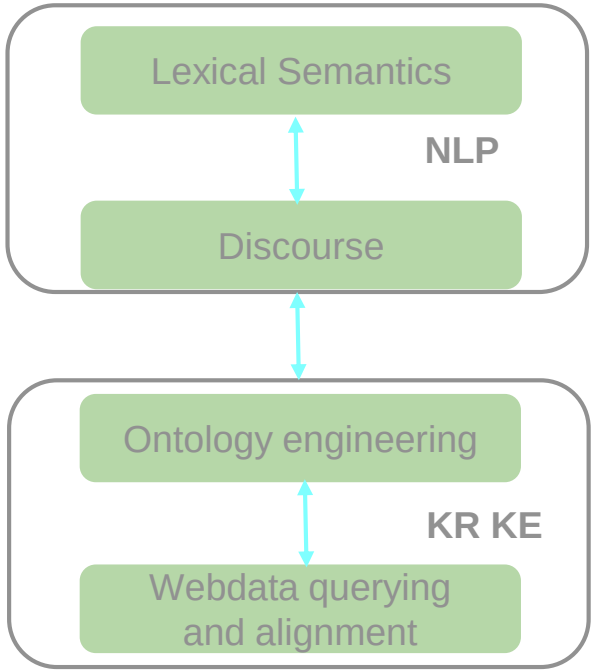
Models and Engineering for Language, Ontology and DIscourse

- Since October 2011
- 14 permanent staff
- 6 PhD. in progress
- 5 post-doctorate members

http://www.irit.fr/MELODI

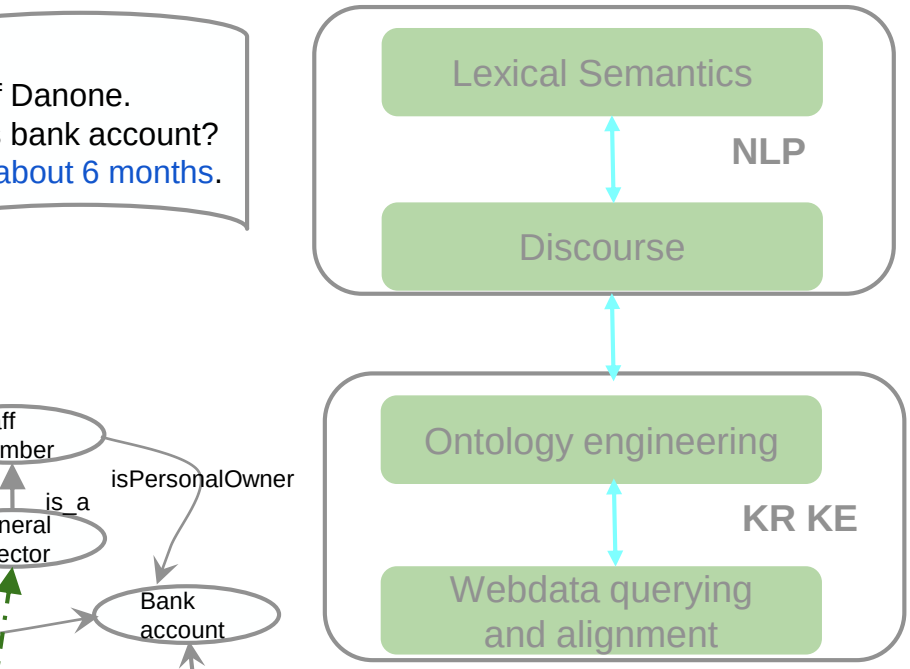
Philippe Muller
Nathalie Aussenac-Gilles

- Research perspective: to articulate linguistic data with knowledge
- 4 main areas in two research fields
  - Natural Language Processing (NLP)
  - Knowledge Representation and Knowledge Engineering (KR and KE)
- Complementarity: Theory and empirical validation in harmony

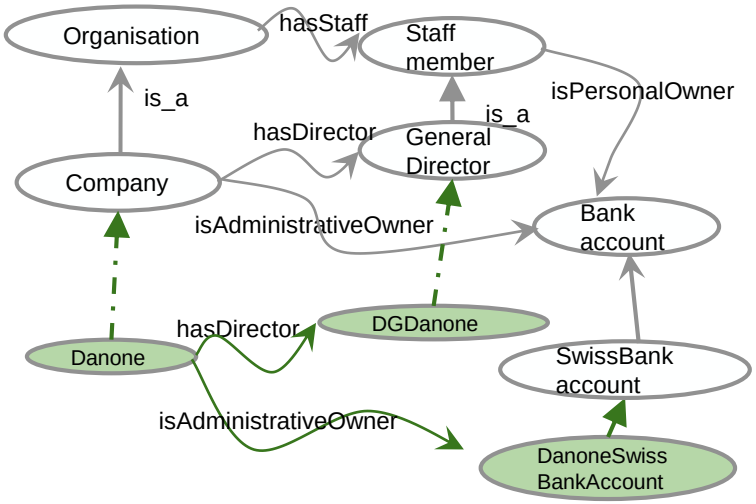


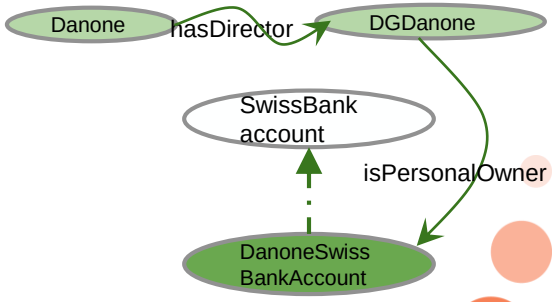
SCIENTIFIC ASSESSMENT

- I am the general director of Danone.
- Have you ever had a swiss bank account?
- The company had one for about 6 months.



Does the director of Danone own a swiss bank account?





SCIENTIFIC ASSESSMENT

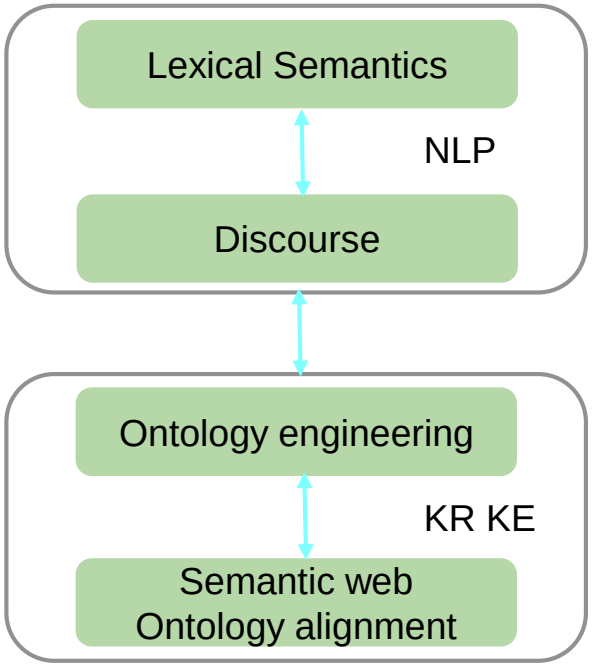
CROSS AREA THEMES

Opinion and preference analysis

Temporal information extraction

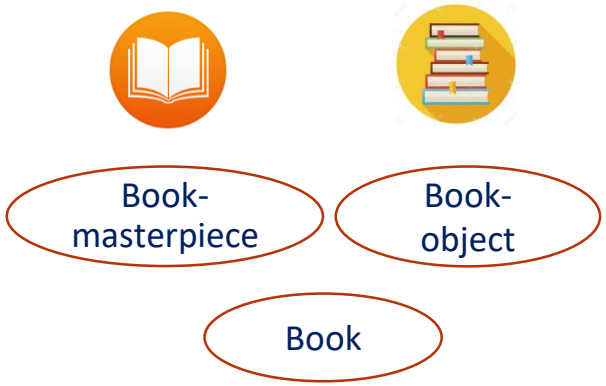
Semantic relation extraction and representation

Integration of lexical resources and ontologies



ONTOLOGY ENGINEERING (1)
Studies in Formal Ontology .

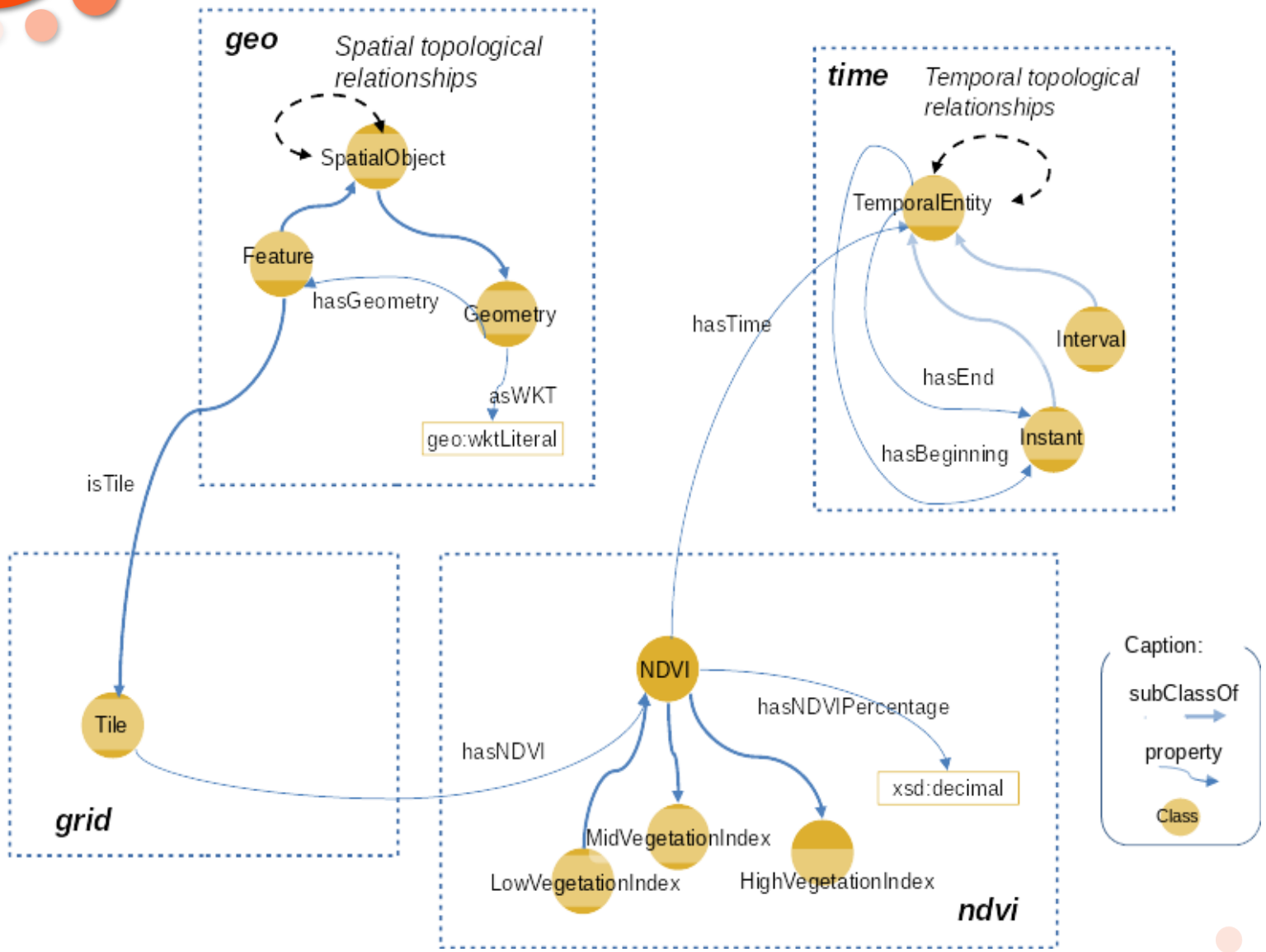
- Formal Ontology
  - foundational properties of conceptual models
  - explicit theorization of the types of entities and relations they contain
- Complex categories
  - ontological counterpart to systematic polysemy
  - characterized by mereological properties, constitution and coincidence
  - Graph-based representation
- From domain ontologies to high level categories
  - Alignments
- Formal ontology and ontology validation
  - Constraints and properties from formal ontology
  - Inference and natural language facts
  - to check domain knowledge



ONTOLOGY ENGINEERING (2)
From data to ontologies and KB .

- Ontology reuse : the MUSKA methodology
  - Ontology selection : trust and confidence
  - Ontology adaptation with design patterns
  - Ontology population with LOD data
  - An ontology for agriculture, collaboration with CIRAD, joint PhD supervision
- Ontologies of spatial and temporal data
  - Integration ontology: reuses OWL-Time and Geo-Sparql
  - Domain ontology for data and metadata semantic representation
  - 2 projects to link contextual data with Earth Observation images: SparkinData – CANDELA
- Ontology for IoT data
  - Integration ontology: IoT-O, an ontology for IoT data
  - Contribution to oneM2M standard, implementation proof of SOSA
- Change representation for spatial and temporal data > Ph.D at THALES
- Ontology of business processes
  - Adaption of BPMN model > AVIREX project
  - Formal ontology and process representation

Integration model for EO tiles and NDVI index



ONTOLOGY ENGINEERING (3)
From text to ontologies and KB .

- Extraction of Semantic relation from text
  - Relation extraction from semi-structured documents using layout and text architecture
    - From web forms to ontologies and knowledge bases: Moano project
    - Hypernymy extraction from vertical item lists
    - Best PhD. ATALA award
  - Combination of techniques including pattern matching and Machine Learning
    - Distant supervision using BabelNet
    - SemPedia project, collab. With CLLE-ERSS
    - Collab. with brasilian univ.
  - Open information extraction
    - deep learning and covolutional networks for relation extraction
    - Relation extraction from scientific scholar papers > RelTEX project
    - Combining patterns and supervised learning > Geotrend project

SEMANTIC WEB (1)
Ontology-based data integration

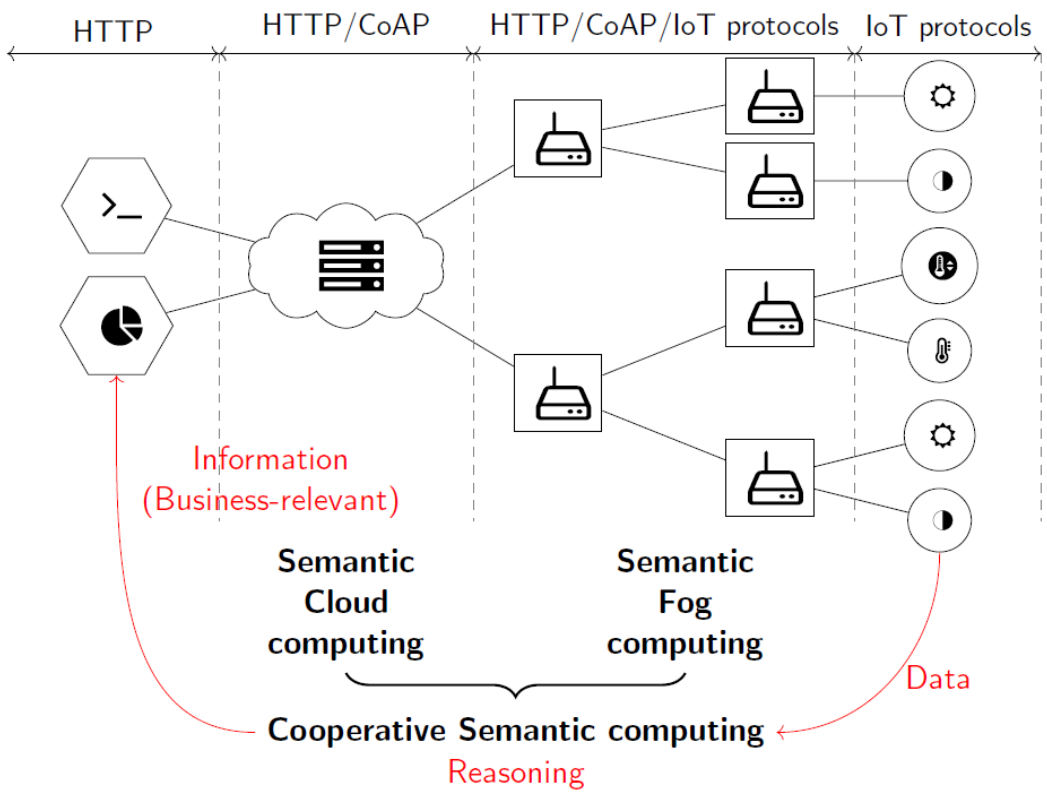
- Ontology-based integration of spatial and temporal data
  - Ontology for Earth Observation data integration
  - from open data to LOD repositories
  - Integration when searching the data with SPARQL queries
  - SparkinData – CANDELA

- Ontology-based integration of IoT data
  - Use IoT-O to represent sensor data collected in a connected apartment
  - Deployed in a system to fully manage the connected devices
  - FIESTA-IoT project

SEMANTIC WEB (2)

Ontology-based data integration

- Performance issues of semantic applications when scaling up with big data sets (IoT context)
  - Semantic Web of Things, collab with LAAS
  - distributed dynamic and opportunistic processing of sensor data
  - Semantic cloud computing



SEMANTIC WEB (3)
Ontology-based data integration

- Intuitive interfaces to query semantic datasets
  - Natural language querying of semantic web data > Swip
  - Ontology based semantic information retrieval
  - > distributed dynamic and opportunistic processing of sensor data (CANDELA)
  - Alignments to rewrite queries > PhD.
- Ontology alignment > Cassia Trojahn talk
  - Multilingual alignment, Complex alignments, Holistic alignments
  - Manipulation and visualization of alignments
  - Ontology alignment evaluation

NLP
LEXICAL SEMANTICS

- Models of formal lexical semantics using a rich theory of types
- Distributional semantics and methods of composition
  - Learning representations (word and sentence embeddings)
  - Automatic methods for semantic compositio n
- Towards hybrid models
- Construction of ontologically coherent semantic lexicons (French FrameNet, open knowledge base for Italian)

NLP
DISCOURSE

- Theoretical modelling
  - Strategic conversation using game theory
  - Topic and presupposition in discourse
  - Socio-pragmatic dimension of discourse relations
- Discourse parsing using
  - Structured prediction (Machine learning)
  - Layout features
- Analysis and extraction of specific knowledge
  - Temporal relations
  - Opinions and preferences (coll. ADRIA)
  - Causation

NLP
Projects in the medical domain

- Automatic detection of depressive behavior on social networks (Anr Project)
- Connecting medical data and social data with documents written by patients (NCDS cohort) (with INSERM)
- Analysis of care pathways for patients with Alzeihmer and similar diseases (with INSERM)

DIRECTIONS FOR THE FUTURE
Ontology and semantic web @ MELODI

- Coupling Formal and domain ontologies
  - Aligning high level and domain ontologies
  - Process ontology
- Ontology based Data integration
  - Change representation
  - Sensor data diagnosis thanks to semantics
  - Distributed reasoning over IoT
- Ontology alignment
  - Complex, holisitic alignments
- From language to ontologies and KB and back
  - Information extraction for ontology population
  - Relation extraction with distant or weak supervision and distributional representations (word / sentence embeddings)
  - Linguistic patterns for ontology alignment
  - Querying Linked Data

SCIENTIFIC DIRECTIONS FOR THE FUTURE

Stakes

- Diversification of methods and theories for discourse parsing and analysis
- The next generation of web analysis and search tools using deep semantic understanding

Scientific challenges

- Investigate further interaction and complementarity
  - Lexical semantics and ontologies
  - Discourse and dialog models
  - Hybrid distributional and formal semantic methods of composition
  - General linguistic information and domain specific information
- Extend results to integrate new data and technologies
  - Complex and cross-lingual ontology alignments for querying the web of data
  - New forms of linguistic communication (chats, blogs, fora, tweets, …)
  - Semantic web technologies to better exploit results from parsers and linguistic annotations