Domain Ontologies Integration for Virtual Modelling and Simulation Environments

Pavel A. Smirnov, Sergey V. Kovalchuk, Alexey V. Dukhanov
ITMO University, St.Petersburg, Russia
smirnp@niuitmo.ru, sergey.v.kovalchuk@gmail.com

Abstract

This paper presents a model of semantic ontologies integration into workflow composition design process via Virtual Simulation Objects (VSO) concept and technology. Domain knowledge distributed over open linked data sources may be usefully applied for new VSO-images design and used for organization computational-intensive simulation experiments. In this paper we describe the VSO-architecture extended with novel functionality regarding integration with linked open data sources. We also provide a computational-scientific example of domain-specific use-case offering a solution for some public-transportation domain problem.

Keywords: Virtual simulation objects, semantic technologies, computational experiment, knowledge base

1 Introduction

According to E-science paradigm computational simulations is field of science where simulation modeling results are considered to be scientific. Simulation modeling is typically characterized by computational-intensive tasks, described in form of workflow or separated tasks. Usually this kind of simulation involves scientists from different domains and requires geographically distributed resources like grids, supercomputers, parallel processing technologies and etc.

In our previous work we have proposed a concept and technology called Virtual Simulation Objects concept (Sergey V. Kovalchuk, Pavel A. Smirnov, Sergey S. Kosukhin, Alexander V. Boukhanovsky). The idea is to offer scientists to create virtual entities of real world objects and use them to construct virtual environment. Simulating objects together within the united environment allows to investigate their behavior and interference on each other. In (Sergey V. Kovalchuk, Pavel A. Smirnov, Sergey S. Kosukhin, Alexander V. Boukhanovsky) we also proposed a concept implementation – two web-applications (further we call it VSO-tools) based on CLAVIRE (CLoud Applications VIRtual Environment) platform. VSO-tools provide knowledge-based user-assistance to processes of virtual environments design and composite applications creation.

2   Backgrounds & related work

The idea of application semantic technologies to domain of E-science is not new. The most part of papers use semantics for organizing knowledge base (KB) to describe resource nodes, software packages, abstract web services and any other operational entities, required for workflow system functioning and control. Knowledge-base due to its reasoning mechanism facilitates complex experiments platforms with artificial intelligence functionality bringing automated code generation and data-validation mechanisms. The several papers describing platforms with semantic-based approach are presented in this section.

Composition Analysis Tool (Kim J., Gil Y., Spraragen M. Principles for interactive acquisition and validation of workflows) assists users in composing computational workflows. The tools uses KB aimed to represent workflow components in terms of input and output parameters, defined in ontology. Using the component ontology the tool automatically reasons about semantics of each parameters and component to helps user to specify valid dataflows designing computational experiments.

Knowledge Analysis project (Kim J., Gil Y., Spraragen M. Principles for interactive acquisition and validation of workflows) assists users in creating and validating process models in domain of biology. The system operates with KB based on domain-independent upper level ontology containing about 80 semantic relations and about 500 generic concepts of elementary entities of processes and actions. Ontology reasoning with AI techniques is used for automatic modeling, validation during workflow experiment design process.

WINGS (Gil Y. et al.) proposes semantic representations of workflow templates - a high-level reusable analysis methods for computational experiments. Knowledge base is filled with datasets and software components described via OWL, RDF and SWRL formants. Automatic reasoning is used for providing user assistance during constructing workflow-based computational experiments.

Workflow-centric research objects (Belhajjame K. et al.) (the project also known as myExperiment or Wf4Ever) uses ontology to store experiment packs, containing everything regarding particular computational experiment: workflows, documents, datasets, and etc. KB is build according Web 2.0 and Linked data principles in order to make scientific experiments reproducible for other scientists. Workflows are presented by WF-templates consisting of abstract tasks.

Paper (Highly dynamic workflow orchestration for scientific applications) proposes usage of Semantic Web standards and RDQL query language to accumulate ontology-transcribed knowledge to fill the ontololgy, alterable by domain expert that models application-related knowledge.

As we can see, most of the papers describe the platforms that fulfill their knowledge-bases with self-produced semantic-annotated content and further reuse it for some goals (workflow composition, experiment reproduction and reuse, etc). The key contribution of out paper is an idea to reuse already existing semantic-annotated knowledge extracted from third-party linked data sources. As a result, new high-level entities for workflow composition will be obtained. This idea brings valuable domain knowledge about entity's internal structures and external dependencies which are highly relevant for scientific computations design process according to VSO-concept (Sergey V. Kovalchuk, Pavel A. Smirnov, Sergey S. Kosukhin, Alexander V. Boukhanovsky).

3   VSO architecture extension

The proposing architectural extension is regarding an ability to operate with existing domain ontological models in order to use them for Virtual Simulation Object design process. Existing linked open data sources (like DBpedia, YAGO, SUMO, etc.) contain lots of ontological models, classifications and hierarchies, which may be benefit to use during VSO-design process. Domain-specific knowledge extracted from data sources published elsewhere can be used as basis and extended

with new properties, relations, simulation models. Such two-side interaction is benefit from computational science and ontology engineering points of view.

Before describing architectural extension implementation we briefly introduce architecture of already existing solution. The Virtual Simulation Objects concept is based according Intelligent Problem Solving Environment (iPSE) and implemented as an extension to the CLAVIRE platform, service-oriented architecture. Functionally the platform allows users to create, run and control computational-intensive workflows using clusters, grids, supercomputers as resources. Technically CLAVIRE consists of set of interconnected subsystems (most part of them implemented as WCF-services) and existing VSO architecture integrated with package base, data storage, workflow execution and monitoring services (see fig.1). VSO concept has been implemented in two web-applications (VSO construct, VSO class editor) and a set of logical units (VSO knowledge base, Graph processing service) partially integrated into web-apps as compiled .Net libraries.

We have extended VSO architecture with VSO-mapping service (see fig. 1). VSO-mapping service is logically new unit aimed to provide functionality of knowledge extraction, filtration and further usage. Knowledge extraction can be done via querying SparQL-endpoints or parsing data from user-uploaded file. Knowledge filtration means selection only necessary entities and relations and hiding other existing-ones. Knowledge mapping is mapping particular entities and their relations to be processed in some way.
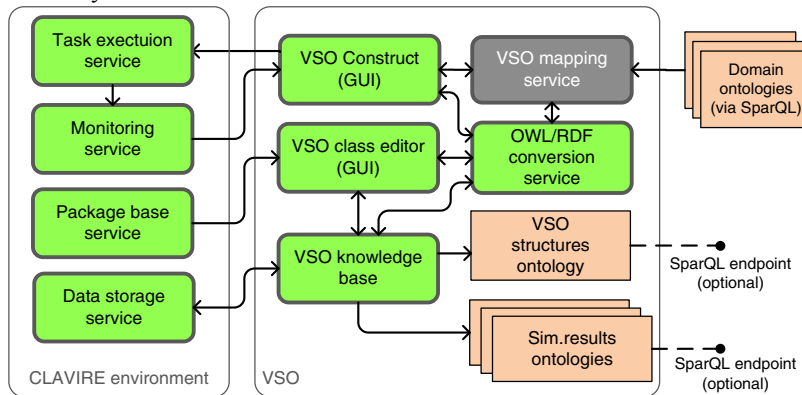

Figure 1: Extended VSO-architecture

Particularly knowledge mapping to during VSO-image design aims to select ontological entities (according particular problem domain) to be mapped on existing entities within VSO-knowledge base. In other words, user uploads some ontological model (class hierarchy) and VSO-mapping service tries to find any matched within existing knowledge-base, consisting of objects ontology and package-base service. Found matches are offered for user confirmation and newly generated virtual object structure inherits relations, selected by user from ontological model. VSO-mapping service will be used through VSO-editor application.

The example of mapping effectively illustrated through domain example, presented in the next section.

4 Case study

The main objective of the section is to illustrate, how Virtual Simulation Objects extended with semantic data support can be applied to automatic workflow generation. We describe a solution of domain-specific problem from end-user's point of view using extended VSO-tools.

The particular problem is devoted public transport schedule optimization. Before any detail descriptions will take place the literature overview should been presented. Within domain of public transportation networks there are several papers based on ontological modeling and semantic web technologies application. For example, paper (Lorenz B., Ohlbach H. J., Yang L.) presents a general ontology of public transportation networks. The ontology is based on GDF (geographic data files) data model and conforms to ISO standards. Ontology classifies city infrastructure objects according separating several layers (Geometric, Feature, Composite attributes and etc.). Several classes from transport layer (PublicTransport class) of the ontology have been used for the use case. Paper (Wang J., Ding Z., Jiang C.) proposes an ontology of transportation system, public transport query algorithm and service implementation based on algorithm. We have used some classes from the ontology building and example of the use case. To conclude short literature overview it is worth to note, that semantic ontology format is rather preferable for urban modeling and simulations. There are also papers devoted to urban systems planning (Gomes J. et al.) and development (Schevers H., Trinidad G. S., Drogemuller R.), city objects visualization (via SVG (pfelkofer F., Lorenz B., Ohlbach H. J.), CityGML (Métral C. et al.)) and etc.

Particular use case is based on attempt to optimize public transport schedule within one of several districts of city Saint-Petersburg (Russia). The bus scheduler optimization is done by genetic algorithm and depends on dynamically changing traffic jams level. Further description will present application of the extended Virtual Simulation Object architecture for the domain problem solution.

As any scientific experiments, the particular use case solution forms a composite application presented as workflow, which executes on distributed computational infrastructure, managed by some execution platform (in our case this is CLAVIRE platform). Initially the workflow was created and by urban scientist and include execution of 4 different software packages, instantiated on computational resources of the platform. Existing VSO-Construct tool already allows user to generate workflow automatically by filling virtual environment with virtual object instances, configuring execution parameters and data flows between them though GUI. To work in this manner, user (being a domain specialist) should defines VSO-image thought GUI-based VSO-editor tool to provide high-level names and descriptions for object parameters, models, methods (scenarios) and implementations. VSO-image creation should be done once, after that image will be stored in knowledge base and automatically extended during further manipulations via VSO-Editor and VSO-Construct.

4.1   VSO design

End-user, being a domain-specialist is aware about data, models, methods used in his package. In case, then a package may (or requires) to process data from semantic data source, it would be enough just to map entities from ontological model on corresponding VSO-image fields. For example, we have an OTN-ontology with a least two classes (StopPoint, Route) which may to be used for definition a VSO-image for transportation modeling simulations (see fig. 2). For the experiment we have to use some more classes (passenger, bus and timetable), which were found in Wordnet-ontology (two last may be also found in ontology of transportation system from (Wang J., Ding Z., Jiang C.)) and compiled into custom ontological model. For this experiment compilation was done via Protégé editor, but for further releases we plan to integrate it into VSO-editor and give to user an ability to uploading and process several ontologies instead of only one.
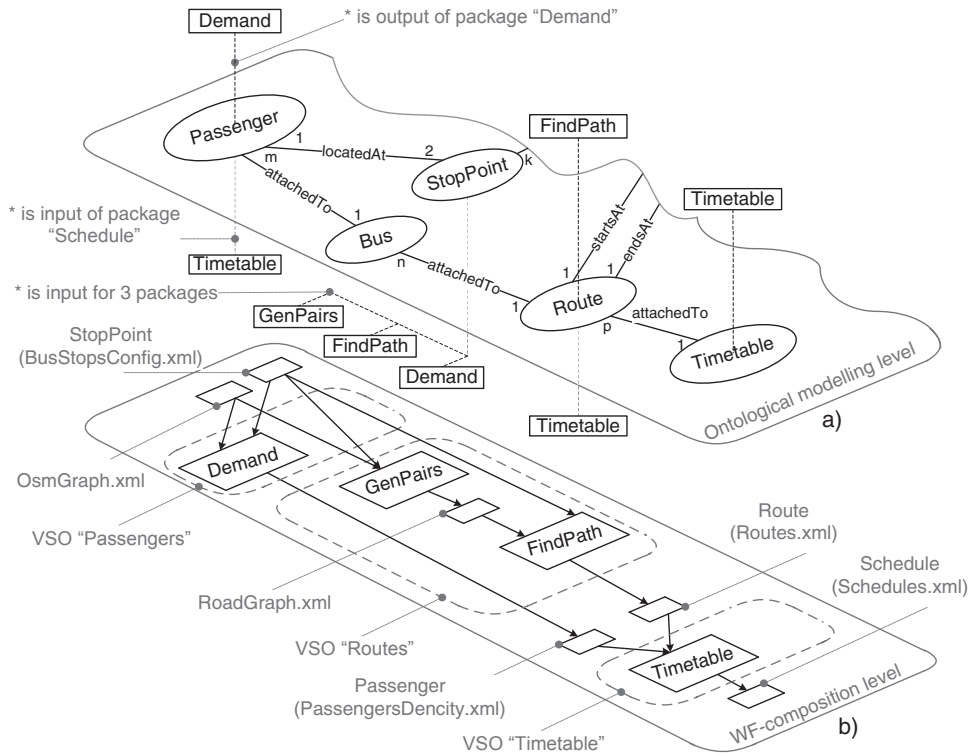
Figure 2: Mapping ontological models (a) onto VSO-level (b)

Then ontological model (see fig. 2a) was compiled and uploaded to VSO-Editor, a list of its entities should be extracted and scanned in order to find any matches with entities existing in knowledge base. At this stage match-searching is done only due to equivalence by name within package definitions in CLAVIRE package base engine. For future we plan to organize be semantic-based search through ontology of existing VSOs. As we can see (see fig. 2a) input and output parameter with names equivalent to ontological entities have been found within several packages (Demand, Timetable, GenPairs, FindPath). All the matches presented as a table are offered for user to be validated and confirmed.

Then matching packages were found, VSO-Editor automatically builds a sequence of packages in form of graph (see fig. 2b) with dataflows between them. To do that VSO-Editor uses the same algorithm, which has been already implemented to automatically compose objects sequence in VSO-Construct application. The result of operation is workflow-template, which is necessary to be validated by user.

Then workflow-template is composed, it may to seem, that final abstract workflow already may be generated. But user should not to forget, that this is only VSO-image in VSO-editor application instead of VSO-instance with particular parameters configuration, selected scenario, implementation, like it looks in VSO-construct application. More than that image-definition is still not finished yet and it cannot be used, like it is used within VSO-construct. At this stage user should group workflow-template steps according to his preferences about their semantic meaning. The example of this grouping is presented on Figure 3, where package Demand is defined as VSO called Passengers, packages GenPairs and FindPath should be defined as two models (according to VSO class hierarchy (Sergey V. Kovalchuk, Pavel A. Smirnov, Sergey S. Kosukhin, Alexander V. Boukhanovsky)) within VSO called Routes, Timetable package may be defined as VSO called Timetable. VSO-editor

functionality performing these operations in user-friendly way is now been developing, but objects that are necessary to run simulation experiment have been defined programmatically.

4.2   Execution & results

As we have mentioned earlier, virtual simulation objects, necessary to build an experiment via VSO-concept have been defined manually and instantiated in VSO-Construct application (see fig. 3). Dataflows between object models have been defined by application automatically and model parameters have been set manually. Automatically generated workflow-script is a result of user's work in VSO-Construct application. The workflow-script has been to be exported as new project in CLAVIRE platform, where its execution have been started.
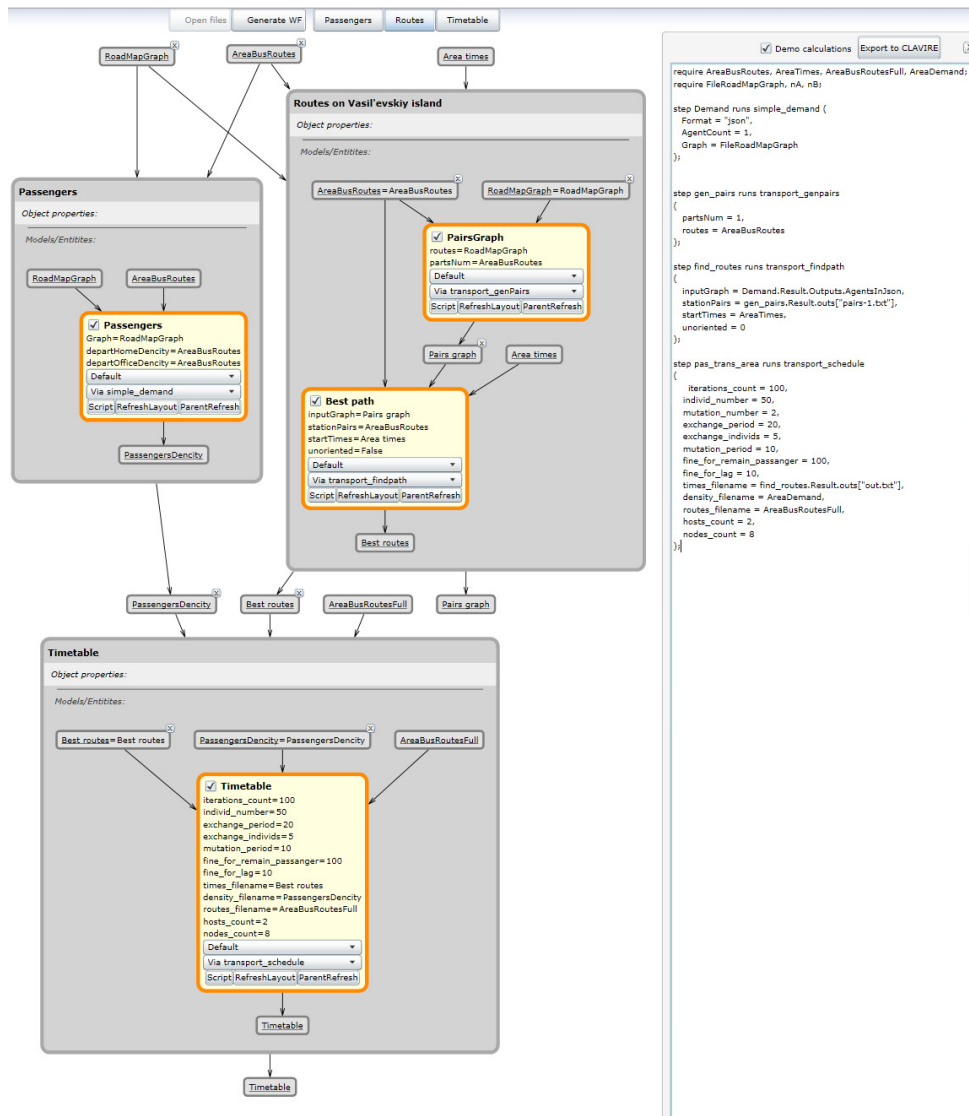

Figure 3: Screenshots of VSO-construct application with 3 objects and generated WF

The workflow execution results perform a timetable optimized with genetic algorithm. For every bus on every route within selected district the timetable declares start times, depending on dynamically changing traffic loading on roads. Optimized timetable was visualized and compared with data of non-optimized timetable (there buses starting periodically with fixed time interval). Comparison diagram is presented on Figure 4. Vertical axis of diagram presents sum of passengers on all bus stops at the moment of time (presented as horizontal axis). Red line (contains maximum verical values) on diagram presents amount of people waiting on bus stops with standard (non-optimized) timetable. Blue line presents the genetically optimized timetable. Amount of refused people (who cannot be transported because current bus is full and continue to wait for the next bus) according optimized algorithm is presented as dark blue line near horizontal axis of the diagram. Blue dots – bus star times according to optimized algorithm.
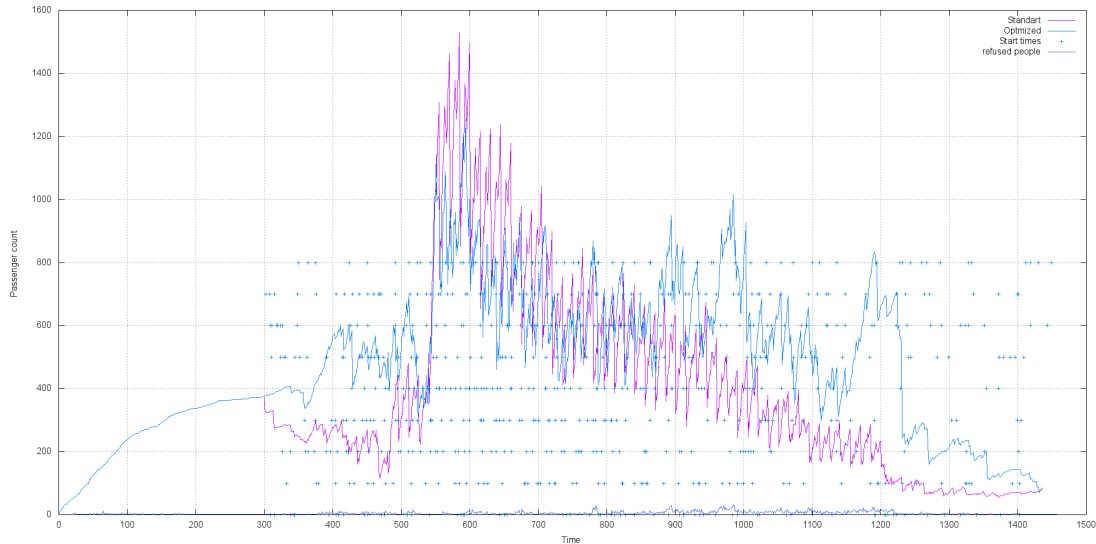

Figure 4: Efficiency comparison of genetic-optimized and standard time-interval timetable

The diagram shows, that optimized timetable shows increased frequency of bus starts during peak in the morning (7-11 am) and decreased frequency for left day (including evening the peak, which is not so explicit as morning).

5 Conclusion and future work

In this paper we have investigated the idea of domain knowledge integration from linked data sources and described architectural extension for Virtual Simulation Objects concept implementing this idea. Such approach allows to cross workflow composition entities (i.e. VSO-images) from computational intensive scientific simulation field with light-weight entities from ontological modeling field in order to use domain knowledge distributed over existing open linked data sources.

For the future investigations we are looking towards automatic semantic formalization of computationally-generated data. To some extent the idea intersects the field of the provenance data mining and management (Missier P. et al.) (Sahoo S. S., Sheth A., Henson C.). We are thinking not about provenance graph annotations, but about automatic production stand-alone semantic entities, performing the result-sets with their properties and relations in order to make them operable as new VSO-images and queryable as any other entities from existing linked data sources.

References

Belhajjame K. et al., Workflow-centric research objects: first class citizens in scholarly discourse. *//Proceedings of the ESWC2012 Workshop on the Future of Scholarly Communication in the Semantic Web. – 2012.*

Gil Y. et al., Wings: Intelligent workflow-based design of computational experiments . *//Intelligent Systems, IEEE. – 2011. – T. 26. – №. 1. – C. 62-72.*

Gomes J. et al., A computer-aided urban planning tool driven by semantic web ontologies. *//Information Systems and Technologies (CISTI), 2012 7th Iberian Conference on. – IEEE, 2012. – C. 1-6.*

Highly dynamic workflow orchestration for scientific applications. , Gubala T., Hoheisel A. *//CoreGRID Intergation Workshop 2006 (CIW06). – 2007. – C. 309-320.*

Kim J., Gil Y., Spraragen M. Principles for interactive acquisition and validation of workflows. , Principles for Interactive Acquisition and Validation of Workflows. *//Journal of Experimental and Theoretical Artificial Intelligence. – 2010. – T. 22. – №. 2. – C. 103-134.*

Lorenz B., Ohlbach H. J., Yang L. , Ontology of Transportation Networks. *– 2005.*

Métral C. et al., Ontology-based approaches for improving the interoperability between 3d urban models. *//Journal of Information Technology in Construction. – 2010. – T. 15.*

Missier P. et al., Janus: from workflows to semantic provenance and linked open data. *//Provenance and Annotation of Data and Processes. – Springer Berlin Heidelberg, 2010. – C. 129-141.*

pfelkofer F., Lorenz B., Ohlbach H. J. , Ontology driven visualisation of maps with svg – an example for semantic programming. *//Information Visualization, 2006. IV 2006. Tenth International Conference on. – IEEE, 2006. – C. 424-429.*

Sahoo S. S., Sheth A. . , Provenir ontology: Towards a framework for escience provenance management. *– 2009.*

Sahoo S. S., Sheth A., Henson C. , Semantic provenance for escience: Managing the deluge of scientific data. *//Internet Computing, IEEE. – 2008. – T. 12. – №. 4. – C. 46-54.*

Schevers H., Trinidad G. S., Drogemuller R. . , Towards integrated assessments for urban development. *//ITcon-Electronic Journal of Information Technology in Construction-Special Issue Decision Support Systems for Infrastructure Management. – 2006. – T. 11. – C. 225-236.*

Sergey V. Kovalchuk, Pavel A. Smirnov, Sergey S. Kosukhin, Alexander V. Boukhanovsky. , Virtual Simulation Objects Concept as a Framework for System-Level Simulation. *// E-Science (e-Science), 2012 IEEE 8th International Conference on. – IEEE, 2012. – C. 1-8.*

Wang J., Ding Z., Jiang C. . , An Ontology-based Public Transport Query System. *//Semantics, Knowledge and Grid, 2005. SKG'05. First International Conference on. – IEEE, 2005. – C. 62-62.*