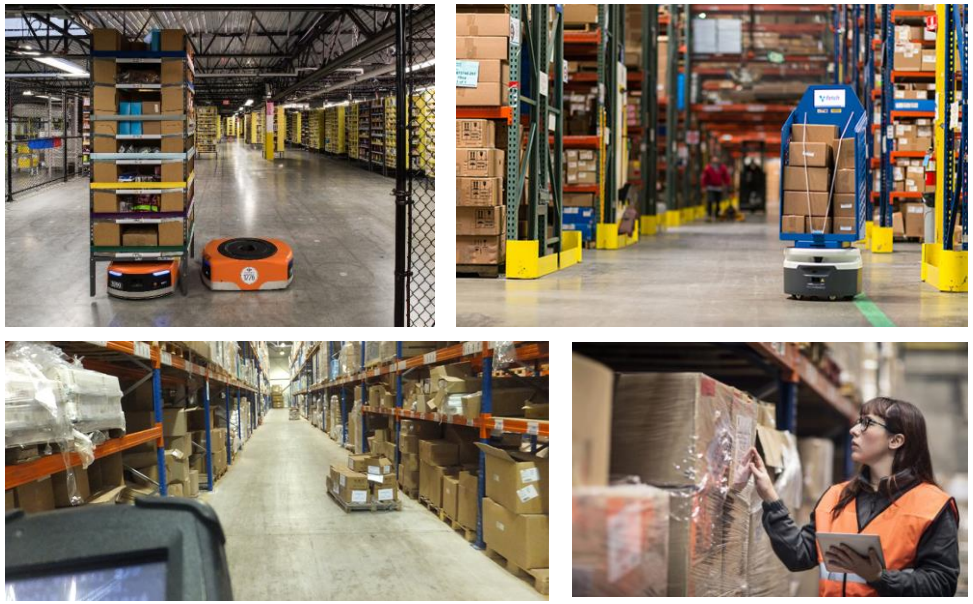
[4] Aligon, J., Golfarelli, M., Marcel, P., Rizzi, S., & Turricchia, E. (2014). Similarity measures for OLAP sessions. Knowledge and information systems, 39(2), 463-489

Evaluation

- Effectiveness
- (Near) Real-time



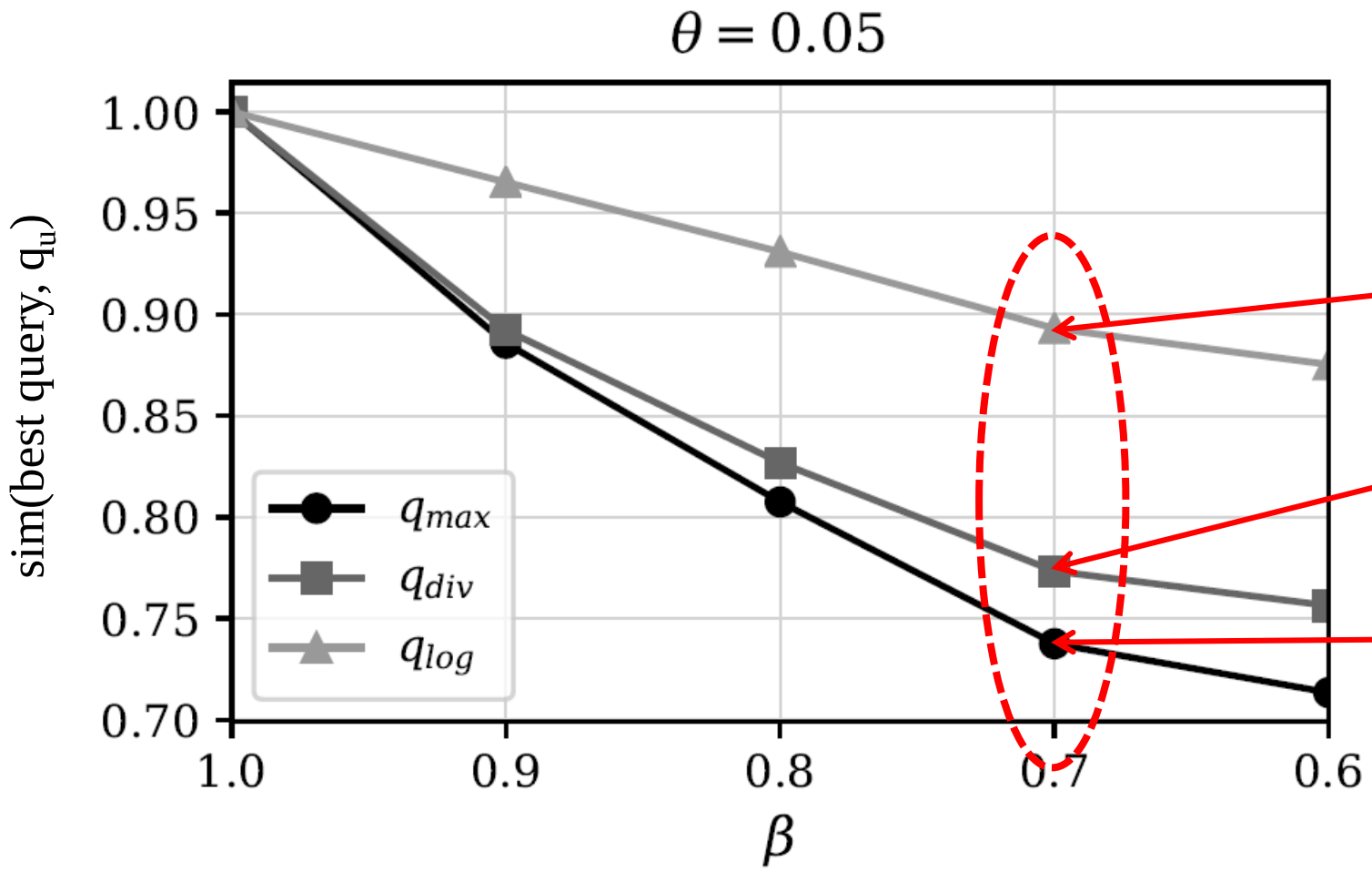
A-BI: Test setup



Examples of context seeds

- Cube
  - 5 linear hierarchies, 5 levels each
  - Maximum cardinality 10^9
  - Dictionary with one entry for md-element
  - One-to-one mappings (entry → md-element)
  - Random context and mapping weights
- Simulate user moving through a factory
  - In 10 different rooms (i.e., 10 context seeds)
  - 5 to 15 recognized entities
- Simulate multiple visits to rooms
  - Generate seed variations

A-BI: Effectiveness



$|T| = 10, N = 4$

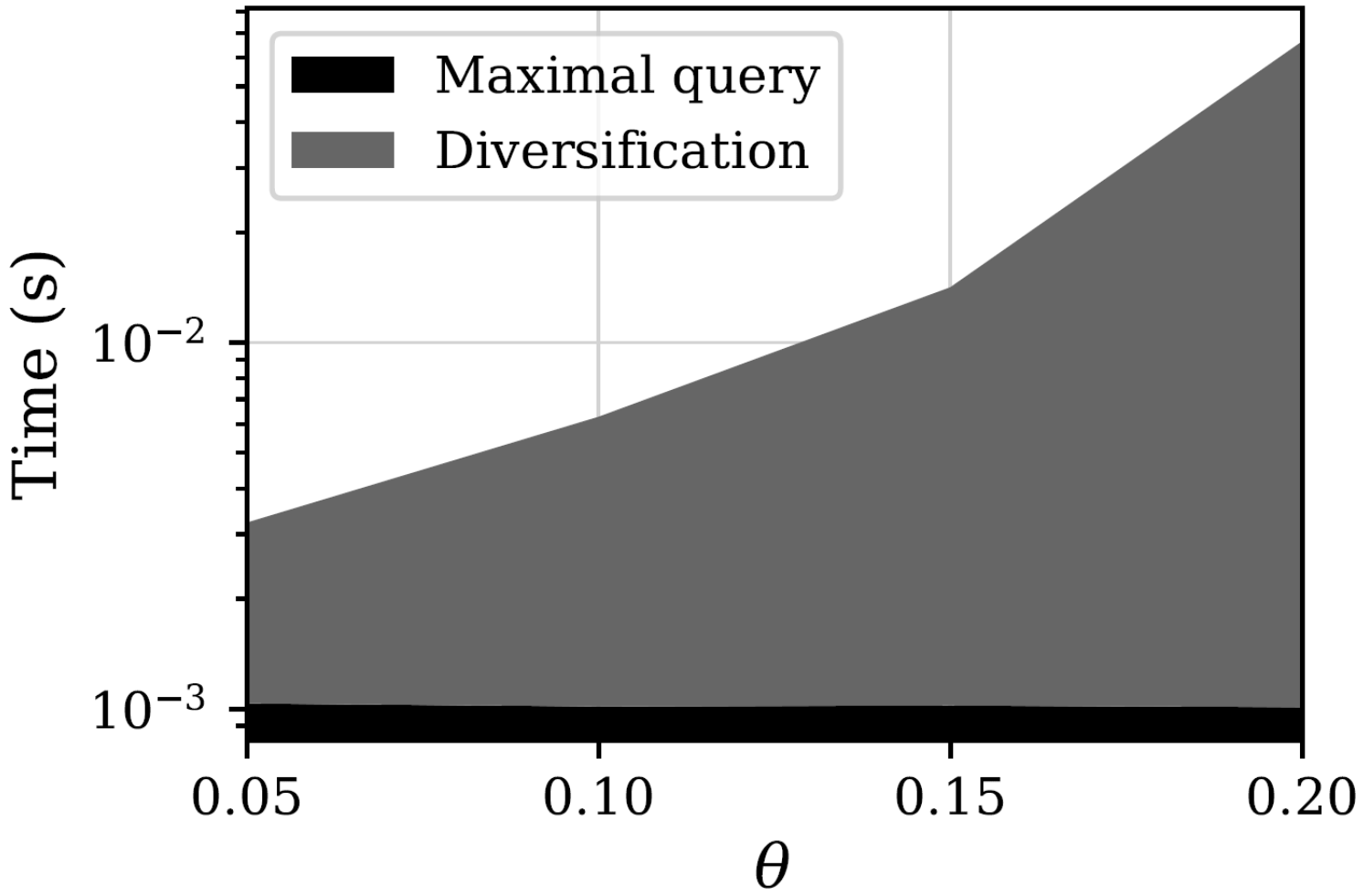
Best query (with log, 1 visit)
After 2 visits: 0.95, 4 visits: 0.98

Best query (no log)

Maximal query

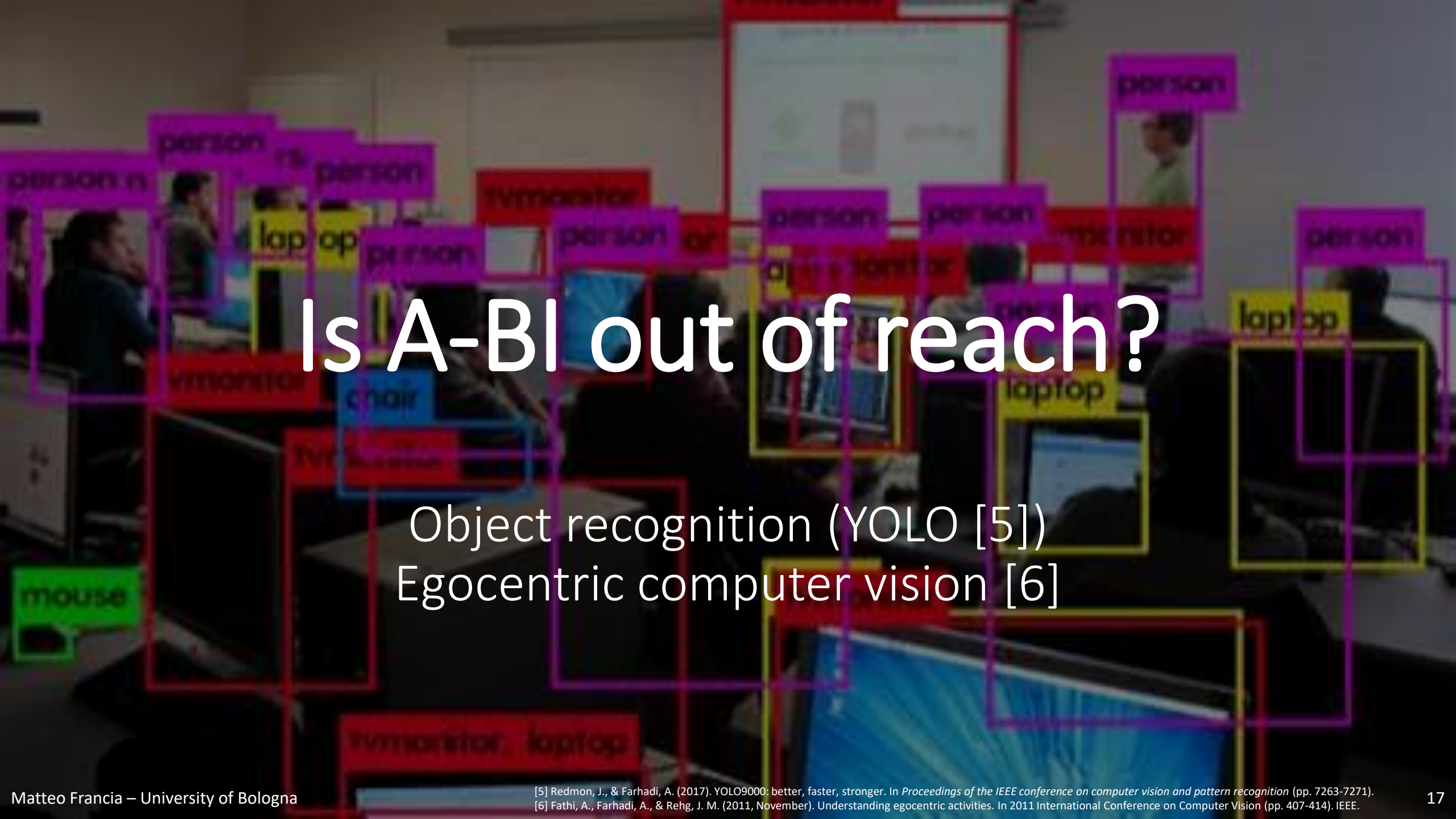
β = target similarity between q_u and q_{max}

A-BI: Efficiency



- Time required to recommend a query set
- Query execution is then demanded to DW system

θ = diversity threshold



Is A-BI out of reach?

Object recognition (YOLO [5])
Egocentric computer vision [6]

[5] Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7263-7271).
[6] Fathi, A., Farhadi, A., & Rehg, J. M. (2011, November). Understanding egocentric activities. In 2011 International Conference on Computer Vision (pp. 407-414). IEEE.

Work in progress: relevance of groups

- Up to now
  - Relevance of single md-elements
  - Recommendation address all the elements together

Proposal: Relevance is about groups of md-elements

- Element a is relevant with b but not with c
- Definition of group relevance rel
- Review formulation to provide recommendation related to groups

- Given $rel(\{\text{Maint.Type}\}) = 1$ and $rel(\{\text{Duration}\}) = 1$
  - $rel(\{\text{Maint.Type, Duration}\}) = 2.5$
  - $rel(\{\text{Device, Month}\}) > rel(\{\text{Device, Date}\})$

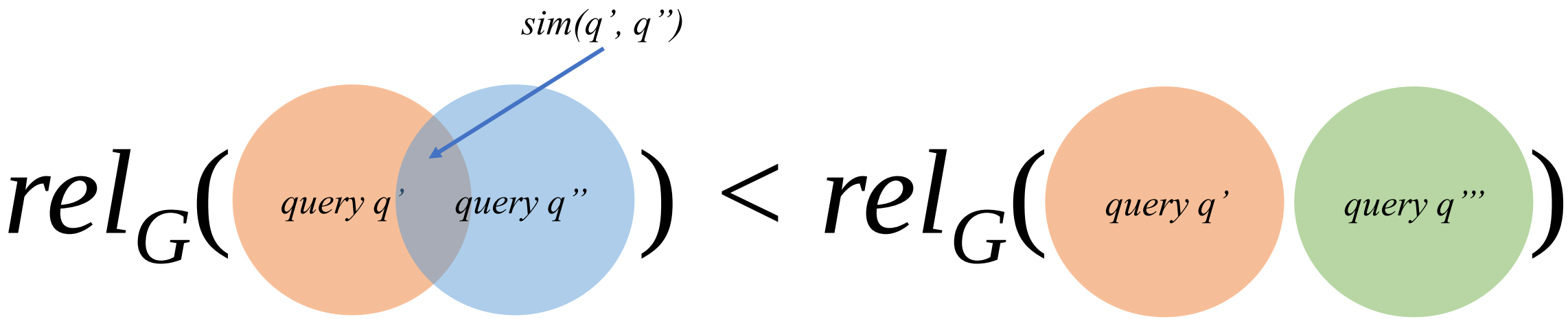
Work in progress: query generation

- Up to now
  - Recommendation as a two-step approach

NEW

Proposal: Optimal formulation for query generation

- Single-step formulation inspired by mutual information
- Minimize amount of information about one query obtained through other query(s)
- Definition and maximization of global relevance rel_G
  - Overlapping queries (i.e., similar queries) → high mutual information



Conclusion

Augmented Business Intelligence

- Recommendation of multi-dimensional analytic reports
- Based on augmented (real) environments
- Under near-real-time and visualization constraints

- Vision
  - Analytics in Health-care
  - Conversational BI

Thanks