32nd Annual INCOSE international symposium

hybrid event

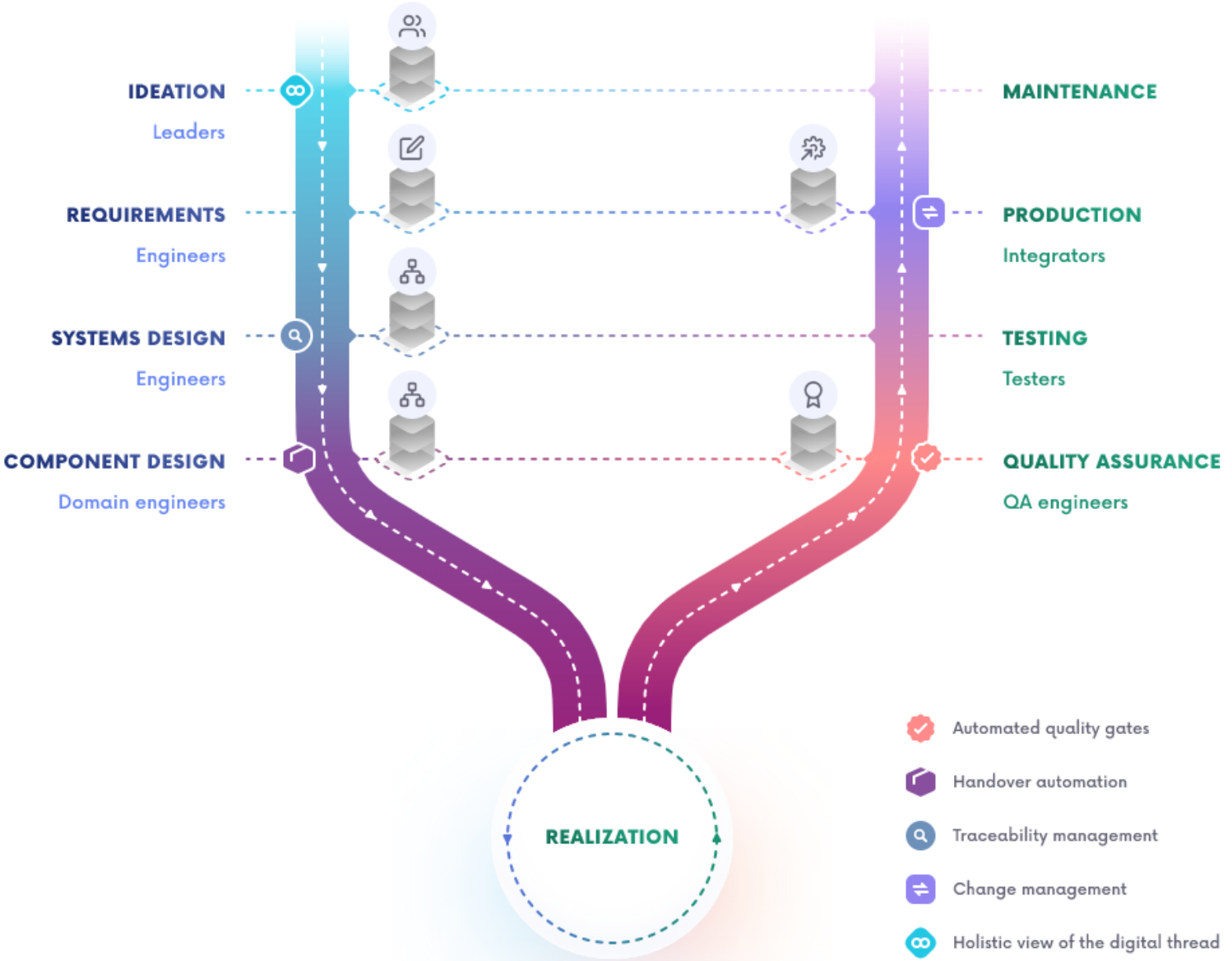
Detroit, MI, USA
June 25 - 30, 2022

Authors: Géza Kulcsár (IncQuery Group), Gaëtan Pruvost (Thales Global Services), Olivier Constant (TGS), István Ráth (IQG), Máté Füzesi (IQG), Dénes Harmath (IQG)

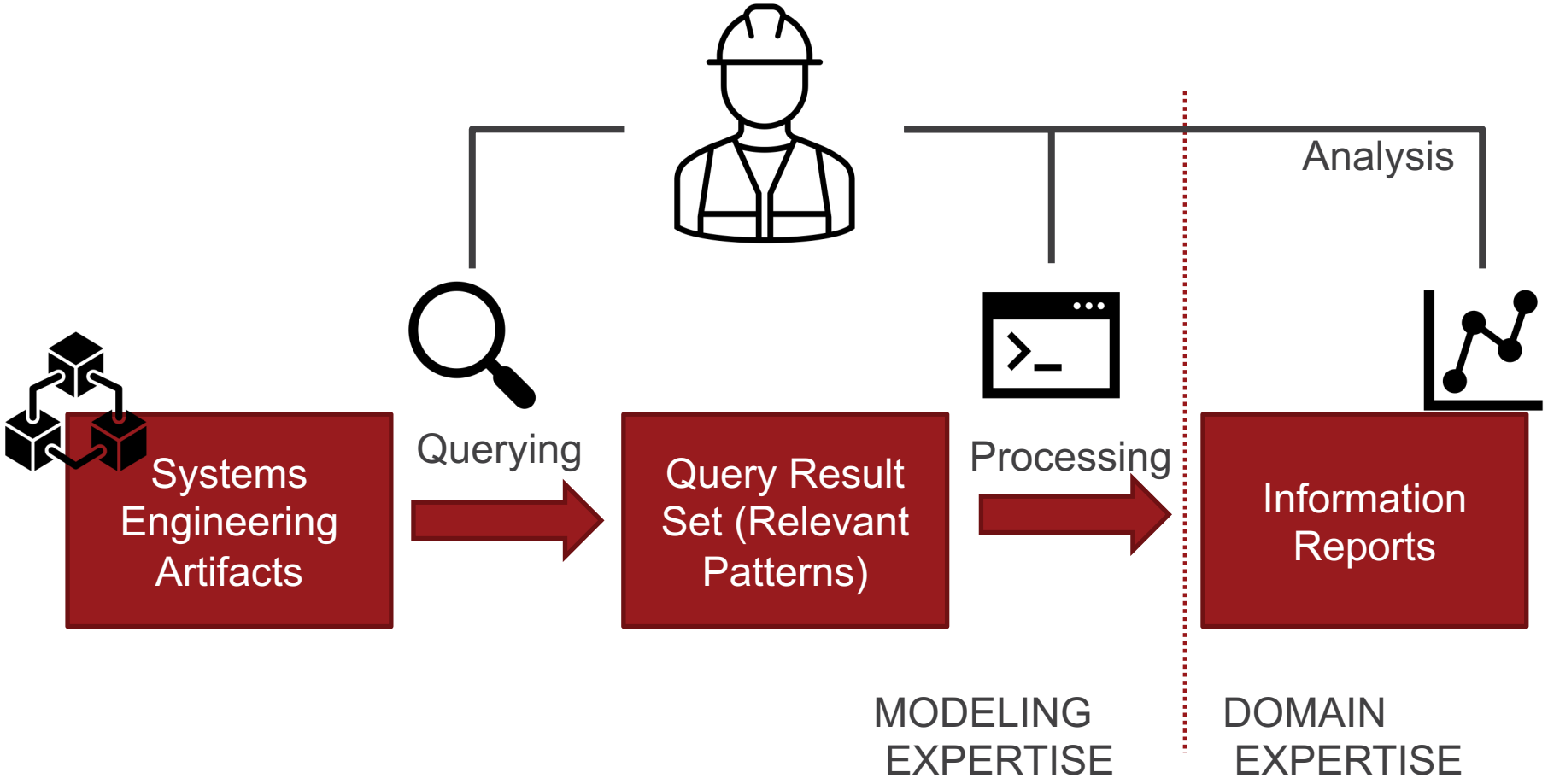
Presenter: *Ákos Horváth IncQuery Group*
akos.horvath@incquerylabs.com

# Natural Language Understanding of Systems Engineering Artifacts

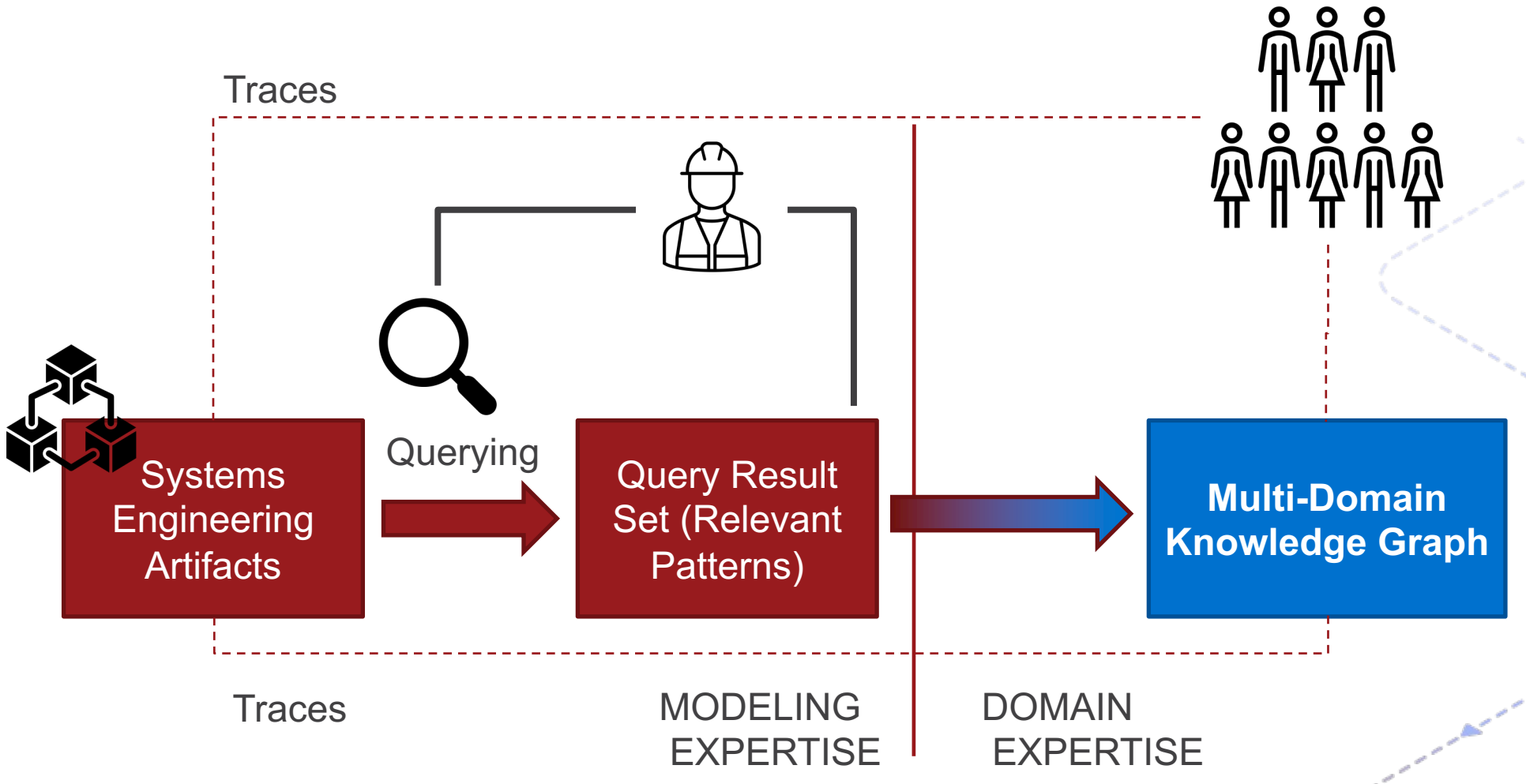
# The Y-Model of Digital Systems Engineering



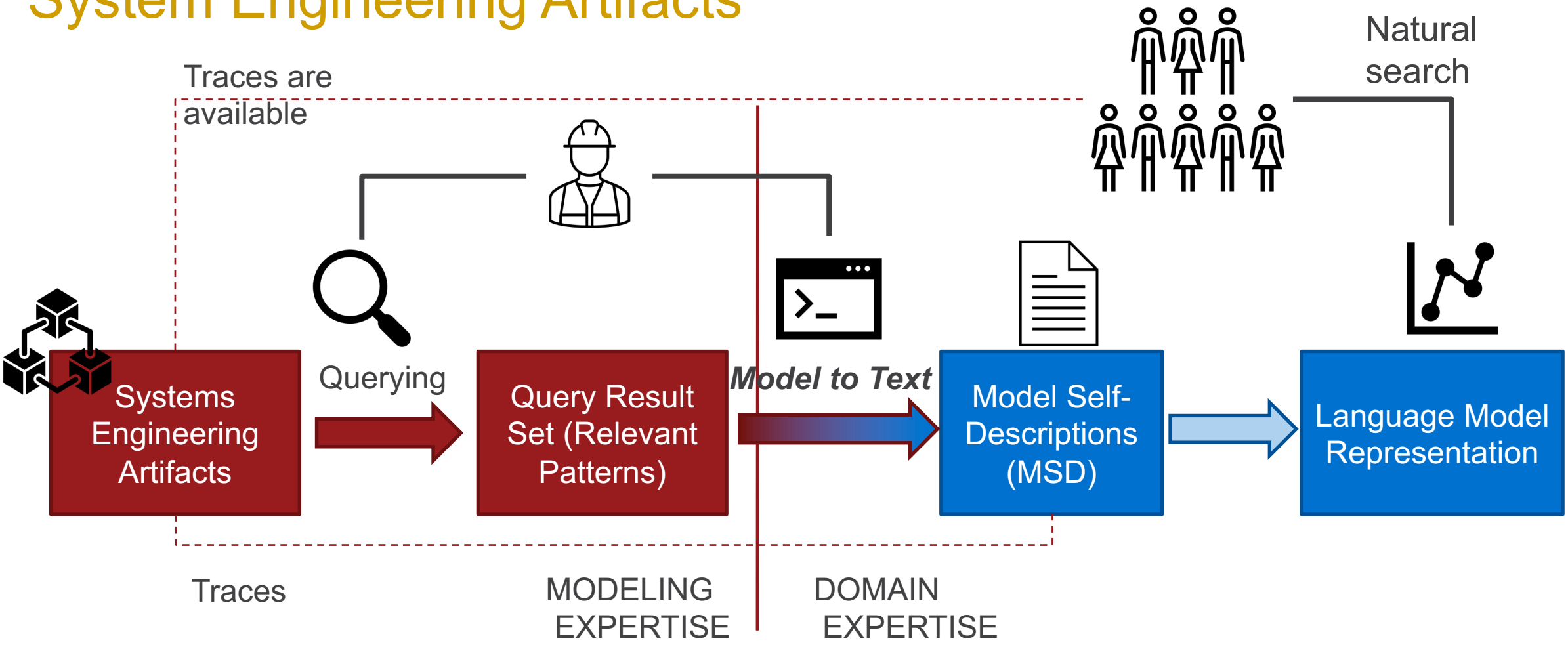
# Systems Model Understanding: Beware, Experts Only



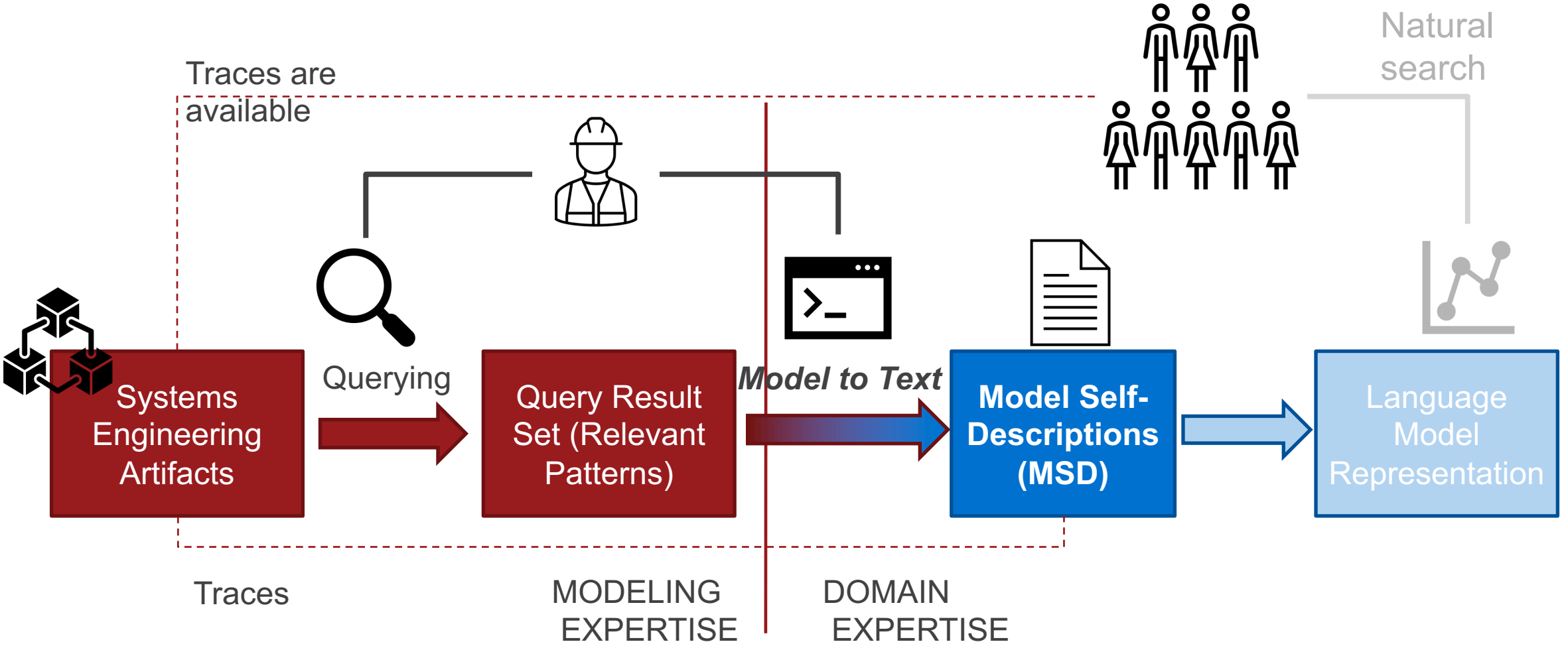
# Vision: Accessible System Model Analysis



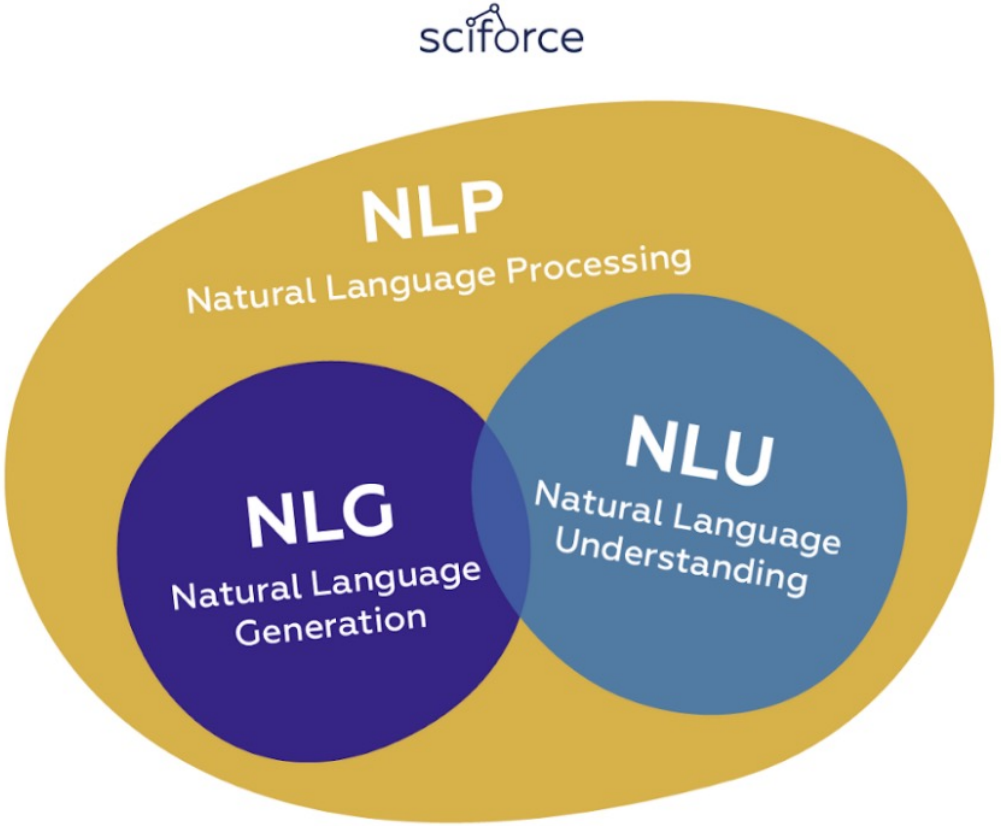

# Overview: Natural Language processing of System Engineering Artifacts



# Contribution 1: Model-Self Descriptions

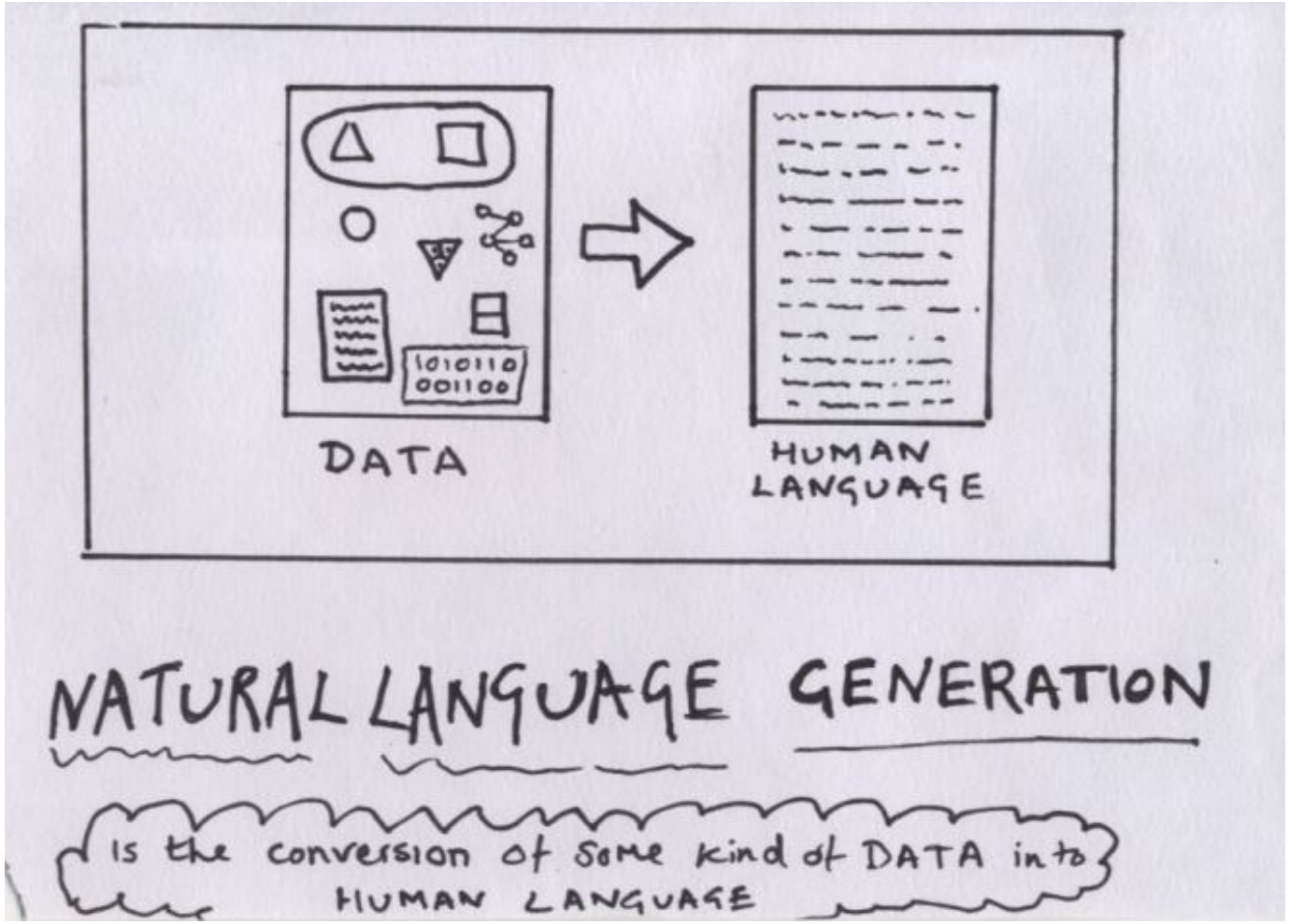


# What is Natural Language Generation?



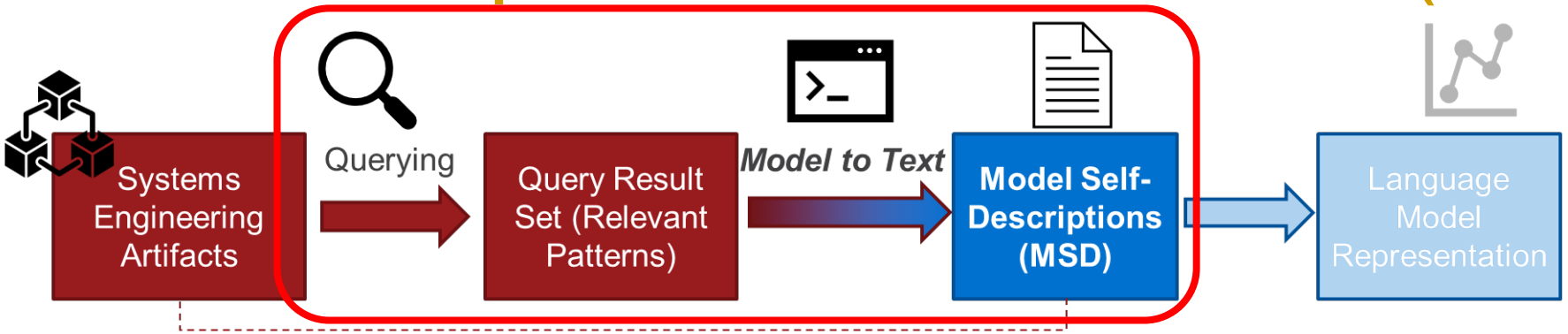

Source: “A Comprehensive Guide to Natural Language Generation.” Medium, Sciforce. 2019

# What is Natural Language Generation?



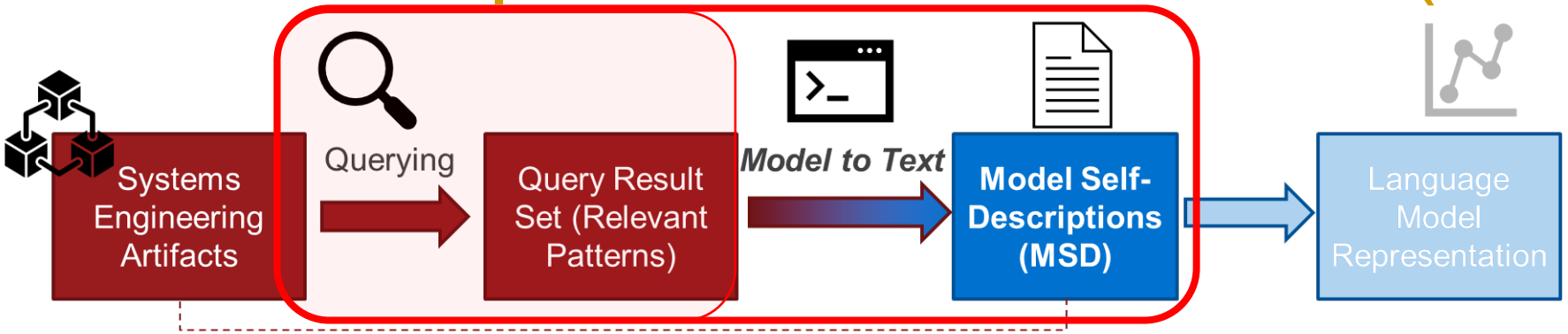

https://bhilluminated.wordpress.com/2017/03/13/nlp-and-nlg/

# Model-Self Descriptions: Models to text (NLG)




- The ongoing revolution in Natural Language Understanding (NLU) results in better solutions for automatically interpreting “general” texts
- That is, if the nature of our input text corresponds to “typical” documents (as found, e.g., on the internet), then we have good chances to have quality results for NLU tasks
- However, sometimes, there is a *specific* meaning to a text, (meant to be) understood by experts only
  - ***Model-self descriptions*** are an intermediate step between NLU algorithms and domain-specific artifacts – here, system models

# Model-Self Descriptions: Models to text (NLG)




- **Current approach:** a query-based information extractor
  - *Traceability is explicitly saved based on query result sets as the input for the text generation*

- **Sample queries**
  - VQL (Viatra Query Language)
    *"look for SysML blocks in a given package"*

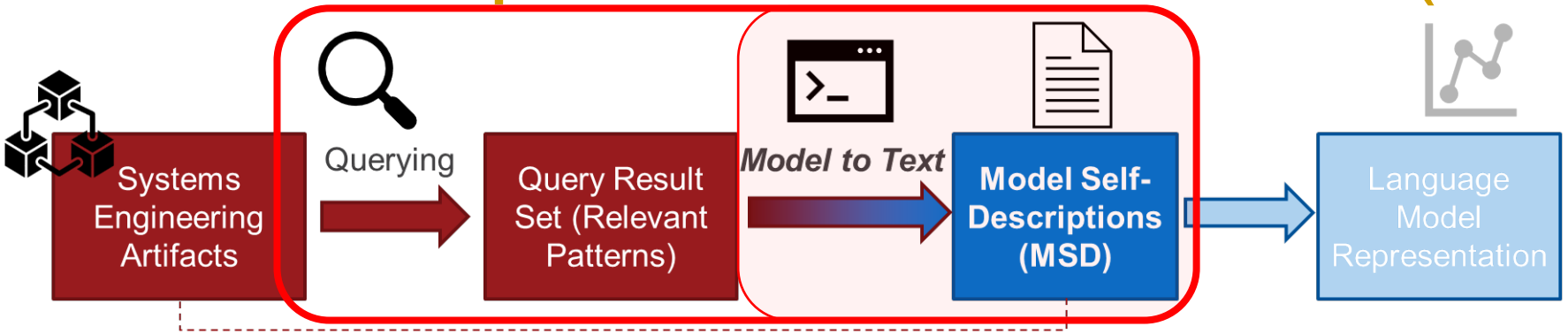
```
pattern blockValueProperties( packageName : java String, block :  Block)

find blockUnderPackages(packageName, block);
```

  - Gremlin (a tool-independent graph query language for the Apache Tinkerpop framework)
    *"get Capella elements with Logical Component or Physical Component as domain type"*

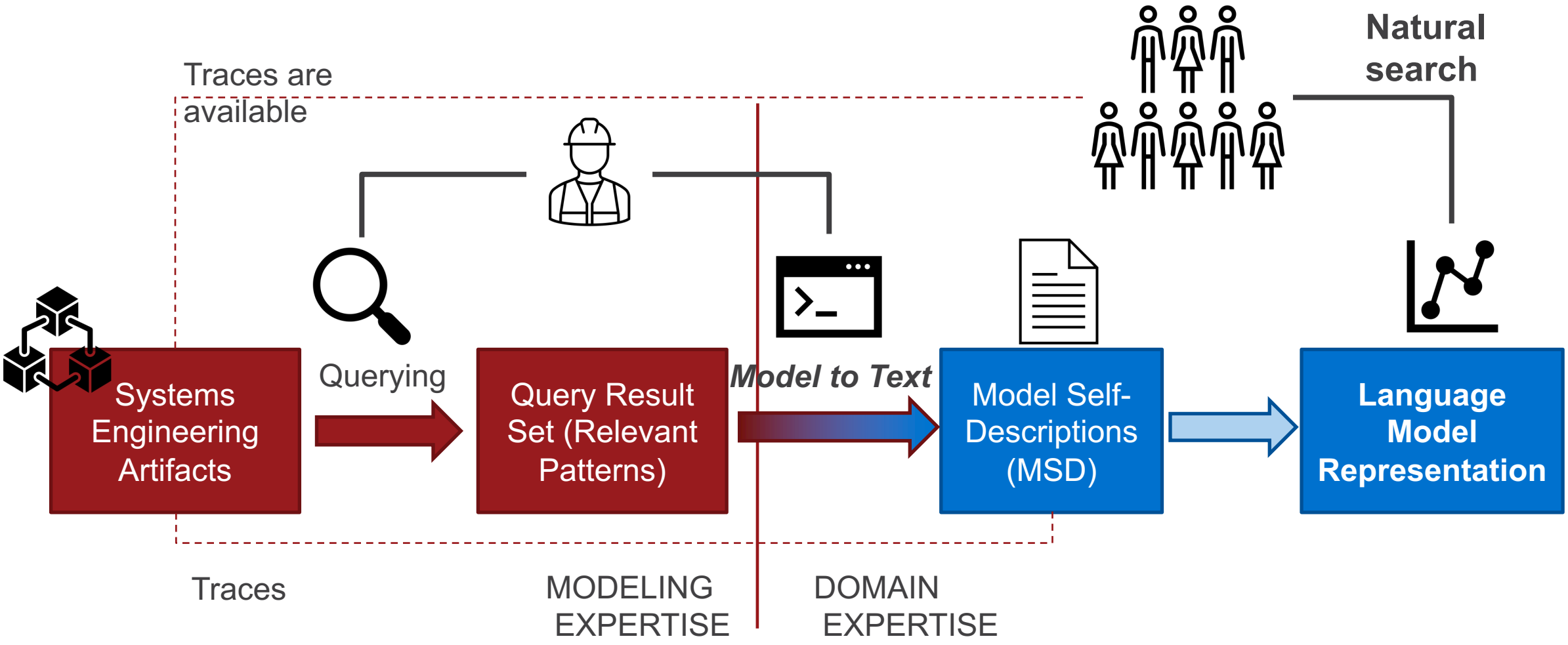
```
g.V().has('domainType', within('LogicalComponent', 'PhysicalComponent'))
```

# Model-Self Descriptions: Models to text (NLG)




- **Current approach:** and a template-driven text generator
  - only presupposing knowledge of SysML but not of the source model or the domain it comes from,
  - For easier understanding: Hybrid-SUV example from OMG
- **Sample generated sentences**
  - *The Environment block does not have any value properties.*
  - *The Environment block has a weather part property with the type of Weather.*
  - *The Weather block is in the HSUVModel package.*
  - *The Transmission block contains 2 part properties, which are the fuelReturn property with the type of Fuel and the fuelSupply property with the type of Fuel.*

# Contribution 2: Natural Search in Models (NLU)



# What is Natural Language Understanding?

- NLU reduces human speech into a structured ontology -- a language model consisting of semantics and pragmatics definitions that is used for recognition:
  - **Intent recognition:** identifying the writer's sentiment and determining his/her objective. Generally, establishes the meaning of the text.
  - **Entity recognition** focuses on identifying entities in a message, like named and numerical entities.
    - People, locations, etc.
    - Dates, percentages, currencies, etc.

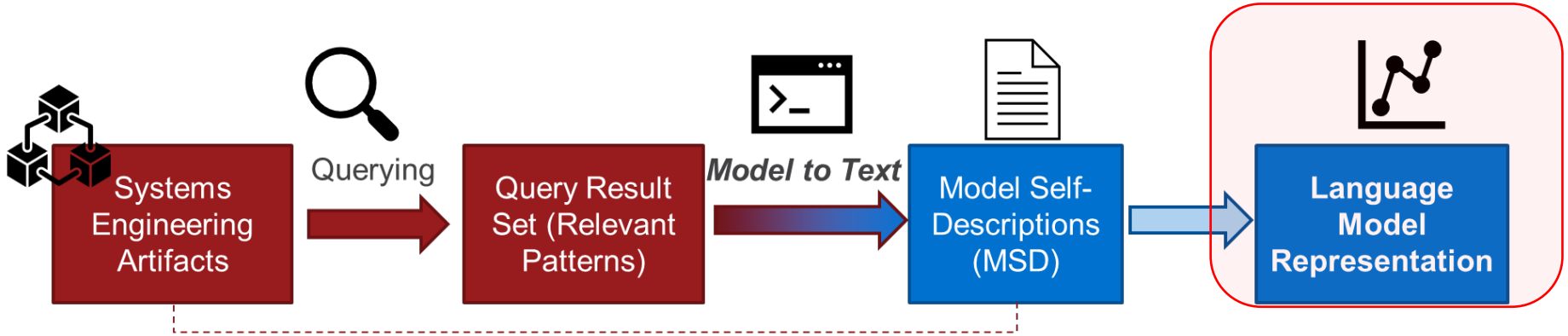

*https://www.kdnuggets.com/2019/03/deconstructing-bert-part-2-visualizing-inner-workings-attention.h*

# Language Models

- Generally, a language model is a representation of natural-language texts in numerical form, allowing for machine interpretations
- However, their complexity ranges from simple, fixed word-to-vector mappings to cutting-edge neural networks
- We experimented using two established language models:

  - **TF-IDF** (term frequency – inverse document frequency)
    - A rather simple approach, with vector representations of words: A term gets **higher weights if it is frequent in the input document** we study, but **rare in the overall language corpus**. Intuitively, this model could support domain-specific search scenarios such as ours.
  - **BERT** by Google (a deep learning model with context awareness features)
    - Pre-trained on Wikipedia and other generic, large-volume English-language sources, BERT can be downloaded by anyone and used for NLP/NLU scenarios. It promises advanced, **human-like understanding in certain contexts by its context-sensitive** vector encodings.

# Experiment: Semantic Search in the HSUV Model




- In this experiment, we use an intuitive set of five search terms, representing various semantic facets of the input model (and its domain), as well as divers search strategies from single-word searches to actual questions:
  - *environment*
  - *environmental awareness*
  - *fuel*
  - *What properties do we store about fuel?*
  - *subsystem of HybridSUV*

# Results: TF-IDF

| Search term | Top 5 Results |
|---|---|
| *environment* | (355, 0.95386463) There is a Environment block in the model.<br>(362, 0.9485242) The Environment block does not have any port.<br>(357, 0.91609514) The Environment block does not have any value properties.<br>(356, 0.7378269) The Environment block is in the HSUVModel package.<br>(81, 0.73392475) The HSUVModel package contains the Environment block. |
| *environmental awareness* | (0, 0.0) There is a HSUV Specification package in the model.<br>(1, 0.0) The HSUV Specification package is in the HSUV Requirements package.<br>(2, 0.0) The HSUV Specification package does not have any subpackages.<br>(3, 0.0) The HSUV Specification package does not contain any blocks.<br>(4, 0.0) There is a HSUV Requirements package in the model. |
| *fuel* | (501, 0.9324345) There is a Fuel block in the model.<br>(508, 0.92490125) The Fuel block does not have any port.<br>(507, 0.8482605) The Fuel block does not have any part properties.<br>(502, 0.6503485) The Fuel block is in the HSUV Structure package.<br>(112, 0.6460559) The HSUV Structure package contains the Fuel block. |
| *What properties do we store about fuel?* | (507, 0.9963838) The Fuel block does not have any part properties.<br>(501, 0.79381794) There is a Fuel block in the model.<br>(508, 0.7874046) The Fuel block does not have any port.<br>(476, 0.6349541) The Transmission block contains 2 part properties, which are the fuelReturn property with the type of Fuel and the fuelSupply property with the type of Fuel.<br>(605, 0.63273764) The ElectricMotorGenerator block contains 2 part properties, which are the fuelSupply property with the type of Fuel and the fuelReturn property with the type of Fuel. |
| *subsystem of HybridSUV* | (374, 0.9375306) There is a HybridSUV block in the model.<br>(389, 0.93050283) The HybridSUV block does not have any port.<br>(1000, 0.71769166) The l part property is owned by the HybridSUV block.<br>(1003, 0.71769166) The bk part property is owned by the HybridSUV block.<br>(1006, 0.71769166) The b part property is owned by the HybridSUV block. |

# Results: BERT

| Search term | Top 5 Results |
|---|---|
| *environment* | (998, 0.33707714) The weather is a Weather type part property.<br>(991, 0.34621608) The object part property is owned by the Environment block.<br>(357, 0.35240495) The Environment block does not have any value properties.<br>(593, 0.360799) The AutomotiveDomain block does not have any value properties.<br>(746, 0.36104614) The EconomyEquation block does not have any part properties. |
| *environmental awareness* | (361, 0.38415742) The Environment block has a weather part property with the type of Weather.<br>(997, 0.3849529) The weather part property is owned by the Environment block.<br>(346, 0.39008915) The Weather block is in the HSUVModel package.<br>(994, 0.3911605) The road part property is owned by the Environment block.<br>(901, 0.39734358) The temperature value property is owned by the Fuel block. |
| *fuel* | (1230, 0.29164267) The Port port is owned by the FuelInjector block.<br>(1304, 0.29222393) The press port is owned by the FuelFlow block.<br>(1312, 0.29304463) The flowrate port is owned by the FuelFlow block.<br>(27, 0.29450905) The Wheel package contains the block.<br>(507, 0.29519367) The Fuel block does not have any part properties. |
| *What properties do we store about fuel?* | (993, 0.2565186) There is a road part property in the model.<br>(898, 0.25655788) The pressure value property is owned by the Fuel block.<br>(996, 0.25691932) There is a weather part property in the model.<br>(432, 0.25707555) The FuelFlow block does not have any value properties.<br>(901, 0.26040572) The temperature value property is owned by the Fuel block. |
| *subsystem of HybridSUV* | (125, 0.25454134) The HSUV Structure package contains the InteriorSubsystem block.<br>(622, 0.2562611) The InteriorSubsystem block is in the HSUV Structure package.<br>(592, 0.2611069) The AutomotiveDomain block is in the HSUV Structure package.<br>(388, 0.26409155) The HybridSUV block has a p part property with the type of PowerSubsystem.<br>(385, 0.26469934) The HybridSUV block has a b part property with the type of BodySubsystem. |

# Qualitative discussion

- The “environment” search term reveals that in a model-based setting, using deeper natural-language semantics does not obviously bring benefits: result with TF-IDF are much cleaner, while BERT results also include nonsense
- With a slightly more complicated term, “environmental awareness”, TF-IDF seems to produce erroneous behavior, while the results of BERT look smart
- Results for “fuel”, again, show that simplicity also has its merits: BERT’s findings feel totally irrelevant
- However, interestingly, the power of BERT is somewhat increased if we look for the same thing in a natural-language questioning scenario: “What properties do we store about fuel?”
- Finally, the word “HybridSUV" is certainly a technical name. As a consequence, TFIDF gives as results where this very form of the word occurs. BERT, is capable of matching that term to the HSUV acronym. Furthermore, the results reflects upon the search intent of sub-structure.

# Conclusion and Future Work

- We have provided a **conceptual framework** for bridging state-of-the-art NLU techniques and systems model comprehension, along with an **experimental prototype**.
- First, preliminary results show that domain-specific semantic comprehension most likely **requires a combination** of simplistic search mechanisms and advanced semantic approaches.

- Obviously, there is need for **further evidence** with more search terms, **diverse corpora** and more engineering domains.
- There is a yet unexploited potential in fine-tuning (**training**) neural networks like BERT **on actual system models** (MSDs) to increase search quality.

# Thank you!

INCQUERY GROUP

32nd Annual INCOSE international symposium

hybrid event

Detroit, MI, USA
June 25 - 30, 2022

www.incose.org/symp2022

akos.horvath@incquerylabs.com

incquery.io

+36 30 5511164

# Thank you!


Meet you at our booth (D1)

akos
ryiabs.com
incque
+36 30 5511164