Semantic Systems Research Lab
-Projects-

Marta Sabou
Fajar J. Ekaputra
Peb R. Aryan
Laura Waltersdorfer
Kevin Haller
Stefani Tsaneva
Gregor Käfer
Ayu Irsyam

Institute of Software Technology and Interactive Systems
Information and Software Engineering Group
Vienna University of Technology

Favoritenstraße 9-11/177,
1040 Vienna, Austria
The Semantic Systems Research Lab performs foundational and applied research in the area of information systems enabled by semantic (web) technologies. Its work is situated at the confluence of Semantic Web and Human Computation research areas, and focuses on the following main research questions:
(1) How to involve people in (semantic) system design both for information and software engineering? We investigate the use of Human Computation methods for acquiring and verifying knowledge structures such as ontologies and knowledge graphs (HOnEst). We also use Human Computation methods for verifying software engineering models.

(2) How to ensure that (semantic) information systems operate in reliable (trustworthy) and ethically acceptable ways? We perform work on using semantic technologies to make information systems auditable (OBARIS, WeilFort, VasQua). Additionally, we extend this research to cyber-physical systems (CPS) where we aim to enable their explainability (ExpCPS).

(3) How can semantic (web) technologies improve information systems? We investigate in particular the benefits of semantic methods from the area of data integration (CitySPIN, CDL-Flex) and exploratory search (OntoTrans, STAR) in information systems from a broad range of domains including Smart City, Industrie 4.0, material engineering, and medical science.
(1) How to involve people in (semantic) system design both for information and software engineering?
Ontologies and Knowledge Graphs

Ontology defects hamper the system’s ability to provide factually correct and unbiased answers.

- Often human involvement is needed to ensure high quality domain models.
Incorrect system diagrams lead to faults in the follow-up artifacts:

- High costs
- Software model inspection
  - on-site verification processes:
  - not scalable to larger models
  - not suitable for virtual organisations
Human Computation harnesses Human Contributions on Large Scale

100 million/day

Digitization of:
13 million NYT articles;
2 million books per year;

“Players took 3 weeks to solve the three dimensional structure of a [...] protein [...], but whose structure had eluded biochemists for more than a decade.”
http://blogs.nature.com/spoonful/2012/04/foldit-games-next-play-crowdsourcing-better-drug-design.html


Motivation

Ontologies, taxonomies, Knowledge Graphs need manual verification

EER diagrams manually checked for completeness and correctness with respect to system specifications

How to propose an HC based solution to the problem of conceptual model verification that is applicable across research areas?

Contributions:
- Problem formalization
- HC Approach
- Evaluation in the Software and Knowledge Eng. Areas

HC techniques primarily used to address cognitively easy tasks (e.g., labeling, object recognition)

Problem Formalization & Generic Approach for Verifying Conceptual Models (VeriCoM)

P1: ME-based. MEs guide the verification.

\[ M = \bigcup_n ME_n \]

P2: Evidence-based. Evidence for me from FR is used as context in the HC Task.

P3: Model as context. Model fragment used as context in the HC Task.

P4. Defect types. Defect types used to design HC Task.

In a Nutshell: Further Results

Crowd-based inspection comparable to traditional inspection [1]

Open ended task design helps harness subtle insights and creativity [2]

Human Computation technique as a novel method for empirical software engineering experimentation [3].

Ongoing and Future Work

HOnEst: Human-centric Ontology Evaluation (FWF)
AIM: “scalable human-centric ontology evaluation”

- FWF Elise Richter
- 2020-2024
- UPM, KMi, UMannheim

Systematic literature review [1]: What semantic resources are evaluated with human-involvement, why and how is the evaluation performed?

Task design, empirical studies, support ontology engineer courses during distance learning [2].

Hybrid human-machine system including reasoners and machine learning systems [3].

Industry Project: SIEMENS OPC UA

**ORE: OPC-UA Rule Editor (ORE) Project**

**Goal:** supporting OPC-UA experts in identifying **constraints** in pdf documents of Companion Specifications and formalizing **rules** to validate if the constraints are satisfied in OPC-UA files.

- Siemens DE
- 2021
- (2022 extension)

**Constraints** as text snippets, extracted from OPC-UA specifications, that express some kind of requirements that the OPC-UA file conforming to those specifications should fulfill.

**EXAMPLE**
Each FiniteStateMachine can have at most one State of type InitialStateType, but a FiniteStateMachine does not have to have a State of this type.

**Rules** are translations of the content expressed by a constraint into a formal representation (e.g., a SPARQL query).
In this survey, rules are represented in a structured natural language so that they can be understood also by OPC-UA experts that are not familiar with the syntax of the formal rule language. Complex constraints can be translated in several rules that we refer to as a **Rule Sets**.

**EXAMPLE**
1. The node FiniteStateMachine exists.
2. The node InitialStateType exists.
3. There should not exist a FiniteStateMachine with more than one State of type InitialStateType.
Research Areas and Projects

(3) How can semantic (web) technologies improve information systems?

- **Research Communities**
  - **FAIR Data**
    - **Research Topics**
      - **Data Citation & Reproducibility**
        - **Research Projects**
          - Siemens ExpCPS (2019-2021)
          - WAW VasQua (2020-2021)
      - **KG-based Explainability**
        - **Research Projects**
          - FFG WellFort (2019-2022)
      - **Auditable Information System**
        - **Research Projects**
          - FFG OBARIS (2020-2022)
      - **Ontology Engineering (OBDI)**
        - **Research Projects**
          - H2020 OntoCommons (2020-2023)
      - **Exploratory Search System**
        - **Research Projects**
          - H2020 OntoTrans (2020-2024)
      - **HC & C**
    - **Research Projects**
      - FWF HoNEST (2020-2024)
      - Quality Software Engineering
      - Software Quality Improvements

**Application Domains**
- Smart Energy Grid
- Travel Management
- Medical
- Environment
- Manufacturing
- Material Engineering
OntoTrans

- EU - H2020
- 2020 - 2024

(A) representing manufacturing process challenges in a standard ontological form (User Case)

(A) connecting User Cases with existing appropriate information sources i.e. available data and materials modelling solutions

(A) recommending consistent materials modelling workflow options

aim: support the development of dedicated Apps delivering a smart guidance for materials producers and product manufacturers
End users and translators may not have a clear view of the information they need but rather require support to discover new information and learn about the domain. OntoTrans will support them with an ESS which will rely on data science algorithms to identify emerging patterns in the available semantic information.
Exploratory Search Platform

Obj1) adaptable to different use cases and legacy environments (without costly changes)

Obj2) allow experimentation with different interface paradigms (without costly changes)

Obj3) easy plugin system for new algorithms for knowledge graph analytics
**Obj3) Analytical Pipeline**

- **Analytical Services** analyze the (semantic) structure of a knowledge graph
  - can make use of:
    i. SPARQL
    ii. Gremlin
    iii. Full-Text Search Index
    iv. Other Analytical Services

- new services can be plugged into the application

- **Centrality metrics** assign an “importance” value to entities in the knowledge graph.

- **Similarity metrics** assign a “similarity” value to a pair of entities in the knowledge graph.
Analytical Services

Centrality Metrics
- Degree
- Betweenness
- Page Rank
- LCA Page Rank

Similarity Metrics
- Resnik Similarity
- Linked Data Semantic Distance
- Peer Pressure Clustering
- KG Embeddings
Exploratory Search Interface (UC1)

Keyword Search

UNIVERSAL ROBOTS URS
- Robot Type
  - Universal Robotics

UNIVERSAL ROBOTS UR10E
- Robot Type
  - Universal Robotics

UNIVERSAL ROBOTS UR10E
- Robot Type
  - Universal Robotics

UNIVERSAL ROBOTS UR3E
- Robot Type
  - Universal Robotics

Related Robot Types
- COMAU AURA
  - Robot Type
  - Universal Robotics
  - Origin: Germany
  - Reach: 2790 mm
  - Handling Payload: 170 Kg
  - Degrees Of Freedom: 6

- COMAU E.DO 6
  - Robot Type
  - Universal Robotics
  - Keine Beschreibung

- STABLI TX2 90XL
  - Robot Type
  - Universal Robotics
  - Keine Beschreibung

- ROZUM PULSE 75
  - Robot Type
  - Universal Robotics
  - Keine Beschreibung

- PRECISE AUTOMATIC PAVS6
  - Robot Type
  - Universal Robotics
  - Keine Beschreibung

Recommendations/Infobox

FAKULTÄT FÜR INFORMATIK
Research Areas and Projects

(2) How to ensure that (semantic) information systems operate in reliable (trustworthy) and ethically acceptable ways?
How to design and implement Auditable Semantic AI Systems (SAIS)?

- G1: Provide a taxonomy and characterization of SAIS
- G2: Establish a technology stack for SAIS
- G3: Provide data integration & management methods for SAIS
- G4: Create an auditability framework for SAIS
- G5: Introduce and evaluate SAIS stack in concrete environmental use cases
1. Semantic Web
   - symbolic knowledge
   - explainability
   - vulnerable to noisy data
   - knowledge acquisition bottleneck

2. Machine Learning
   - knowledge from sparse data
   - broad applicability
   - intransparency
   - often lots of training data required

3. Semantic Web Machine Learning
   - Focus on semantic web resources and machine learning approaches
   - = SWeMLS, subset of neurosymbolic AI

Context: 3rd Wave of AI

G1: Provide a taxonomy and characterization of SAIS
Systematic Mapping Study: Methodology

G1: Provide a taxonomy and characterization of SAIS
Query Date: October 2020
Research Team: 10 participants

Search Queries

G1: Provide a taxonomy and characterization of SAIS

<table>
<thead>
<tr>
<th>Sub-Query</th>
<th>Used Search Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 (SW module)</td>
<td>knowledge graph, linked data, semantic web, ontolog*, RDF, OWL, SPARQL, SHACL</td>
</tr>
<tr>
<td>Q2 (ML module)</td>
<td>deep learning, neural network, embedding, representation learning, feature learning, language model, language representation model, rule mining, rule learning, rule induction, genetic programming, genetic algorithm, kernel method</td>
</tr>
<tr>
<td>Q3 (System)</td>
<td>Natural Language Processing, Computer Vision, Information Retrieval, Data Mining, Information Integration, Knowledge Management, Pattern Recognition, Speech Recognition</td>
</tr>
</tbody>
</table>
Systematic Mapping Study: Result Set

G1: Provide a taxonomy and characterization of SAIS

Bibliographic analysis:
1) Scopus accounts for largest portion
2) Strong increase of papers over last years
**G1: Provide a taxonomy and characterization of SAIS**

**System Pattern Model**

- Statistical Training Type
- Symbol Usage
- Maturity
- Reasoner: SPARQL query engine
- Classical ML: Bayesian; KNN; SVM; Decision Trees; Mixture Models; Regressions; Markov-process Models; Clusterings; Dim. Reduction; Formal Concept A.; Rule Learning; Topic Models; SGM; Genetic Algorithms; Factorization Machines
- Deep Learning: Transformers; LSTM; Trad. RNN Models; Trad. CNN; GAN; Plain Encoder-based; Trad. FFNN; Matrix Factorization; Graph Deep Learning; Translational Distance; Rec. GNN; Conv. GNN; Graph AutoEncoders; Graph FFNN

**Domain**

- Text: Analysis; Annotation; QA & conversational; Information Extraction; Information Retrieval; Other
- Image and Video: Other
- - (impl/eval) infrastructure
- - (impl/eval) process-steps
- - (impl/eval) software
- - (eval) data
- - (eval) data-split
- - (eval) parameter
- - (eval) metrics
- - Provenance (yes/no)

**Task**

- Reasoner; SPARQL query engine;
- Text: Analysis; Annotation; QA & conversational; Information Extraction; Information Retrieval; Other
- Image and Video: Other
- - (impl/eval) infrastructure
- - (impl/eval) process-steps
- - (impl/eval) software
- - (eval) data
- - (eval) data-split
- - (eval) parameter
- - (eval) metrics
- - Provenance (yes/no)

**Documentation**

- - Atomic Pattern
- - T-Pattern
- - I-Pattern
- - Fusion Pattern
- - Y-Pattern
- - Other

**Pattern**

- Domain: General (Cross-Domain; Dictionaries); Natural Sciences; Admin. & Politics; Human Culture & Edu.; Geo. & Economics Software & Tech; Other
- Type: Thesaurus; Taxonomy; Ontology; Dataset; KB; Linked dataset; KG
- Size: <500; 500-1K; 1K-10K; 10K-100K; 100K-500K; >1M
- Formalism: owl/owl-2; rdf/rdf-s; other

**Model**

- Classical ML: Bayesian; KNN; SVM; Decision Trees; Mixture Models; Regressions; Markov-process Models; Clusterings; Dim. Reduction; Formal Concept A.; Rule Learning; Topic Models; SGM; Genetic Algorithms; Factorization Machines
- Deep Learning: Transformers; LSTM; Trad. RNN Models; Trad. CNN; GAN; Plain Encoder-based; Trad. FFNN; Matrix Factorization; Graph Deep Learning; Translational Distance; Rec. GNN; Conv. GNN; Graph AutoEncoders; Graph FFNN

**Instance**

- Symbolic Model
- Symbol
- Data

**Legend:**

- 1..1 object properties with cardinality
- 0..N subsumption (subClassOf)
- indirect relations via SHACL constraints
Based on [1], compositional design patterns for hybrid systems

SMS retrieved papers beyond these patterns, e.g. [2]


G1: Provide a taxonomy and characterization of SAIS

Follow for more detailed analysis in upcoming publications:

http://semsys.ifs.tuwien.ac.at
http://www.obaris.org
https://twitter.com/LaWaltersdorfer
Auditing SWeML

G4: Create an auditability framework for SAIS

Ongoing PhD work of Laura Waltersdorfer
(2) How to ensure that (semantic) information systems operate in reliable (trustworthy) and ethically acceptable ways?
Explainable Cyber-Physical Systems (ExpCPS)

Explainable Cyber Physical Energy Systems based on Knowledge Graph

Peb Aryan, Fajar Ekaputra, Marta Sabou, Daniel Hauer, Ralf Mosshammer, Alfred Einhalt, Tomasz Miksa, Andreas Rauber

Information & Software Engineering Research Group,
Institute of Information Systems Engineering,
Faculty of Informatics - TU Wien

Siemens AG

MSCPES 2021 18.05.2021

funded by the Austrian Research Promotion Agency (FFG)
in the PoSyCo project (FFG No. 3036508)
Motivation

Explanations could support the work of several Smart Grid stakeholders:
- Customers
- Grid operators (sales, operations, planning)
- Third parties (EVCS operators)

Explanations are the result of combining data from various sources:
- CPES assets sensors
- Open data (weather)
- Legacy systems
Flexibility Scenario

**Features**
- M1: energy market
- S1: Flex application server
- W1: weather
- H1-H4: residential houses
- B1-B2: commercial buildings
- C1-C2: EV charging stations

**Events**
- e7: trafo overload
- e4, e5, e6: high consumption
- e3: flex notification
- e1: energy market (real-time)

**Example explanation**
e7 because e(4,5,6)
e(4,5) because of e3; because of e1
e6 because of e2
Explanation Generation Process

- Event log / time series
- System states (dynamic)
- System Model (static)
- Causality Annotation
- Context KG
- Event Causality KG
- Event Annotation
- Explain

Event Causality KG

Explanation
Implementation Result

- Overload (TRAFO-BUILDING) caused by PeakConsumption (EV-STATION, RESIDENTIAL-SINGLE, COMMERCIAL-FACTORY)
- PeakConsumption (RESIDENTIAL-SINGLE) caused by LowSolarOutput (SOLAR-PANEL)
- LowSolarIntensity (AIRSHIP)
- PeakConsumption (RESIDENTIAL-SINGLE) caused by FlexRequest (AIRSHIP)

Explanation details:

```json
{
  "event": "OverloadEvent",
  "location": "TRAFO-BUILDING:09376278"
}
```

```json
{
  "causedBy": [
    0: {
      "event": "PeakConsumptionEvent",
      "location": "RESIDENTIAL-SINGLE:0938560"
    },
    1: {
      "event": "FlexRequestEvent",
      "location": "AIRSHIP:09316650"
    }
  ]
}
```
Summary: Research Areas and Projects

Research Communities
- FAIR Data
  - AR, TMi
- Semantic Web Research
  - MS, FJE, PRA, LW, KH

Research Areas
- RQ2
  - AR, TMi
    - Data Citation & Reproducibility
  - PRA, MS, TMi
    - KG-based Explainability
  - FJE, LW, MS, TMi
    - Auditable Information System
- RQ3
  - FJE, PRA, KH
    - Ontology Engineering (OBDI)
  - KH, FJE, MS
    - Exploratory Search System
- RQ1
  - HC & C
  - MS, KH
  - HC for KG Quality Improvement

Research Projects
- Siemens ExPCPS (2019-2021)
- WAW VasQua (2020-2021)
- FFG WellFort (2019-2022)
- FFG OBARIS (2020-2022)
- H2020 OntoCommons (2020-2023)
- H2020 OntoTrans (2020-2024)
- FWF HoNEST (2020-2024)

Application Domains
- Smart Energy Grid
- Travel Management
- Medical
- Environment
- Manufacture
- Material Engineering
Semantic Systems Research Lab
Thank you!
Happy 2022!