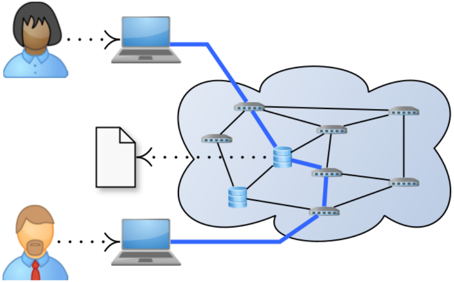
Leveraging Knowledge Graphs For Classifying Incident Situations in ICT Systems

GRASEC @ ARES 2023

Lionel Tailhardat, Orange, lionel.tailhardat@orange.com
Raphaël Troncy, EURECOM, raphael.troncy@eurecom.fr
Yoan Chabot, Orange, yoan.chabot@orange.com

2023-08-30

Context & motivations: alarm spreading & heterogeneous networks



Scenario Networking / online collaboration

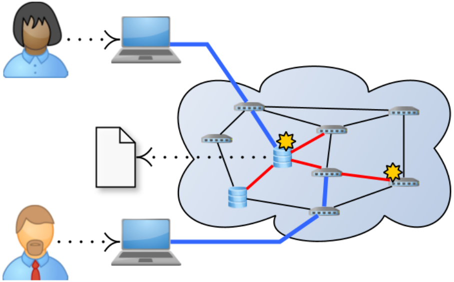
Situation Impaired network service

Observables Alarms and logs from multiple monitoring systems

Diagnosis Situation understanding through causal models

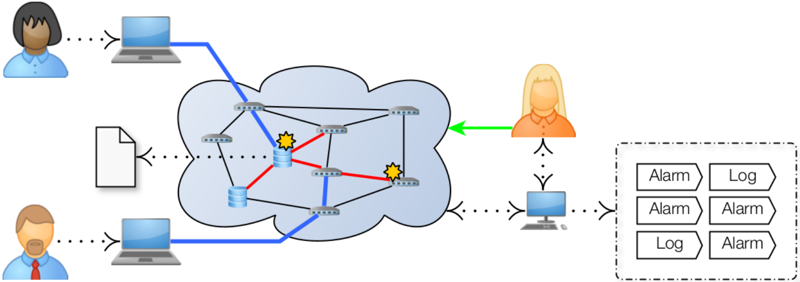
Real world Alarm spreading phenomenon, heterogeneous networks (multi-technology, multi-vendor)

Context & motivations: alarm spreading & heterogeneous networks



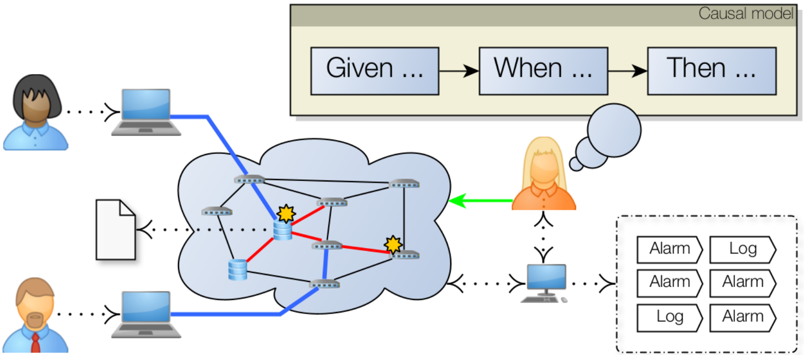
Scenario	Networking / online collaboration
Situation	Impaired network service
Observables	Alarms and logs from multiple monitoring systems
Diagnosis	Situation understanding through causal models
Real world	Alarm spreading phenomenon, heterogeneous networks (multi-technology, multi-vendor)

Context & motivations: alarm spreading & heterogeneous networks



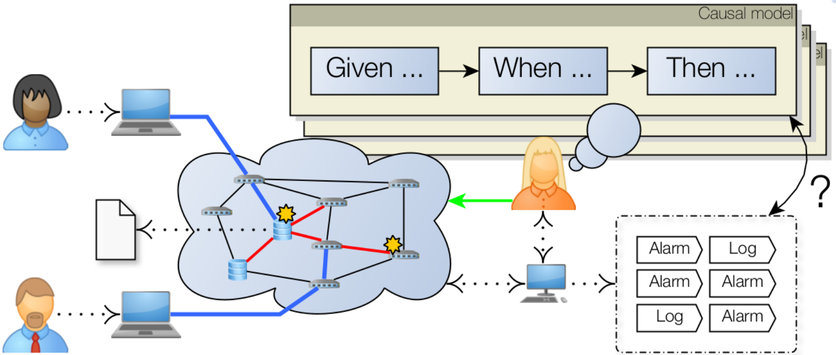
Scenario	Networking / online collaboration
Situation	Impaired network service
Observables	Alarms and logs from multiple monitoring systems
Diagnosis	Situation understanding through causal models
Real world	Alarm spreading phenomenon, heterogeneous networks (multi-technology, multi-vendor)

Context & motivations: alarm spreading & heterogeneous networks



Scenario	Networking / online collaboration
Situation	Impaired network service
Observables	Alarms and logs from multiple monitoring systems
Diagnosis	Situation understanding through causal models
Real world	Alarm spreading phenomenon, heterogeneous networks (multi-technology, multi-vendor)

Context & motivations: alarm spreading & heterogeneous networks



Scenario	Networking / online collaboration
Situation	Impaired network service
Observables	Alarms and logs from multiple monitoring systems

Diagnosis	Situation understanding through causal models
Real world	Alarm spreading phenomenon, heterogeneous networks (multi-technology, multi-vendor)

Problem statement: explicit representation of anomaly models

Incident Management How can we provide a unified approach to the diagnostic stage?

Anomaly Modeling Which techniques include the notion of time and explainability capabilities?

Decision support How do we learn/use a manipulable representation of anomalies?

Approach

1. Formalizing knowledge representation and inference needs, using expert opinions.
2. Developing a method to explicitly represent anomaly models based on RDF knowledge graphs
   - Predict the category of a trouble ticket using graph embeddings,
   - Link anomaly models to a logical representation through a qualitative analysis of incident tickets.

Working hypothesis

- Shared vocabulary for describing ICT systems $\rightsquigarrow$ easier situation understanding.
- Relational structure for each type of incident $\rightsquigarrow$ phenomena that occur in network operations.

Problem statement: explicit representation of anomaly models

Incident Management How can we provide a unified approach to the diagnostic stage?

Anomaly Modeling Which techniques include the notion of time and explainability capabilities?

Decision support How do we learn/use a manipulable representation of anomalies?

Approach

1. Formalizing knowledge representation and inference needs, using expert opinions.
2. Developing a method to explicitly represent anomaly models based on RDF knowledge graphs
   - Predict the category of a trouble ticket using graph embeddings,
   - Link anomaly models to a logical representation through a qualitative analysis of incident tickets.

Working hypothesis

- Shared vocabulary for describing ICT systems $\rightsquigarrow$ easier situation understanding.
- Relational structure for each type of incident $\rightsquigarrow$ phenomena that occur in network operations.

Problem statement: explicit representation of anomaly models

Incident Management How can we provide a unified approach to the diagnostic stage?

Anomaly Modeling Which techniques include the notion of time and explainability capabilities?

Decision support How do we learn/use a manipulable representation of anomalies?

Approach

1. Formalizing knowledge representation and inference needs, using expert opinions.
2. Developing a method to explicitly represent anomaly models based on RDF knowledge graphs
   - Predict the category of a trouble ticket using graph embeddings,
   - Link anomaly models to a logical representation through a qualitative analysis of incident tickets.

Working hypothesis

- Shared vocabulary for describing ICT systems $\rightsquigarrow$ easier situation understanding.
- Relational structure for each type of incident $\rightsquigarrow$ phenomena that occur in network operations.

Challenges: incident diagnosis use cases

Goal Providing a unified approach to the incident diagnostic stage.

Approach Get specific on the nature of the analysis and responses that are performed (scoping the diagnostic phase) based expert panel interviews (16 NOC/SOC/field experts from Orange $\simeq$ 150 operational team members).

#	Description
1	Circumscribe assets and causes search space for multi-applications incident situations ★
2	Alert on impaired service situations occurring on (distributed) fail-over architectures
3	Assess legitimacy of a given network flow
4	Track single identity from a set of various activity traces
5	Analyze false-positive and recurrent cyber security alerts
6	Analyze compliance of web navigation traces from institutional website

Focus case #1

- Most challenging.

Challenges: incident diagnosis use cases

Goal Providing a unified approach to the incident diagnostic stage.

Approach Get specific on the nature of the analysis and responses that are performed (scoping the diagnostic phase) based expert panel interviews (16 NOC/SOC/field experts from Orange $\simeq$ 150 operational team members).

#	Description
1	Circumscribe assets and causes search space for multi-applications incident situations ★
2	Alert on impaired service situations occurring on (distributed) fail-over architectures
3	Assess legitimacy of a given network flow
4	Track single identity from a set of various activity traces
5	Analyze false-positive and recurrent cyber security alerts
6	Analyze compliance of web navigation traces from institutional website

Focus case #1

- Most challenging.
- Encompasses the other use cases (generalizes the heuristic established in the incident diagnostic phase).

Experimental setup and methodology

Data integration

Data as an RDF Knowledge Graph

- Orange internal data sources (network topology, alarms, trouble tickets, etc.)
- Knowledge graph-based platform [1]
- NORIA-O RDFS/OWL data model [2]

Statistical Learning

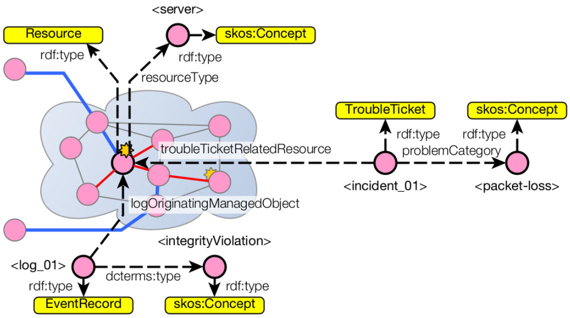
Decision support as a classification problem

- Predict the category of a trouble ticket
- Graph embeddings
  (random walk + CBOW model)
- Multiclass classifier
  (random forest, F1 weighted score model selection)

Model-based AD

Link anomaly models to a logical representation

- Analyze trouble tickets qualitatively
- Highlight corresponding SPARQL queries
- Compare queries with the classifier
  (embeddings' similarity graph + reciprocal alignment of groups with the Szymkiewicz-Simpson coefficient)



[1] Tailhardat, et al. 2023. "Designing NORIA: a Knowledge Graph-based Platform for Anomaly Detection and Incident Management in ICT Systems"

[2] Tailhardat, et al. 2022. "NORIA-O: An Ontology for Anomaly Detection and Incident Management in ICT Systems" (under review)

Experimental setup and methodology

Data integration

Data as an RDF Knowledge Graph

- Orange internal data sources (network topology, alarms, trouble tickets, etc.)
- Knowledge graph-based platform [1]
- NORIA-O RDFS/OWL data model [2]

Statistical Learning

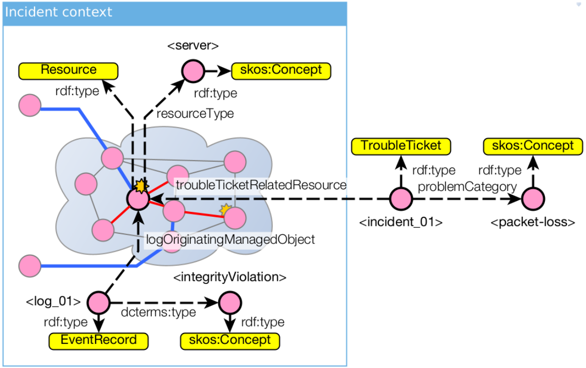
Decision support as a classification problem

- Predict the category of a trouble ticket
- Graph embeddings
  (random walk + CBOW model)
- Multiclass classifier
  (random forest, F1 weighted score model selection)

Model-based AD

Link anomaly models to a logical representation

- Analyze trouble tickets qualitatively
- Highlight corresponding SPARQL queries
- Compare queries with the classifier
  (embeddings' similarity graph + reciprocal alignment of groups with the Szymkiewicz-Simpson coefficient)



[1] Tailhardat, et al. 2023. "Designing NORIA: a Knowledge Graph-based Platform for Anomaly Detection and Incident Management in ICT Systems"

[2] Tailhardat, et al. 2022. "NORIA-O: An Ontology for Anomaly Detection and Incident Management in ICT Systems" (under review)

Experimental setup and methodology

Data integration

Data as an RDF Knowledge Graph

- Orange internal data sources (network topology, alarms, trouble tickets, etc.)
- Knowledge graph-based platform [1]
- NORIA-O RDFS/OWL data model [2]

Statistical Learning

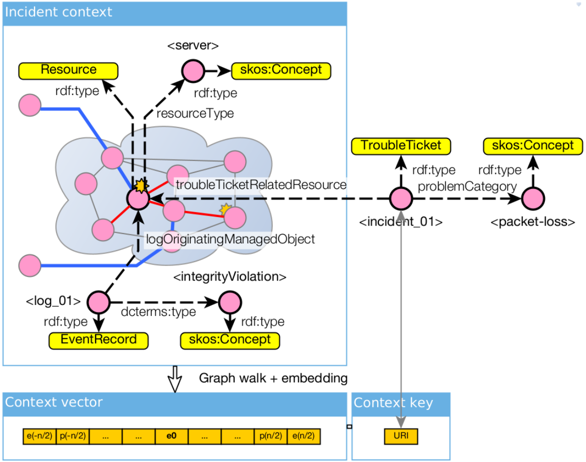
Decision support as a classification problem

- Predict the category of a trouble ticket
- Graph embeddings
  (random walk + CBOW model)
- Multiclass classifier
  (random forest, F1 weighted score model selection)

Model-based AD

Link anomaly models to a logical representation

- Analyze trouble tickets qualitatively
- Highlight corresponding SPARQL queries
- Compare queries with the classifier
  (embeddings' similarity graph + reciprocal alignment of groups with the Szymkiewicz-Simpson coefficient)



[1] Tailhardat, et al. 2023. "Designing NORIA: a Knowledge Graph-based Platform for Anomaly Detection and Incident Management in ICT Systems"

[2] Tailhardat, et al. 2022. "NORIA-O: An Ontology for Anomaly Detection and Incident Management in ICT Systems" (under review)

Experimental setup and methodology

Data integration

Data as an RDF Knowledge Graph

- Orange internal data sources (network topology, alarms, trouble tickets, etc.)
- Knowledge graph-based platform [1]
- NORIA-O RDFS/OWL data model [2]

Statistical Learning

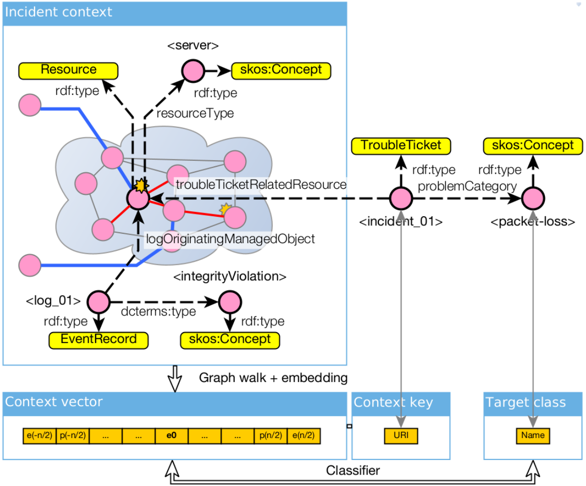
Decision support as a classification problem

- Predict the category of a trouble ticket
- Graph embeddings
  (random walk + CBOW model)
- Multiclass classifier
  (random forest, F1 weighted score model selection)

Model-based AD

Link anomaly models to a logical representation

- Analyze trouble tickets qualitatively
- Highlight corresponding SPARQL queries
- Compare queries with the classifier
  (embeddings' similarity graph + reciprocal alignment of groups with the Szymkiewicz-Simpson coefficient)



[1] Tailhardat, et al. 2023. "Designing NORIA: a Knowledge Graph-based Platform for Anomaly Detection and Incident Management in ICT Systems"

[2] Tailhardat, et al. 2022. "NORIA-O: An Ontology for Anomaly Detection and Incident Management in ICT Systems" (under review)

Experimental setup and methodology

Data integration

Data as an RDF Knowledge Graph

- Orange internal data sources (network topology, alarms, trouble tickets, etc.)
- Knowledge graph-based platform [1]
- NORIA-O RDFS/OWL data model [2]

Statistical Learning

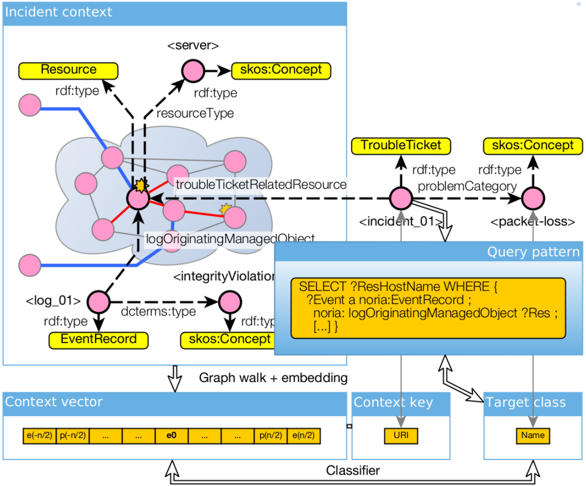
Decision support as a classification problem

- Predict the category of a trouble ticket
- Graph embeddings
  (random walk + CBOW model)
- Multiclass classifier
  (random forest, F1 weighted score model selection)

Model-based AD

Link anomaly models to a logical representation

- Analyze trouble tickets qualitatively
- Highlight corresponding SPARQL queries
- Compare queries with the classifier
  (embeddings' similarity graph + reciprocal alignment of groups with the Szymkiewicz-Simpson coefficient)



[1] Tailhardat, et al. 2023. "Designing NORIA: a Knowledge Graph-based Platform for Anomaly Detection and Incident Management in ICT Systems"

[2] Tailhardat, et al. 2022. "NORIA-O: An Ontology for Anomaly Detection and Incident Management in ICT Systems" (under review)

Evaluation

Classifier

Data integration 15 sources $\rightarrow$ 4M triples (400K entities)
138 noria:TroubleTicket entities
5 target class (noria:troubleTicketCategory property)

Best model 0.81 F1 weighted score
Embeddings: Walk Depth $= 8$, Walk Count $= 30$ (WD08-WC30)
Random forest: max tree depth $= 5$, tree count $= 20$, max feature count $=$ SQRT, information gain criterion $=$ gini

	WC10	WC20	WC30
WD04	0.64 gini-05-SQRT-030	0.59 gini-05-SQRT-020	0.73 gini-05-SQRT-030
WD08	0.49 gini-05-SQRT-100	0.75 gini-05-SQRT-050	**0.81** gini-05-SQRT-020
WD10	0.52 gini-05-SQRT-020	0.60 gini-05-SQRT-020	0.76 gini-05-SQRT-020

Strengths The classifier shows a reasonably good performance in terms of precision and recall for a first attempt.

Caveats The dataset is too small (for some classes in particular) + available context for trouble ticket entities is not systematically consistent.

Evaluation

Logical representation

Data integration

(same as previously)
139 noria:TroubleTicket entities

Query patterns

12 patterns
0.09 reduction factor / dataset
0.12 average overlap / classifier

Strengths Low number of patterns + polyvalent patterns + n − to − m pattern/class relationship.

Caveats Unclear pattern/class relationship due to dataset inconsistencies.

Pattern name	Example	Count	Overlap
AlarmState	Service disruption on Optical Network Terminal (ONT). Unable to access http://example.org	**49**	0.30
AuthError	Authentication error. User does not have access to the 'xxx' role. Please check my rights.	**9**	0.13
CoFailure	Co-occurring alarm in a network device neighborhood and creation of a parent/child relationship between trouble tickets for Service Level Agreement (SLA) tracking.	**3**	0.13
Complex	Requires further expertise for providing a pattern.	**15**	0.13
Debug	Non-relevant trouble ticket entity, present for debugging purposes of the ticketing system.	**6**	0.13
ErroneousRes.	The processing flow references a resource that does not exist.	**9**	0.13
HeartBeat	The number of failed calls has increased significantly. No response to SNMP polling and Ping. Agent not running or cannot communicate. Extreme slowness or even unavailability of the service when opening and closing documents on the platform.	**13**	0.13
Overbilling	-	**6**	0.00
RecurringFai.	Repeated occurrence of the same type of failure on a device within a short period of time.	**1**	0.13
RequestForIn.	Please decommission the 'xxx' system. The Customer is calling about the Request For Change (RFC) status.	**17**	0.14
RiskPreventi.	Automated deployment flow triggered on resource.	**17**	0.13
RMA	Return Merchandise Authorization (RMA) for redundant Power Supply Unit (PSU).	**10**	0.00
	16	**139**	0.12

With respect to the WD08-WC30 model / k $=3$ similarity graph

Summary & future work

Problem Learning and use of a manipulable representation of anomalies for decision support.

Our approach Knowledge representation using SemWeb technologies, multiclass classifier with graph embeddings, model-based anomaly detection.

Next Reliable data integration, semantic annotation of unstructured data, situation diagnosis through incident models.

Paper

Leveraging Knowledge Graphs For Classifying Incident Situations in ICT Systems.
https://doi.org/10.1145/3600160.3604991

Code repository

- NORIA-O
  https://w3id.org/noria/
- SMASSIF-RML
  https://github.com/Orange-OpenSource/SMASSIF-RML
- ssb-consum-up
  https://github.com/Orange-OpenSource/ssb-consum-up
- grlc
  https://github.com/Orange-OpenSource/grlc

Appendices

Challenges: anomaly modeling technique families

Principles	Strengths	Weaknesses
	Model-Based Design	
Query the graph to retrieve anomalies and their context.	Detecting anomalies “recorded” somehow in the graph thanks to the alarm system; straightforward translation of simple anomaly detection rules; multiple abstraction levels (subsumption).	Relies on expert knowledge; lack of probabilistic reasoning; hard to represent sequential decisions; may require to infer more prior information about the anomaly, e.g. its type using classification.
	Process Mining	
Align a sequence of entities to activity models, then use this relatedness to guide the repair.	Detecting anomalies with multiple alerting signals and sequential decisions; replayable models.	Relies on expert knowledge; may require denoising models; probabilistic relatedness.
	Statistical Learning	
Relate entities based on context similarities, then use this relatedness to alert and guide the repair.	Detecting anomalies with multiple alerting signals.	Requires fine tuning of the context definition depending on use case and temporality requirements; probabilistic relatedness.

This work:

- **Focus** Model-Based Design and Statistical Learning
- **Set aside** Process mining approach, because it only captures local processes and therefore misses out on the need for learning from a larger context that is enabled by graph embeddings.

Towards reasoning services for decision support

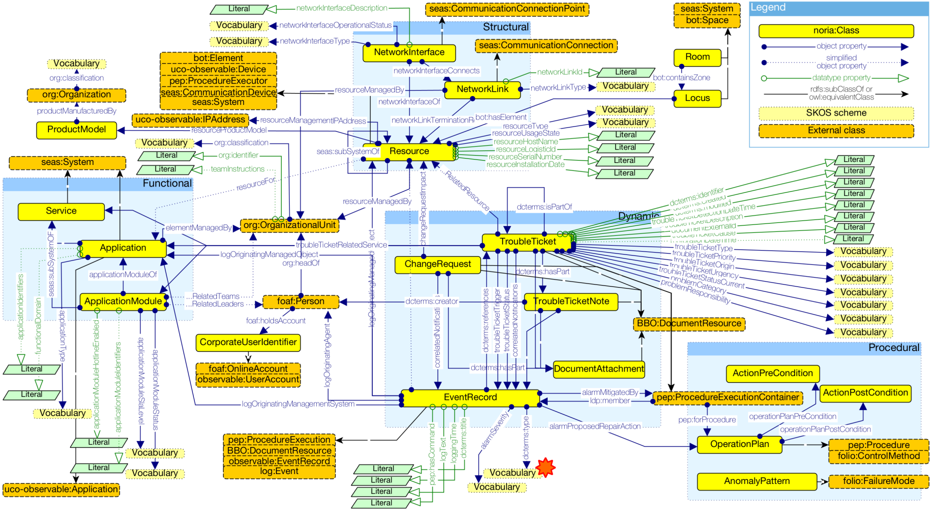
Stages of the incident management process where a recommendation system can be useful:

- Before the ticket creation (early detection),
- At the ticket opening (cause/solution similarity based on ticket descriptors and context),
- During the resolution (cause/solution refinement and proposal of next action based on the actions taken).

Reasoning services (proposal)

1. Predicting the category of a trouble ticket,
2. Predicting the probable cause of a trouble ticket,
3. Detecting anomalies before a trouble ticket is even created,
4. Adding comments to a given trouble ticket (e.g. next best action to undertake),
5. Calculate the n closest anomalies given an observed anomaly.

Overview of the NORIA-O v0.3 data model



Implementation NORIA-O → https://w3id.org/noria/ (open source release under BSD-4 license)

Paper Tailhardat, et al. 2022. "NORIA-O: An Ontology for Anomaly Detection and Incident Management in ICT Systems"

Who's who

Lionel Tailhardat AI R&D Engineer

- Dynamic Systems, Dependability and Knowledge Engineering
- genears.github.io

Dr. Raphaël Troncy Associate Professor

- Knowledge Engineering, Knowledge Graphs and Data Science
- www.eurecom.fr/~troncy

Dr. Yoan Chabot AI Researcher

- Knowledge Engineering, Knowledge Graphs and Data Science
- yoanchabot.github.io

Orange

Intl. Telecommunication infrastructure and service provider (and more ...)

- www.orange.com
- hellofuture.orange.com

EURECOM

Graduate School and Research Center in Digital Science

- www.eurecom.fr

Our proposition: combine AI and Knowledge Engineering techniques for Complex Networks Resilience and Data Security concerns.