Optique 1.0: Semantic Access to Big Data[*]
The Case of Norwegian Petroleum Directorate's FactPages

E. Kharlamov[1],[**], M. Giese[2], E. Jiménez-Ruiz[1], M. G. Skjæveland[2], A. Soylu[2],
D. Zheleznyakov[1], T. Bagosi[3], M. Console[5], P. Haase[4], I. Horrocks[1],
S. Marciuska[3], C. Pinkel[4], M. Rodriguez-Muro[3], M. Ruzzi[5], V. Santarelli[5],
D. F. Savo[5], K. Sengupta[4], M. Schmidt[4], E. Thorstensen[2], J. Trame[4], and
A. Waaler[2]

[1] University of Oxford, UK; [2] University of Oslo, Norway;
[3] Free University of Bozen-Bolzano, Italy; [4] fluid Operations AG, Germany;
[5] Sapienza Università di Roma, Italy

Abstract. The Optique project aims at developing an end-to-end system for semantic data access to Big Data in industries such as Statoil ASA and Siemens AG. In our demonstration we present the first version of the Optique system customised for the Norwegian Petroleum Directorate's FactPages, a publicly available dataset relevant for engineers at Statoil ASA. The system provides different options, including visual, to formulate queries over ontologies and to display query answers. Optique 1.0 offers installation wizards that allow to extract ontologies from relational schemata, extract and define mappings connecting ontologies and schemata, and align and approximate ontologies. Moreover, the system offers highly optimised techniques for query answering.

1 Introduction

Accessing the *relevant* data in Big Data scenarios is increasingly difficult both for end-user and IT-experts, due to the *volume*, *variety*, *velocity*, and *complexity* dimensions of Big Data. This brings a high cost overhead in data access for large enterprises. For instance, in the oil and gas industry, engineers spend 30–70% of their time gathering and assessing the quality of data. The Optique project[1] [1, 2] advocates for a next generation of the well known *Ontology-Based Data Access* (OBDA) approach to address the data access problem. The project aims at solutions that reduce the cost of data access dramatically. In our demonstration we present the first version of the Optique system which we customised for the Norwegian Petroleum Directorate's (NPD) FactPages.[2]

OBDA systems address the data access problem by presenting a general ontology-based and end-user oriented query interface over heterogeneous data sources. The core elements in a classical OBDA systems are an *ontology*, describing

---

[**] Corresponding author: `evgeny.kharlamov@cs.ox.ac.uk`
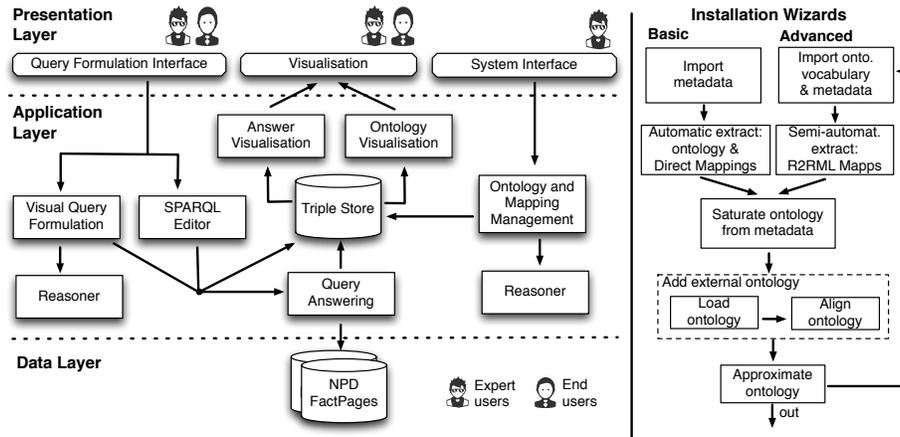[1] `http://www.optique-project.eu/`
[2] `http://factpages.npd.no`

Fig. 1. *Left*: General architecture of the Optique 1.0 system; *Right*: installation process

the application domain, and a set of *mappings*, relating the ontological terms with the schemata of the underlying data sources. End-users formulate queries using the ontological terms and thus they are not required to understand the structure of the data sources. These queries are then automatically translated using the ontology and mappings into an executable code over the data sources.

State of the art OBDA systems, however, have shown among others the following limitations:

- The *usability* of OBDA systems is hampered by the need to use a formal query language. Even if the users know the ontological vocabulary, they may find difficult to formulate queries with several concepts and relationships.
- The *prerequisites* of OBDA, i.e., ontology and mappings, are in practice expensive to obtain. Additionally, they are not static artefacts and should evolve according to the new end-users' information requirements.
- The *efficiency* of the translation process and the execution of the queries is usually not sufficiently addressed in OBDA systems.

The first version of the Optique system, i.e., Optique 1.0, aims at partially overcoming the above limitations. Demonstration videos are available at following address: `http://www.cs.ox.ac.uk/isg/projects/Optique/demos/iswc2013/`.

2 System Overview

A general three-layer architecture of the Optique system is depicted in Figure 1 (*Left*). The current version of the system offers two main functionalities: to query/visualise data and install/maintain the ontology and the mappings. At the backend, the system also offers an efficient query processing mechanism.

Optique 1.0 allows to pose queries via a visual query formulation (VQF) interface, a SPARQL editor, or from a query catalog. VQF exploits reasoning in order to show both explicit and implicit domain knowledge to guide the formulation of the query.

Queries are executed by the Query Answering module based on Ontop system.[3] Ontop provides functionalities for rewriting SPARQL queries using the system's ontology and mappings, syntactic and semantic query optimisation, and query unfolding. Thus, high efficiency of query answering is guaranteed. Rewritten and unfolded queries are in SQL and they are executed over the NPD FactPages data, which is stored in a relational database. The query answers are converted into triples in order to confirm the format of the system's ontology, temporally stored in the system's triple store, and displayed to the user in a tabular way or on maps (using OpenStreetMap).

The installation and maintenance of the ontology and the mappings is done via the Ontology and Mapping Management component. Currently, this component includes two installation wizards: basic and advanced. In Figure 1 (*Right*) we depict workflows of the wizards. The basic wizard exploits the relational database metadata and automatically extracts an initial version of the ontology and direct mappings[4] to the ontology entities. The advanced wizard, unlike the basic one, requires the user intervention and an ontology vocabulary as input in order to (manually) create and edit R2RML mappings.[5] Both the basic and advanced wizards provide functionalities to align the bootstrapped ontology with a state of the art domain ontology and approximate the resulting ontology if it is outside the desired OWL 2 QL profile.[6] Alignment is performed using the ontology matching system LogMap,[7] which has shown to work well in practice and also includes mapping repair facilities.

Optique 1.0 is built on top of the Information Workbench[8] (IWB), a generic platform for semantic data management. The IWB provides a shared triple store for managing the assets of Optique 1.0, such as, ontologies, mappings, query logs, (excerpts of) query answers, database metadata, etc. The IWB also provides generic interfaces and APIs for semantic data management, e.g., ontology processing APIs. In addition to these backend data management capabilities, the IWB provides a flexible user interface which follows a semantic wiki approach, based on a rich, extensible pool of widgets for visualisation, interaction, mashup, and collaboration.

Finally, Optique 1.0 is customised for the NPD FactPages, which is a public, freely available dataset created to regulate and overlook the petroleum activities on the Norwegian Continental Shelf (NCS) and contains information collected from a wide range of activities on the NCS, e.g., operating companies, fields, discoveries, facilities, pipelines, and seismic surveys—both historic and current data. Its data has been converted and published as semantic web data [3], of which parts have been fed into the Optique 1.0 system.

---

[3] http://ontop.inf.unibz.it/

[4] http://www.w3.org/TR/rdb-direct-mapping/

[5] http://www.w3.org/2001/sw/rdb2rdf/r2rml/

[6] http://www.w3.org/TR/owl2-profiles/

[7] http://code.google.com/p/logmap-matcher/

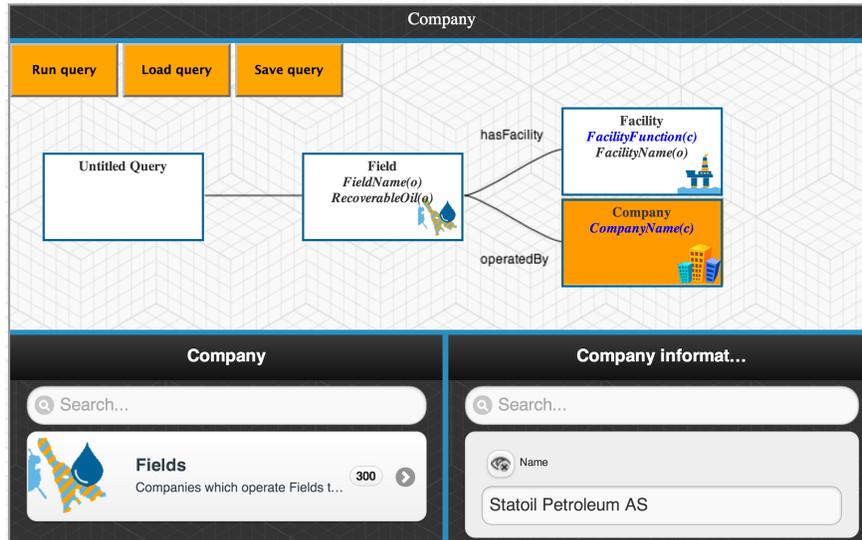[8] http://www.fluidops.com/information-workbench/

Fig. 2. Optique 1.0 System, visual query formulation component

3 Demonstration Details

During the demonstration we will describe the NPD FactPages and present functionalities of the Optique 1.0 system, with the focus on the following aspects: query formulation and execution, and system installation. These aspects will be illustrated on the NPD FactPages data. For the query formulation we will stress our visual query formulation tool that currently supports construction of tree-shaped conjunctive SPARQL queries. The demonstrated queries will be from the oil industry domain. An example query is: *"Find all fields that are operated by 'Statoil Petroleum AS' and which have a facility that produces oil"*; it can be seen in the screenshot of the VQF in Figure 2. We will run queries and present results both in tables and maps, e.g., the location of "Fields" and "Oil facilities" will be displayed on maps. Regarding the system's installation, we will present both basic and advanced wizards and guide through their steps, that is, loading metadata, extraction of an ontology and mappings, alignment with the domain ontology, and approximation of the integrated ontology. We will also show how to edit extracted direct mappings and define new R2RML mappings.

References

1. M. Giese et al. "Scalable End-user Access to Big Data". In: *Big Data Computing*. Ed. by R. Akerkar. Chapman and Hall/CRC, 2013.
2. E. Kharlamov et al. "Optique: Towards OBDA Systems for Industry". In: *ESWC postproceedings volume: Best Workshop Papers*. 2013.
3. M. G. Skjæveland, E. H. Lian, and I. Horrocks. "Publishing the Norwegian Petroleum Directorate's FactPages as Semantic Web Data". In: *The Semantic Web – ISWC 2013*. Ed. by H. Alani et al. Vol. 8219. LNCS. 2013.