FACULTY OF ENGINEERING

OTAGen: A tunable ontology generator for benchmarking ontology-based agent collaboration

F. Ongenae, S. Verstichel, F. De Turck, T. Dhaene,
B. Dhoedt, P. Demeester,

July 28, 2008

Department of Information Technology - Broadband Communication Networks research group (IBCN)

Overview

- **Introduction**
  - Problem statement
  - Example
- **Related work**
  - Ontology technologies
  - Benchmarking tools
- **OTAGen**
  - Parameters
  - Workflow
  - Advantages
- **Future work**
- **Conclusion**

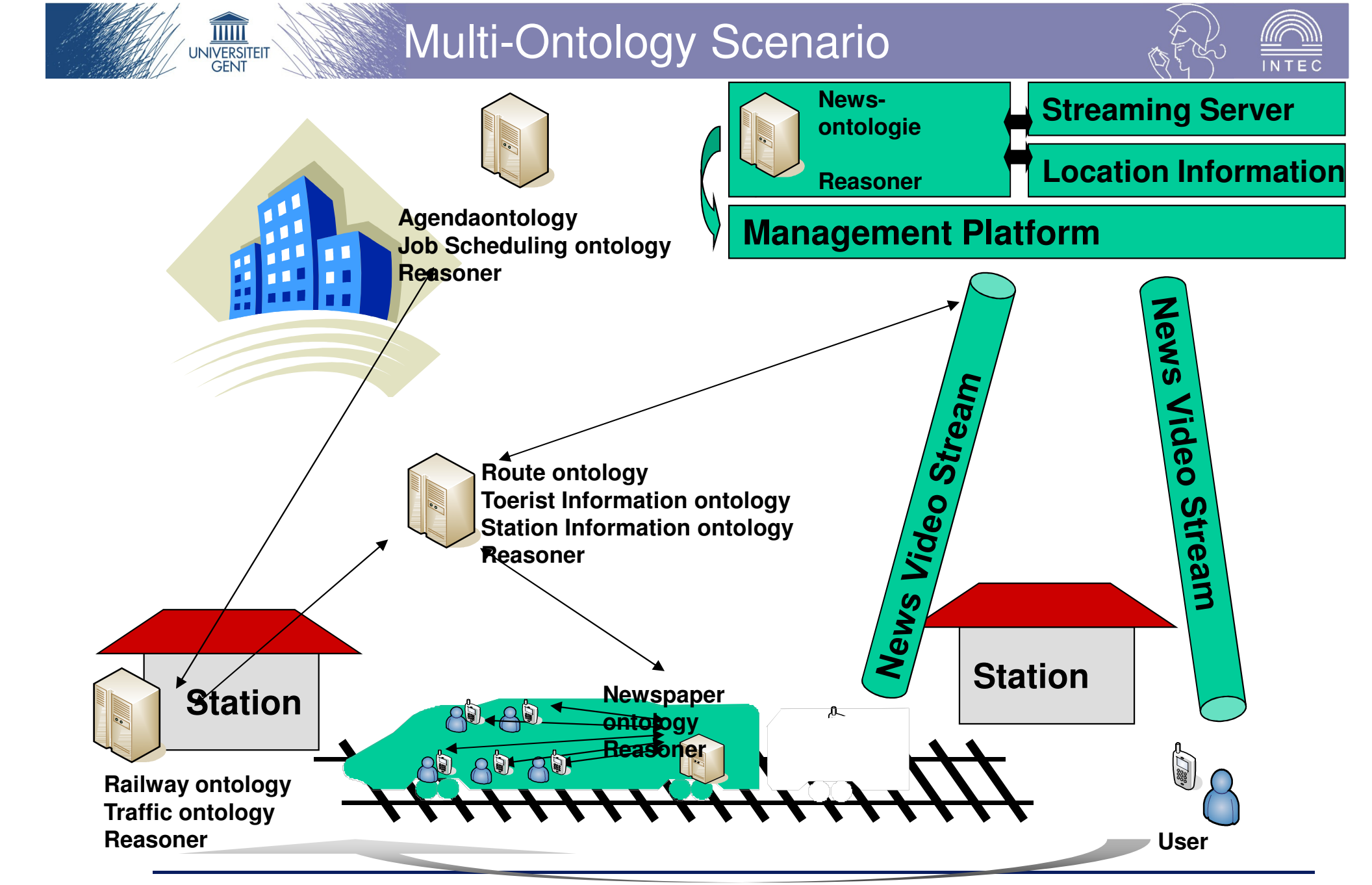
Problem statement

- **Development of a multi-agent framework (using ontologies) with various scheduling and monitoring algorithms**
  - Online repositioning algorithms
  - Repartitioning algorithms
  - Algorithms for query decomposition
  - …
- **"Islands" of ontologies**
  - Cannot use one test ontology → would introduce unnecessary correlations
  - Need for a large amount of ontologies
  - With varying complexity

Example:

- **Solution**: development of **OTAGen**, a highly tunable ontology generator
- A large amount of ontologies can be generated with **varying complexity** and **size**
- These ontologies can be used, to **test** and **benchmark** the multi-agent framework and the algorithms

Multi-Ontology Scenario



Overview

- **Introduction**
  - Problem statement
  - Example
- **Related work**
  - Ontologies
  - Benchmarking tools
- **OTAGen**
  - Parameters
  - Workflow
  - Advantages
- **Future work**
- **Conclusion**

Ontology

- **Goal: Formulate a complete and strictly conceptual model over a certain domain**
- **Describes the entities (e.g. Person), properties (e.g. Name) and relations (e.g. HasSibling)**
- **A strong formal ontology can be processed by a machine (queries, reasoning,…)**
- **2 parts:**
  - T-Box: Terminology layer
  - A-Box: Instantiation layer (data)
- **Application areas: Semantic Web, Context-Aware applications, Location Based Services,…**

Ontology: OWL

- **Ontology Web Language (OWL)**
- **Well-defined vocabulary for describing a domain**
- **Three sublanguages:**
  - OWL-Lite
  - OWL-DL
  - OWL-Full
- **OWL-DL: Foundation in Description Logics**
  → reasoning to check consisteny and infer new knowledge
- **Reasoning**
  → resource intensive and often time-consuming

Benchmarking tools: LUBM

- **Aim: benchmark Semantic applications and profile their behaviour with different sizes and complexity of the used ontology**

- **Lehigh University Benchmark (LUBM)**
  - A university domain ontology
  - T-Box statically defined
  - Includes a set of 14 queries
  - Size of A-Box can be specified and varied to generate different ontologies
  - Behaviour of the applications can be measured by executing the queries on the generated ontologie

Benchmarking tools: LUBM

- **Disadvantages:**
  - T-Box is static
  - T-Box covers only a subset of the OWL-Lite inference → many ontologies are more complex
  - The influence of the T-Box complexity on the reasoning/algorithms cannot be tested
  - Adding explicit knowledge to the A-Box does not add implicit knowledge
  - The generated data (A-Box) form multiple relatively isolated graphs and lack necessary links between them

Benchmarking tools: UOB

- **University Ontology Benchmark (UOB)**
  - Extension of LUBM
  - Consists of 2 ontologies:
    - **UOB-Lite: OWL-Lite constructs in the T-BOX**
    - **UOB-DL: OWL-DL constructs in the T-BOX**

- **Disadvantages**
  - Still a more or less static T-Box
  - Complexity of the T-Box cannot be varied (increased) across different tests

Overview

- **Introduction**
  - Problem statement
  - Example
- **Related work**
  - Ontology technologies
  - Benchmarking tools
- **OTAGen**
  - Parameters
  - Workflow
  - Advantages
- **Future work**
- **Conclusion**

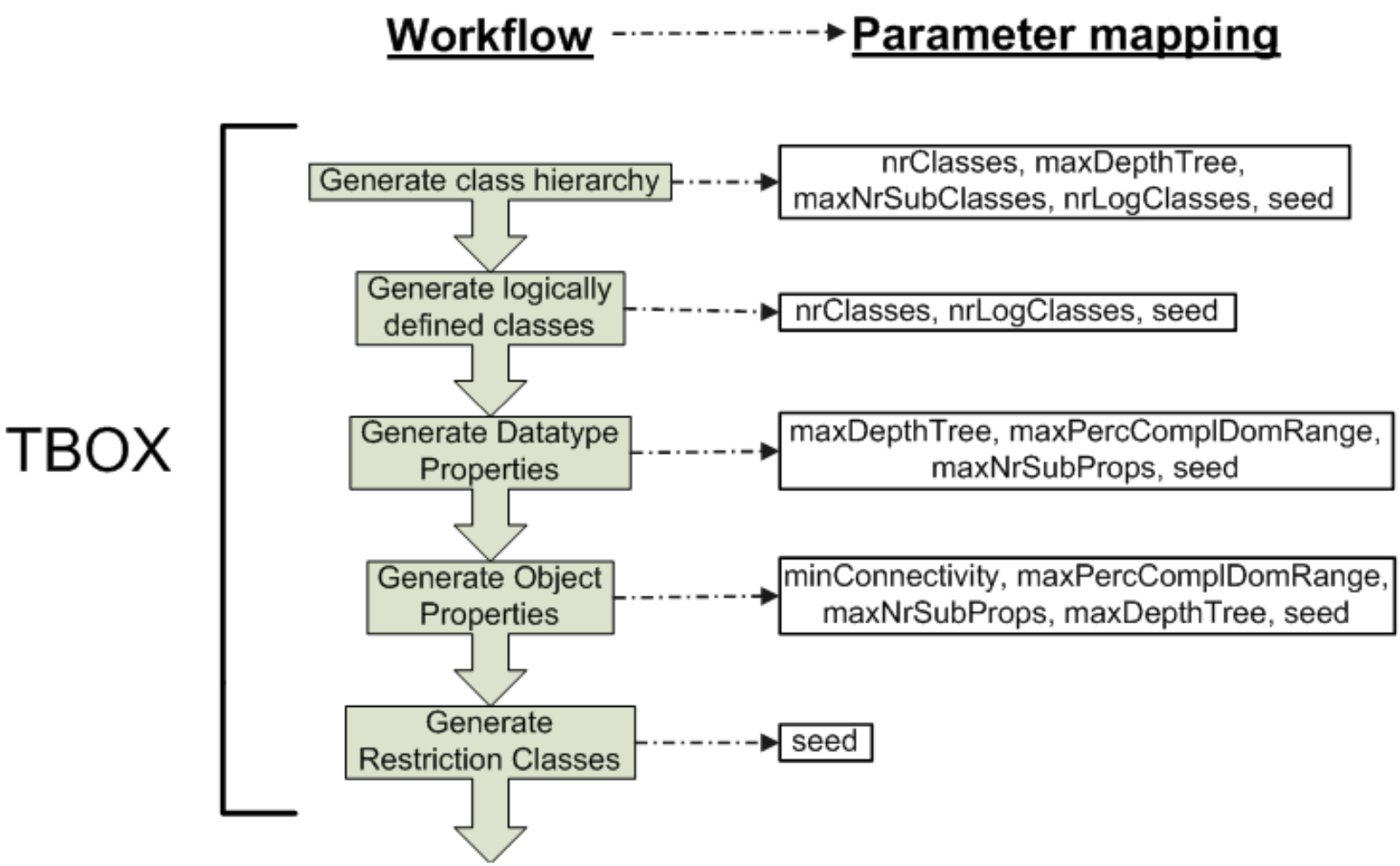
OTAGen: Introduction

- **Input**
  - User specifies parameters for the conceptual level (T-Box)
    e.g. nr. of (logical) classes, minimum connectivity,…
  - User specifies parameters of the instance level (A-Box)
    e.g. nr. of individuals, obj. prop. instances,…
  - User specifies characteristics of the queries
    e.g. the nr. of queries, their depth,…
  - This can all be inputted through a properties file
- **Output**
  - The T-Box (conceptual level) and A-Box (instance level) of a ontology are randomly and automatically generated
  - Some queries are generated for this ontology
- **A deterministic property is added to the generation process by using a seed**

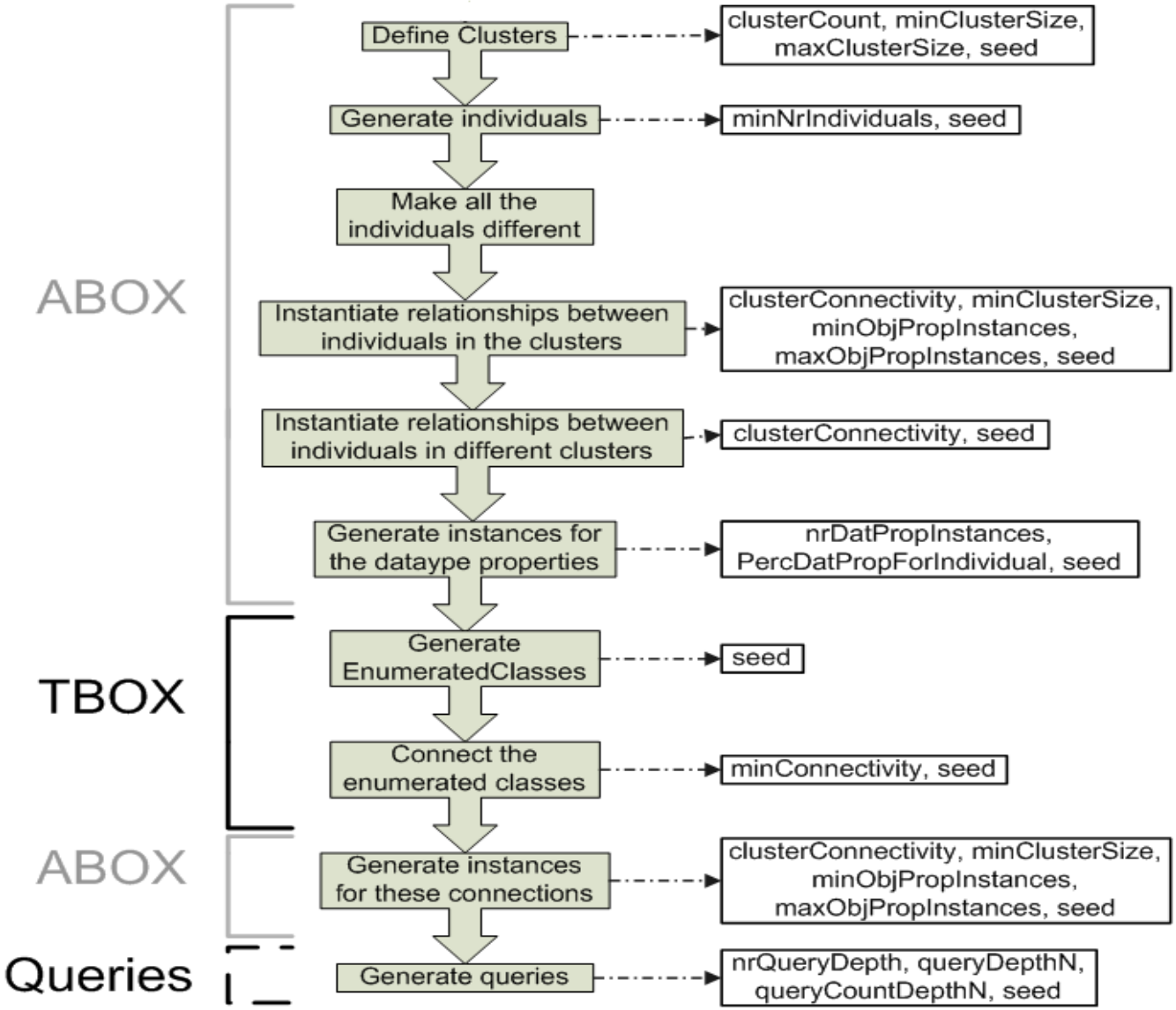
OTAGen: Parameters

Parameter:	Abbreviation:
The seed	seed
The number of classes	nrClasses
The maximal depth of the classes and properties tree	maxDepthTree
The maximal number of direct subclasses	maxNrSubClasses
The maximal number of direct subproperties	maxNrSubProps
The number of logically defined classes	nrLogClasses
The minimal connectivity	minConnectivity
The maximal percentage of object properties or datatype properties with complicated domain/range specifications	maxPercComplDomRange
The number of clusters	clusterCount
The minimum number of classes that have to be included in a cluster	minClusterSize
The maximum number of classes that have to be included in a cluster	maxClusterSize
Specifies the percentage of fully connected instances amongst the instances of the concepts in the cluster	clusterConnectivity
The minimum number of individuals	minNrIndividuals
The percentage of individuals with instantiated datatype properties	PercDatPropForIndividual
The number of times a non-functional datatype property is instantiated	nrDatPropInstances
The minimum number of times a non-functional object property is instantiated (intra-cluster)	minObjPropInstances
The maximum number of times a non-functional object property is instantiated (intra-cluster)	maxObjPropInstances
The number of query depths	nrQueryDepth
The depth of query set number n	queryDepthN
The number of queries generated for set number n	queryCountDepthN

OTAGen: Workflow



OTAGen: Workflow



OTAGen: advantages

Advantages

- A-Box can be gradually increased in size while the size and the complexity of the T-Box remains constant
- The connection degree of the A-Box can be varied to create a very connected or a sparse graph
- Adding explicit knowledge to the A-Box can possibly add a large amount of implicit knowledge (e.g. transitive properties)
- T-Box complexity can be gradually increased
- Includes all the OWL-Lite and OWL-DL inference constructs
- A set of queries with varying depth is generated for each generated ontology

Overview

- **Introduction**
  - Problem statement
  - Example
- **Related work**
  - Ontologies
  - Benchmarking tools
- **OTAGen**
  - Parameters
  - Workflow
  - Advantages
- **Future work**
- **Conclusion**

Future work

- **Initial studies have shown the same results as earlier studies**
  - OTAGen works
  - Ontologies are generated correctly
- **OTAGen will be used in the development of the multi-agent framework**
  - Provides a large variety of ontologies to test the algorithms on
  - Ontologies can be generated for the different “Islands”.

Overview

- **Introduction**
  - Problem statement
  - Example
- **Related work**
  - Ontology technologies
  - Benchmarking tools
- **OTAGen**
  - Parameters
  - Workflow
  - Advantages
- **Future work**
- **Conclusion**

Conclusion

- OTAGen: a highly tunable ontology generator
- An extensive number of parameters can be configured
- Can easily generate multiple ontologies with different properties
- Can be used to measure the performance and behaviour of various applications that use ontologies

UNIVERSITEIT GENT

FACULTY OF ENGINEERING

OTAGen: A tunable ontology generator for benchmarking ontology-based agent collaboration

Femke Ongenae, Stijn Verstichel

otagen@intec.ugent.be

Thanks for the attention!

Questions?