Analytical Solutions and Reasoning

and

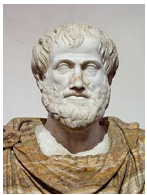
The Sirius Center

MSc topics 2021

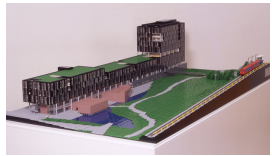
Universitetet i Oslo

18 October 2021

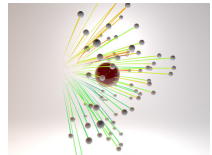
ASR – Analytical Solutions and Reasoning



Logic,
Reasoning

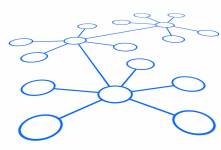


Modeling



Simulation

Databases



Semantic Web
Technologies

Model-driven
Systems

SIRIUS

Centre for Scalable Data Access
in the Oil and Gas Domain

SIRIUS: A Centre for Research-Driven Innovation

- 8 Years
- 19 Partners
- 3 Universities
- 20+ PhD
- 50+ Workers

Academia

UiO : University of Oslo

University of Oxford Department of Computer Science

simula

NTNU Norwegian University of Science and Technology

SINTEF

Industrial end-users

Schlumberger

equinor

TechnipFMC

aibel

AkerSolutions

BOSCH

DNV·GL

Vendors

ONTOPIC

KADME Knowledge and data management expertise

OSIsoft.

prediktor

Dolphin

SAP

IBM

EVRY

NUMASCALE Bigger Data Analytics

Oxford Semantic Technologies

BPT

computas

ASR and SIRIUS **MSc topics**

ASR:
https://www.mn.uio.no/ifi/studier/masteroppgaver/asr/

SIRIUS:
https://www.mn.uio.no/ifi/studier/masteroppgaver/sirius/

MSc project in Brazil?

Two Possibilities

- Federal Univ. Rio Grande do Sul (Porto Allegre)
  - Research in ontologies
  - How to structure a good ontology?
  - Ontology-based Information systems
  - Applications in Geosciences
- Univ. Espírito Santo, Vitória
  - Similar topics, ask!
- Travel plus stipend

Topic areas

- Machine Learning, Data Science and Digital Twins
- Ontology-Based Information Systems
- Modeling and Analysis
- Data Streams and Applications
- Logic and Theoretical Computer Science

Data Science Topics

How trainable are GNNs?

(Comparing trainability of Graph Neural Networks for logic-expressible functions)

- Graph Neural Networks (GNNs):
  - a family of modern NN architectures
  - work with structured data (e.g., Knowledge Graphs)
- (Theoretical) expressivity of GNNs:
  - different GNN architectures express different sets of functions on graphs
  - well-understood by connecting to logics
- (Practical) trainability of GNNs:
  - "can express in theory" does not imply "can be trained"
  - Question:
    Which GNNs architectures can be trained for which logic-expressed functions?

Supervisors: Egor V. Kostylev (egork@ifi.uio.no), Roxana Pop (roxanap@ifi.uio.no)

Predictive Analytics in Marketing & Supply

equinor

The Marketing and Supply Solutions (TDI EDT MSS) unit is developing and maintaining the software and analytics solutions to distribute, transport, and sell the gas, liquids, and electricity produced by Equinor's facilities. It's a challenging domain with many interesting optimization and prediction tasks.

If you chose a thesis with MSS, you will

- Work in a very dynamic environment in Equinor and collaborate with the shipping and trading units.
- Possibly work a few days per week from Equinor's offices in Fornebu, with possible visits to trading desks (Stavanger, London, …).
- Learn about commodity markets and trading.
- Choose between a number of possible topics in shipping, distribution, or trading.
- Learn about industrial scale machine learning, data safety, and best practices, from Equinor's data science team.

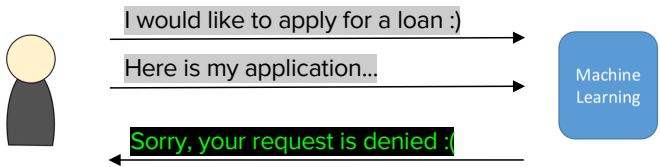
Supervisor: Dirk Hesse, Vice President TDI EDT MSS

eXplainable AI: Actionable Recourse in Machine Learning

Advisors: Ingrid Chieh Yu, Peyman Rasouli

https://www.mn.uio.no/ifi/studier/masteroppgaver/asr/actionable-recourse-in-machine-learning-classifica.html

When you are denied a loan...



Do you know what to do to get approved?

Actionable Recourse

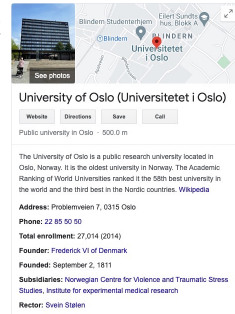
Recourse: ability to obtain a desired prediction from a machine learning model by changing actionable input variables

- Example:

	Age	Balance	# Credit Cards	Occupation	Sex	Prediction
Actual Instance	28	1000$	2	Developer	Male	Rejected!
Actionable recourse	_	**2000$**	**1**	_	_	Accepted!
Non-actionable recourse	**45**	_	_	_	_	Accepted!
Non-actionable recourse	_	_	_	_	**Female**	Accepted!

Ontology Summarization via Machine Learning Techniques

- Problems
  - Ontologies are getting bigger
  - Limited capacities of human user to understand contents
- Goal: extract top-k most relevant axioms that related to users' interests
  - Combine with machine learning techniques to extract relevant information
  - Apply tools on real-world ontologies/knowledge graphs
- Relevant courses:
  - IN3060/IN4060 Semantic Technologies
  - IN3050/IN4050 Artificial Intelligence and Machine Learning
- Supervisors: Jieying Chen (jieyingc@ifi.uio.no)



Reinforcement Learning for Knowledge Graphs

- Problems
  - Difficult to main and develop large-scale knowledge
- Tasks: (two topics)
  - Entity summarization: extract top-k most relevant triples that relate to users' interest.
  - Link prediction: eg, predict how likely there exists a relation between two entities
- Relevant courses
  - IN3060/IN4060 Semantic Technologies
  - IN3070/IN4070 Logikk
  - IN-STK5100 Reinforcement Learning and Decision Making Under Uncertainty
  - IN-STK5000 Adaptive methods for data-based decision making
- Supervisors: Jieying Chen (jieyingc@ifi.uio.no),
  Thomas Kleine Büning (thomkl@ifi.uio.no)
  Christos Dimitrakakis (chridim@ifi.uio.no)

Topics in Reinforcement Learning, Fairness or Privacy

- Supervisor: Christos Dimitrakakis (chridim@ifi.uio.no)

Building the future of drug discovery

Average cost of developing only one drug: 3.5 BILLION US$
Average time to develop only one drug: 13 YEARS!
Or sometimes (too often) “forever”!

Computer-Aided Molecular Modelling & Simulation (CAMMS) helps more and more.

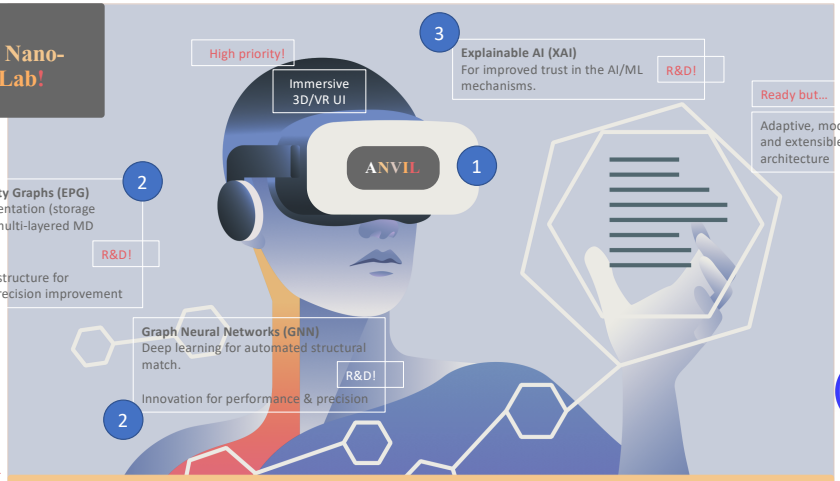
by
In-Virtualis

Challenge!

(many)

1. The user has no “intimate” relation to the data: Non-intuitive UI hinders adoption and requires expertise
1. Current tools are specific to few areas of use. They do not cover all needed workflows
2. Performance & precision always a MAJOR challenge
3. Automation using machine learning is hindered due to the “unexplainable” nature of ML
4. Many needs, no simple extension/tailoring

the Advanced Nano-Virtual Lab!



Join the fantastic journey!

Contact persons:

- Assoc. Prof. M. Naci Akkøk (nacia@ifi.uio.no), also CEO @ In-Virtualis AS (naci@in-virtualis.com) & Assoc. Prof. at OsloMet (mehmetna@oslomet.no)
- Assoc. Prof. Egor V. Kostylev (egork@ifi.uio.no)
- Assoc. Prof. Dumitru Roman (dumitrur@ifi.uio.no), also SINTEF (Dumitru.Roman@sintef.no)

- with a little help from
  Dr. Thibaud Freyd and other In-Virtualis resources

with
In-Virtualis

Analysis and Optimization of Logistics Data

Advisor: Rudi Schlatte

- Logistics Data from Equinor: Ship schedules, maintenance plans, etc.
- We are currently working on making data available for master students
- students can choose topics of interest together with the thesis advisor and the industry representative.

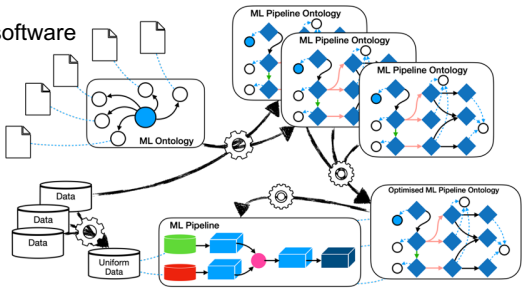
SIRIUS

Neuro-Symbolic AI-based Intelligent System for Geoscience and Sustainable Industry: Energy and Manufacturing

equinor

BOSCH
Invented for life

Scope: We develop an intelligent software system for industrial data analysis



Interdisciplinarity:

- Machine learning
- Sematic technologies
- Geoscience

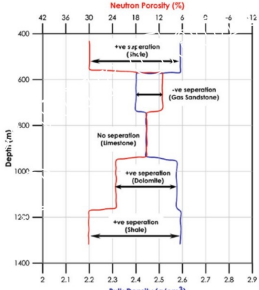
Step-by-step guidance:

- Scientific reading &writing
- Programming (Python)
- Knowledge modelling (OWL 2)

Further opportunities:

- Industrial internship
- Industrial PhD study

Lithologic logs	Fluid indicator logs
• Gamma-ray • Litho-density • Neutron-density	• Resistivity log • Induction log • NMR
Porosity logs	**Other logs**
• Neutron • Density • Sonic • NMR	• Caliper • Temperature • Dipmeter/FMI

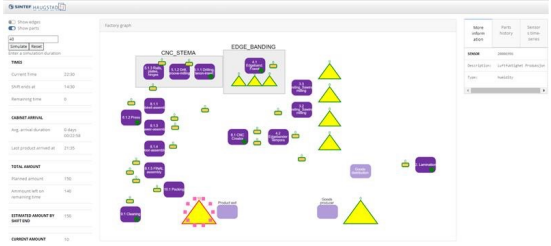


Supervisors:
Baifan Zhou
Yuanwei Qu
Evgeny Kharlamov
Jieying Chen

Contact: baifanz@ifi.uio.no

Knowledge Graphs-based real-time Digital Twins with analytics support

https://www.mn.uio.no/ifi/studier/masteroppgaver/asr/knowledge-graphs-based-real-time-digital-twins-wit.html

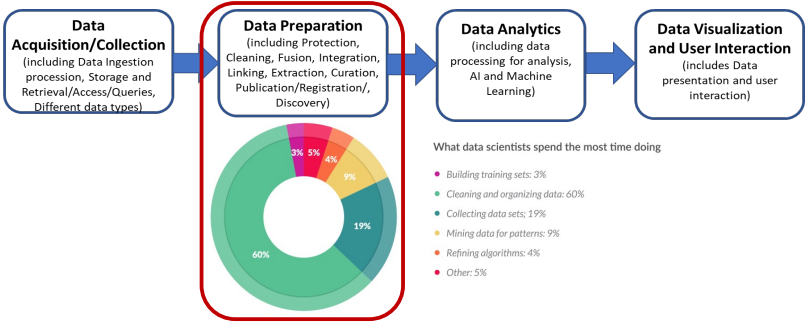


The thesis aims to build upon the existing prototype (see screenshot of the current UI above) and explore various techniques and tools for combining knowledge graphs representations of Digital Twins with real-time and analytics aspects, focusing on scalability and genericity of the solution. Examples of concrete tasks could include (but are not limited to):

- Identify and implement at least one feature for a Digital Twin for example time series analysis, optimizing resources in a manufacturing factory or prediction of machine failure.
- Identify and implement at least one feature to improve the quality of the existing prototype such as unit testing, continuous integration and deployment or logging.
- Evaluate implemented feature according to previously defined benchmarks.

Data prep

https://www.mn.uio.no/ifi/studier/masteroppgaver/asr/data-prep-|-data-pipelines.html



- **Data preparation** (also known as **data prep**, **wrangling**, **transformation**) is the most time consuming phase in data science projects
- Potential MSc theses topics: exploring and finding **intelligent mechanisms** to support data scientists in the data preparation phase, examples including: use of machine learning for intelligently suggesting data transformations, automated data quality assessment and recommendations for improving data quality, application of semantics and formal reasoning in the data preparation phase, support for data extension, enrichment and interlinking, use of (knowledge) graph representation and analytics techniques in data preparation

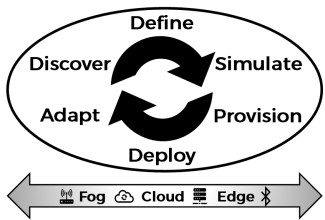
Data pipelines

https://www.mn.uio.no/ifi/studier/masteroppgaver/asr/data-prep-|-data-pipelines.html

Big Data pipelines are composite pipelines for processing data with non-trivial properties and characteristics, commonly referred to as the *Vs of Big Data.*

Example: 30GB file

Unzip
TSV to CSV
Split
Transform
Store



Potential MSc theses topics: exploring and finding mechanisms that support both data scientist and domain experts in the complete lifecycle of managing data pipelines, covering AI-driven data pipeline discovery, languages for data pipelines modeling, and techniques for simulation, deployment, and adaptation of data pipelines. Of particular interest could be support for data pipelines on the Computing Continuum (how heterogenous infrastructures such as cloud, fog, edge could be used to support the complete data pipeline lifecycle)

Declarative vs procedural data manipulation languages – state of the art analysis

Supervisors:

Prof. Ahmet Soylu (OsloMet, UiO) & Dr Francisco Martin-Recuerda (SINTEF)

Abstract:

Data transformation is a crucial and time-consuming phase in data analytics projects. A common approach to define the necessary transformations to produce data suitable for machine learning algorithms is using a data manipulation language (DML). A DML is a computer programming language used for adding (inserting), deleting, and modifying (updating) data stored in files or databases. A DML can be categorized in two main groups: declarative and procedural. A very popular example of declarative DML is the language SQL (Structured Query Language), which is commonly used to query and manipulate relational databases. Similarly, SPARQL is the SQL counterpart for RDF (graph) databases. With the adoption of Python in many data analytics projects, the library Pandas became one of the most successful examples of procedural DMLs. The aim of this thesis is to investigate and compare the most promising DMLs currently available, including declarative and procedural languages. The result of this analysis will be highly relevant for data scientists and data engineers. Several research and innovation projects will provide real examples for the analysis.

Keywords:

"Data Manipulation Language", "Data Engineering", "Python", "Pandas", "RDF", "SPARQL", "SQL"

Topics in Ontology-Based Information Systems

- Internal Supervisors: Dag Hovland, Leif Harald Karlsen, Titi Roman, Martin G. Skjæveland
- Topics with external supervision:
  - DNV

Engineering Ontology-Based Information Systems

- Conceptual Models (Ontologies) can drive software
- Relies on ontologies, mappings, databases, queries,...
- Complex artefacts, and complex interrelationships
- Difficult to develop and maintain

- Several topics, short or long theses, from very theoretical to very practial. Examples in following slides, but adaptions to fit the interest of the student is possible.

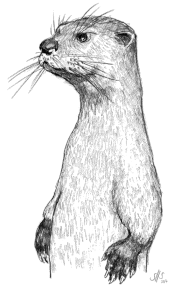
- The work will be carried out in close collaboration with ongoing efforts at Equinor and DNV and possible start-up company.

Supervisors: Martin G. Skjæveland, Dag Hovland,
Johan W. Klüwer (DNV GL)

Implement efficient updates for OTTR

Supervisor: Leif Harald Karlsen – leifhka@ifi.uio.no

- OTTR templates are macros for RDF and OWL
- Need the data on RDF-format (reasoning, queries)
- Also keep the OTTR-instances for easy updates of modelling
- Updating instances requires recompute full expansion of all instances
- Thesis: Investigate efficient ways of updating only the affected expanded RDF-graph



```
:Pizza[owl:Class ?pizza, xsd:string ?label] :: {
    o-rdf:Type(?pizza, owl:Class),
    o-rdfs:SubClassOf(?pizza, :NamedPizza),
    o-rdfs:Label(?pizza, ?label)
} .
```

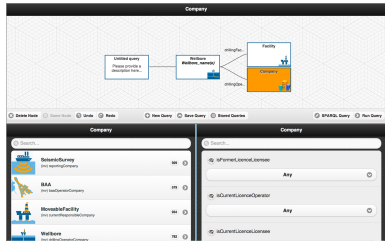
```
:Pizza(ex:margherita, "Margherita") .
:Pizza(ex:grandiosa, "Grandiosa") .
```

⇓

```
ex:margherita a owl:Class ;
    rdfs:subClassOf :NamedPizza ;
    rdfs:label "Margherita" .
ex:grandiosa a owl:Class ;
    rdfs:subClassOf :NamedPizza ;
    rdfs:label "Grandiosa" .
```

Topics in Query Formulation

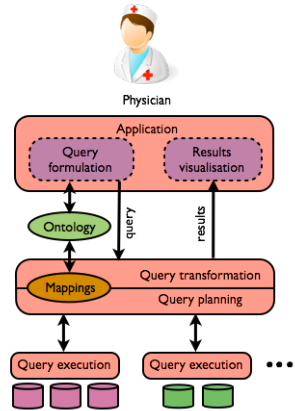
- OptiqueVQS allows building relational queries
- Controlled by an Ontology
- Currently 3 MSc students on this
- Lots to do, depending on your strengths!



- Supervisors: Vidar Klungre, Ahmet Soylu (SINTEF & OsloMet), Martin Giese

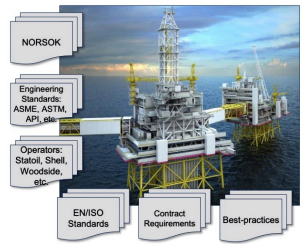
Big Data Access In Health

- Topic: Search Interfaces for Healthcare (Health Registry Data Use
- The student may focus on:
  - Semantic alignment of biomedical ontologies to databases
  - Integration of datasets for clinical research
  - UI design of search interfaces in ontology-driven information systems
  - semantic enrichment of datasets, including image annotations
- Supervisors: Laura Slaughter and Martin Giese



Knowledge Extraction from Large Scale Ontologies with Aibel

- Aibel: leading service company within the oil, gas and offshore wind industries
- MMD ontologies: in OWL 2.0, > 1.8m axioms
- Pains
  - Poor reasoning performance
  - Difficult to maintain
- Goal: exact relevant knowledge but still maintain logical guarantee
  - Apply theorical results in industry
  - Design a framework for MMD
  - More research possibilities



Supervisor: Jieying Chen (jieyingc@ifi.uio.no)

Ontology Diagnosis: Computing Justifications for OWL Ontologies

- Problems
  - Increasing number of large-scale ontologies are being developed
  - Difficult to debug ontologies and reuse their knowledge
- Goal: extract relevant ontology part to explain why certain conclusions follow from the ontology (axiom pinpointing / justification)
  - Extend JUST to deal with range restriction by using an extended calculus
  - Apply tools on real-world ontologies, e.g., MMD ontology from Aibel, SNOMED ontology
  - Supervisors: Jieying Chen (jieyingc@ifi.uio.no)
    Martin Giese (martingi@ifi.uio.no)

Topics in Modelling and Analysis

- Supervisors: Lizeth Tapia

Verification of Meta-controllers

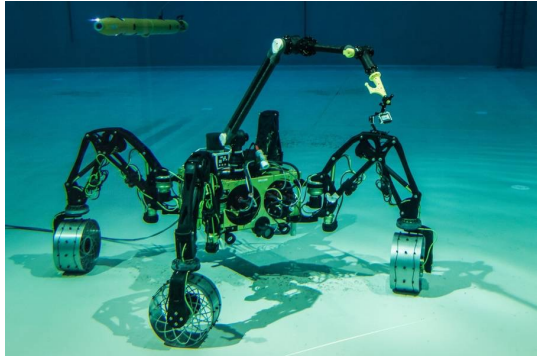
REMARO
RELIABLE AI FOR MARINE ROBOTICS

Marie Skłodowska-Curie Actions
Innovative Training Networks (ITN)

Meta-controllers are components which are in charge of recovery from faults and task-related contingencies. They continuously diagnose the current status of the different controllers of an autonomous system and reconfigure them depending on the current circumstances.

How can we guarantee that the reconfiguration process is only done when needed? E.g., reconfiguration does not start under normal circumstances, coverage of the failure detection process, successful reconfiguration, etc.

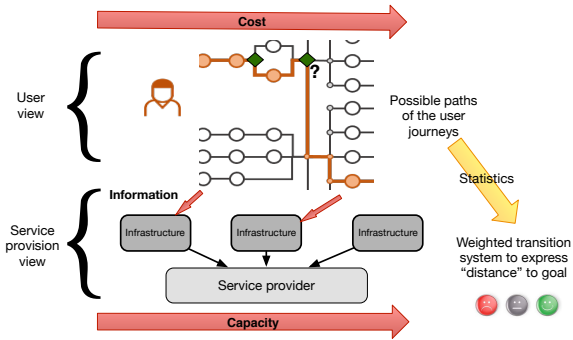
This master topic will contribute with the modelling of a meta-controller to facilitate reliability analysis.



Contact person:
Silvia Lizeth Tapia Tarifa
sltarifa@ifi.uio.no

Towards the digitalization of user-centric service processes

Smart Journey Mining



A user journey is the experiences a person has when interacting with different systems, typically digital services (e.g., online shopping, tax declaration, online booking, etc.).

In this research topic we will use logic-based methods towards the automatize analysis of the user experience when interacting with such services to guide users towards successful journeys.

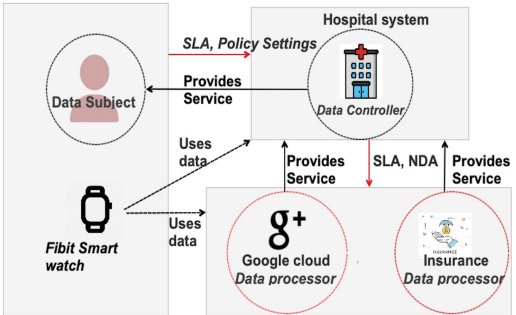
We plan to use executable modelling languages and their associated analysis tools to describe, predict, and prescribe user journeys as concurrent processes.

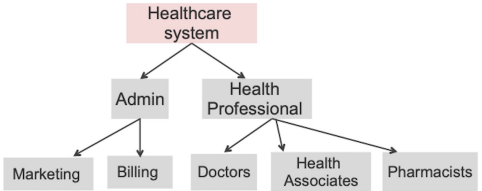
Contact person:

Silvia Lizeth Tapia Tarifa

sltarifa@ifi.uio.no

Privacy policies to Capture Privacy Concerns





Contact person:
Silvia Lizeth Tapia Tarifa
sltarifa@ifi.uio.no

This master topic is in the topic of **designing of a privacy policy language to express General Data Protection Regulations**. In this language, we define policies as sets of tuples with attributes that we use to capture privacy preferences at different level of granularity.

We plan for the master topic to contribute in one on the following aspects:

- Policy compliance between users and distributed services
- Static/runtime analysis to guarantee that the services are not violating the privacy preferences of users

Topics on Data Streams and Applications

- Supervisors: Thomas Plagemann, Vera Goebel, Espen Volnes

Privacy Protection in Data Stream Processing

Multiple theses
Details:
- Parrot project
- Video presentation at ASR
- MSc pages

Contact:
- Vera Goebel
- Thomas Plagemann

- Many privacy protecting mechanisms have been developed for database systems
- Data stream processing is becoming more and more important
- Privacy protection in stream processing is different
- Challenges:
  - Understand the differences
  - Adapt existing mechanisms to work for stream processing
  - Analyze stream processing systems for suitability to implement the mechanisms
  - Implement, experiment, evaluate
    - Measure protection level?
    - Measure data quality resp. information loss?

Sleep Apnea Detection with consumer electronics and AI

Multiple theses
Details:
- Cesar project
- Video presentation at ASR
- MSc pages

Contact:
- Vera Goebel
- Thomas Plagemann

- Consumer electronics:
  - Smart watches with oximeter
  - Sweetzpot.no sensor as respiration belt
  - Somnofy.no radar-based sensor

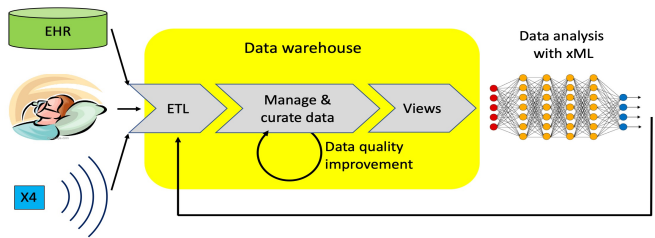
Sweetzpot.no

Somnofy.no

- Develop SW for sleep monitoring
- Collect data in lab and with patients
- Apply existing ML models for classification
- Train new models
- Improve ML solutions
- Evaluate

Data Warehouse

- Data sources:
  - Electronic health records (EHR)
  - Certified devices: polysomnography (PSG), polygraphy (PG), continuous positive airways pressure (CPAP), non-invasive ventilation (NIV)
  - Low-cost consumer electronics: respiration belt for athletes, radar-based sleep monitor
  - ML results, e.g.,augmentation
  - Starting point: MSc Thesis from Garth Theron in duo.no



Modeling execution of stream processing systems

- Testing distributed applications in a replicable way is challenging
- Simulation is a good way to start
- Smart cities, Industry 4.0, Smart health, Internet of Things
  - Distributed real-time stream processing
- Network simulators exist, but simulation models of systems lack

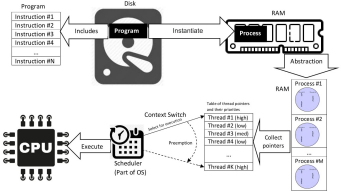


Figure: Process (Hooman Mallahzadeh, wikipedia)

Challenges:

- Understand what affects the temporal behavior
  - Application, OS, CPU caching, JIT optimization, background processes
- Create simulation models

One SQL to Rule Them All

- Real-time data processing can be done using streaming SQL
  - Smart cities, Industry 4.0, Smart health, Internet of Things
- Streaming SQL has yet to be standardized
- Several varieties exist and lack of common agreement
- Due to competing solutions, emerging technologies



Challenges:

- Understand existing varieties
- Express their features using a single language
- Convert from one SQL to existing streaming solutions
- Use in streaming simulator

Topics on Logic and Theoretical Computer Science

- Supervisors: Roger Antonsen, Henrik Forssell, Dag Hovland, Lars Kristiansen, Jens Otten, Jieying Chen

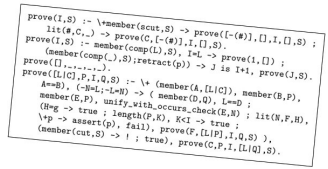
Master Topics in Logic & Automated Reasoning

- Logical reasoning is fundamental task, not only in mathematics and (computer) science, but in everyday life

- GOAL: automate logical reasoning
- based on formal mathematical logic
- one of the oldest core research fields in AI

Several Master Topics: develop tools to automate reasoning:

- based on ideas of the leanCoP/MleanCoP reasoning tools ("provers") for classical/modal first-order logic
- help lawyers to reason automatically given a legal text and its logical representation
- integrate machine learning into leanCoP
- specify requirements in multi-modal logic



```
prove(I,S) :- \+member(scut,S) -> prove([-(#)],[],I,[],S) ;
    lit(#,C,_) -> prove(C,[-(#)],I,[],S).
prove(I,S) :- member(comp(L),S), I=L -> prove(1,[]) ;
    (member(comp(_),S);retract(p)) -> J is I+1, prove(J,S).
prove([],[],_,_,_).
prove([L|C],P,I,Q,S) :- \+ (member(A,[L|C]), member(B,P),
    A==B), (-N=L;-L=N) -> ( member(D,Q), L==D) ;
    member(E,P), unify_with_occurs_check(E,N) ; lit(N,F,H),
    (H=g -> true ; length(P,K), K<I -> true ;
    \+p -> assert(p), fail), prove(F,[L|P],I,Q,S) ),
    (member(cut,S) -> ! ; true), prove(C,P,I,[L|Q],S).
```

Useful courses: "Logic", "Semantic Technologies", ...

Contact: Jens Otten (jeotten@ifi.uio.no ; www.jens-otten.de)

Topics by Lars Kristiansen

- He can supervise students on topics related to
  - logic,
  - computability theory and
  - complexity theory.
- Please, get in touch for more information.
- (The student should have taken at least one of the courses MAT-INF3600 and INF5840.)

A better grep

Improving search with regular expressions

- The standard implementations for searching with regular expressions (e.g. grep) have exponential behavior on some input.
- For example, using “numerical constraints” it is easy to make grep one-liners that consume over 8 GB of memory.
- There are polynomial-time algorithms for a subset of regular expressions with numerical constraints.
- Is it possible to combine the (usually extremely efficient) optimisations and extensions done in grep with algorithms that do not have exponential behavior?
- The work is theoretical and mathematical and you should have an interest in this.

Supervisor: Dag Hovland