ADBIS 2019

Data Integration of Legacy ERP System based on Ontology Learning from SQL Scripts

Chuangtao Ma

machuangtao@caesar.elte.hu

Department of Information Systems,
Eötvös Loránd University

This PhD project was guided by Dr. Molnár Bálint

Outline

- Introduction & Motivation
- Research Question Statement
- Related Work
- Proposed Solutions & Plan
- Conclusion & Future Work

Introduction & Motivation

Legacy ERP system is a kind of enterprise management systems, were developed in several decades ago, that is no longer being enhanced.

Diversity of Code Language & platform

- C++, Java, Php, etc .
- Delphi, IBM AS/400, Visual Studio, etc.

Various DBMS

- Visual FoxPro, Access, SQL Server, Oracle, MySQL, etc.

Unfriendly user Interface

- Character-based user interface.
- Unavailability of the data-access interface.

Outdated Technology

- Obsolete hardware.
- Poor modularity.

Introduction & Motivation

Business
Intelligence

An increasing number of the enterprise decide to build the BI system for responding the dynamic business environment.

01

Data mining & analysis

- Build the centralized data warehousing.
- Integration of the heterogenous data from existing system.

02

Business Process Reengineering(BPR)

- Redesign and improve the business process and architecture of the ERP.
- Improve data dissemination and decision making based on advanced technology.

Introduction & Motivation

Replace all of the existing legacy ERP system

Pros

- Advanced technologies and system.
- Unified and centralized data center .

Cons

- Invest more budget and time.
- Potential risk for upgrading.

Modernize and Integrate the existing legacy ERP system

Pros

- Save the costs, effort, and time.
- Reduce the risk of the project.

Cons

- Limited performance of the BI system.
- Put more effort to integrate legacy system.

How to build BI system If it is lacking enough budget?

Research Question Statement

- It is a challengeable task to integrate the legacy ERP system efficiently and effectively, since the diversity of code language, various DBMS and unavailability of APIs.

Q1: Unified description of Business Process

How to achieve the unified description of business processes (BP) and efficient integration among different sub-organizations?

Q2: Data Integration

How to access and integrate the data from the various DBMS of legacy ERP systems efficiently?

Q3: Result Evaluation

How to check the consistency of the ontologies and evaluate the correctness of the integration results?

Related Work

☐ Data access technology of legacy system

Knowledge Discovery meta-Model (KDM)	Common Business-Oriented Language(COBOL)	Ontology-based Data Access(OBDA)
• Pérez-Castillo, et al, (2011) propose KDM to represent the artifacts of legacy systems as entities, relationships and attributes.	• Millham, R, et al (2009) employed the COBOL to access the data from the relational database of legacy systems.	• Calvanese, D, et al (2017) introduced the OBDA technology to extract the log data from legacy information.

- However, ontology-based data access from distributed data-source still requires the data access interface to be available.

Related Work

☐ Ontology learning & Knowledge Extraction

2009

Extraction algorithm based on Common RDF Model

Extraction algorithm based on RDF (Resource Description Framework) were designed to extract the knowledge from legacy systems.

2011

Process Mining & Sequential Pattern Mining

A dynamic knowledge extraction approach based on process mining and sequential pattern mining are proposed respectively .

2014

Ontology Learning

Semi-automated generation ontology approach from existing textual documents based on ontology learning is proposed .

- However, the knowledge extraction based on ontology learning is still in the early phase to be explored.

Related Work

☐ Semantic data integration

Data integration

The traditional data integration approaches, including , rule-based, middleware framework, and so forth.

Ontology-based semantic integration (OBA-SI)

The heterogeneous data was integrated by the semantic mapping of the ontologies.

Linked data based semantic data integration

A semantic integration approach exploiting linked data are presented to achieve RDF data integration based on query rewriting.

- For the ontology-based semantic integration, the efficiency of the semantic integration is limited by the efficiency and quality of the constructed ontology.

Proposed Solutions & Plan

- This PhD project focus on the integration of legacy ERP system based on ontology learning framework.

Business Process Integration

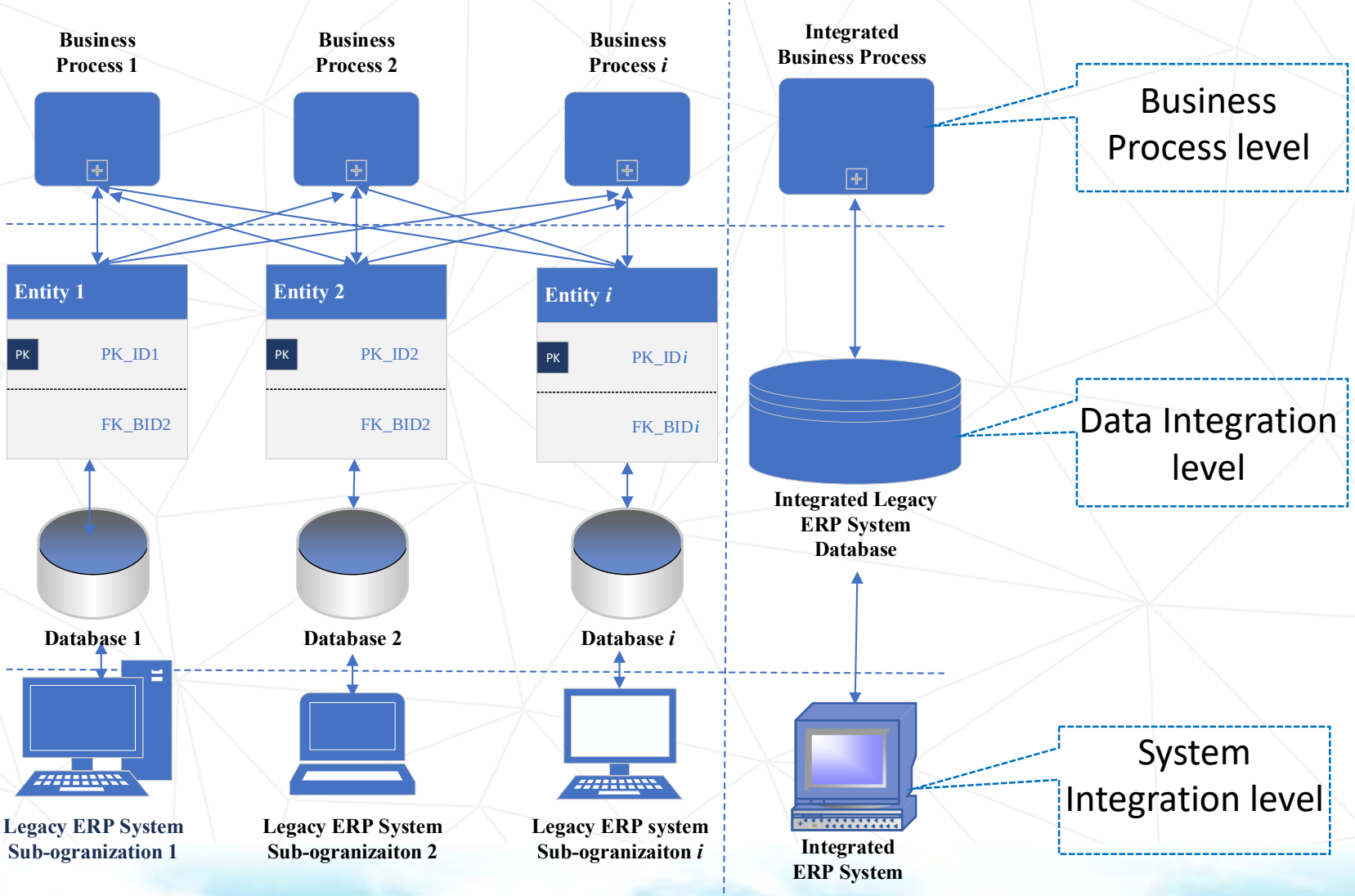
- Unified description of the business process.
- Integration of the business process.

Data Integration

- Ontology learning model from SQL scripts.
- Semantic data integration based on ontology.

Efficient and effectiveness Integration of the legacy ERP systems

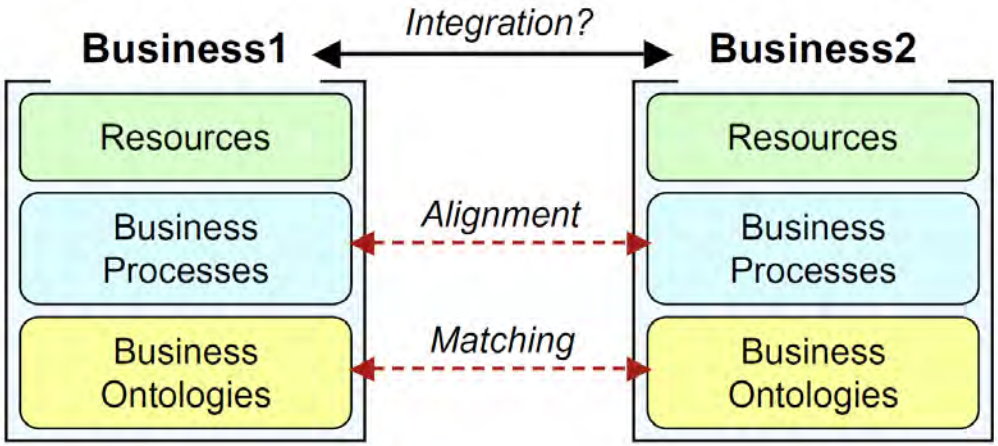
Proposed Solutions & Plan

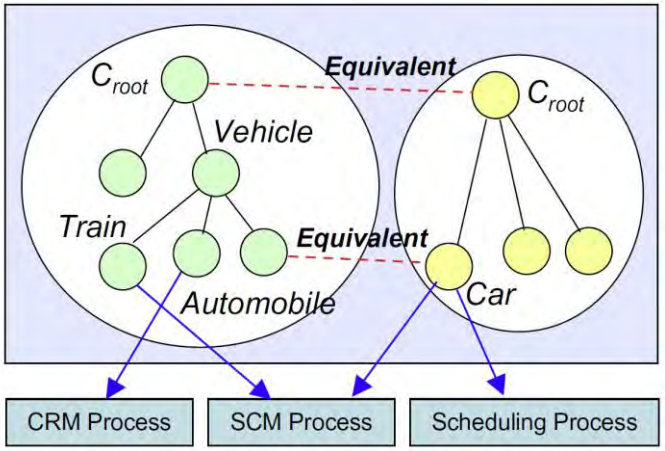


Proposed Solutions & Plan

☐ Unified description & integration of business process

- The ***description logic language*** will be used to achieve the unified description and representation of the business process.
- The ***ontology alignment technology*** will be adopted to achieve the integration of the business process.



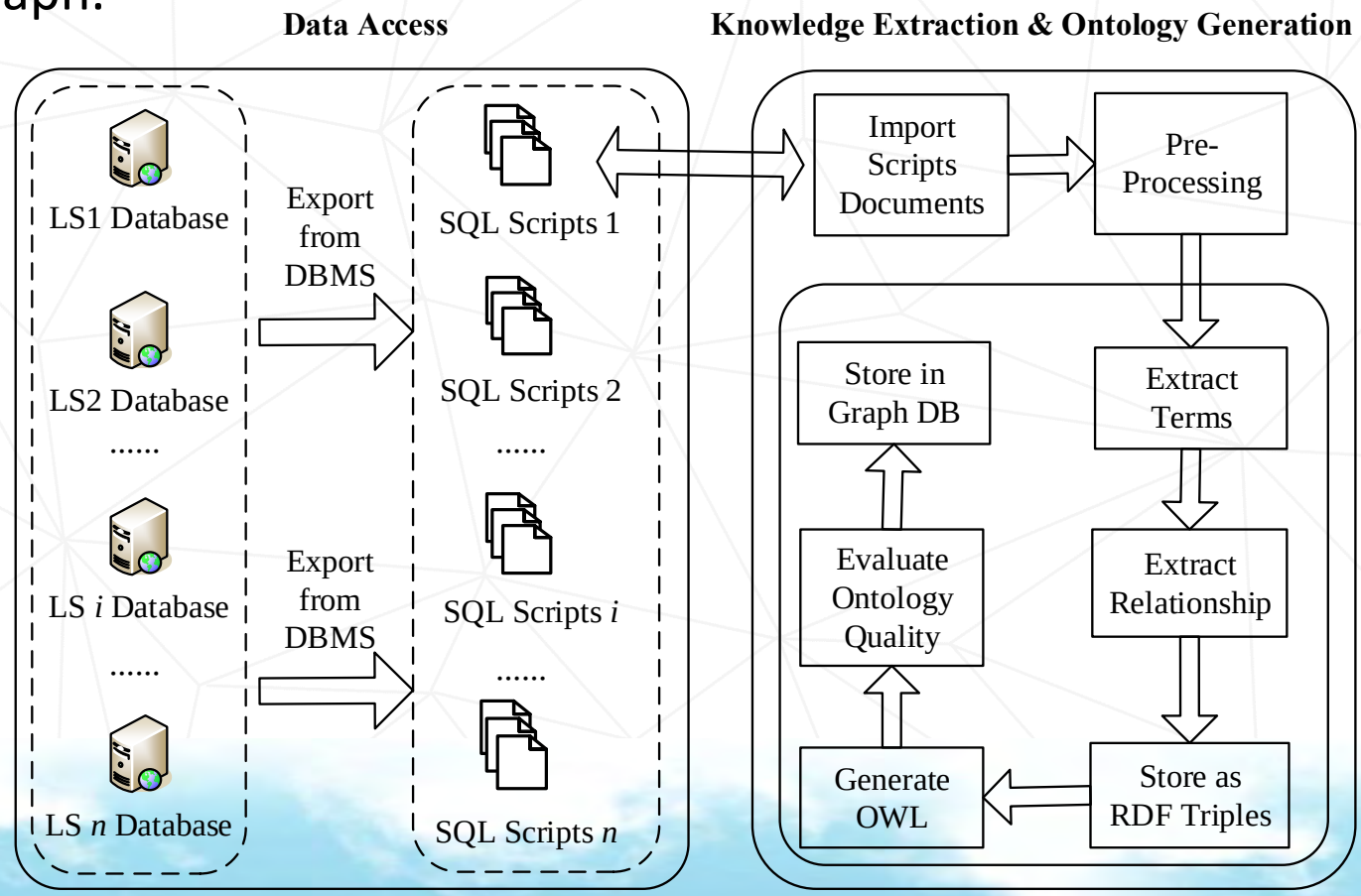


1. Jung, J. J. (2009) 'Semantic business process integration based on ontology alignment', Expert Systems with Applications. Elsevier Ltd, 36(8), pp. 11013–11020. doi: 10.1016/j.eswa.2009.02.086.

Proposed Solutions & Plan

Data integration based on ontology learning

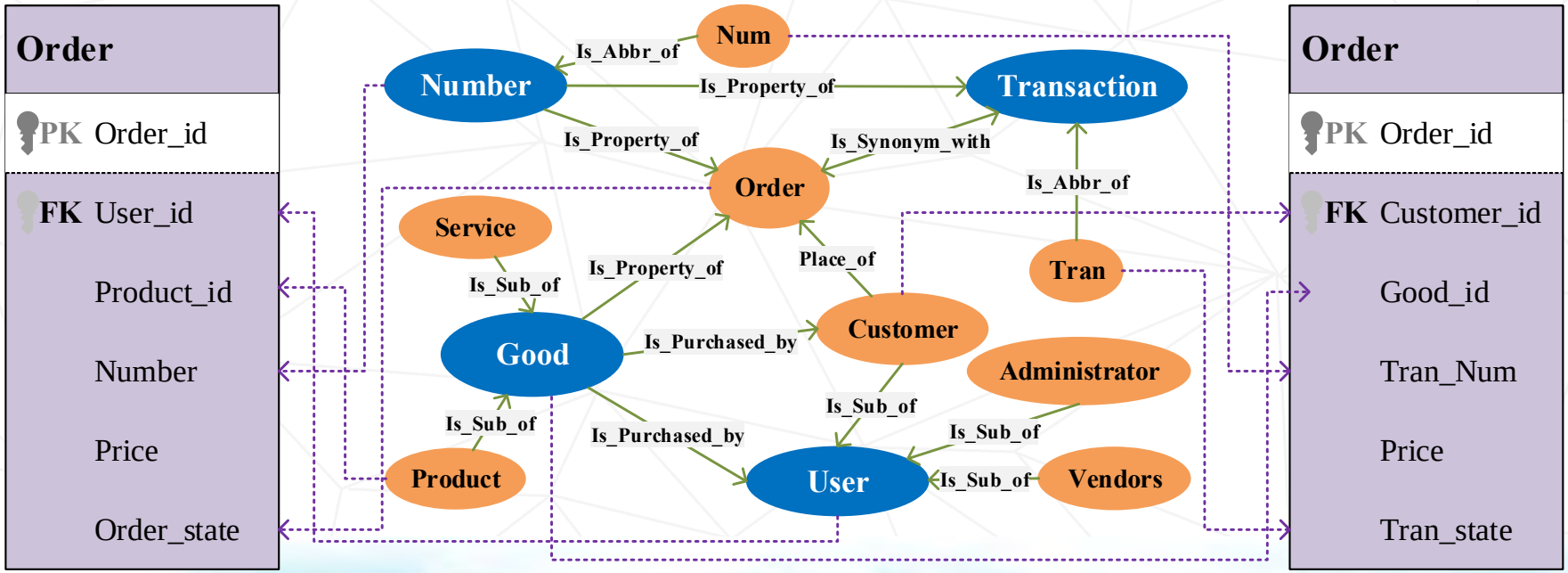
- The input of the ontology learning model is SQL scripts document and the output is the corresponding ontologies and knowledge graph.



Proposed Solutions & Plan

☐ Data integration based on ontology learning

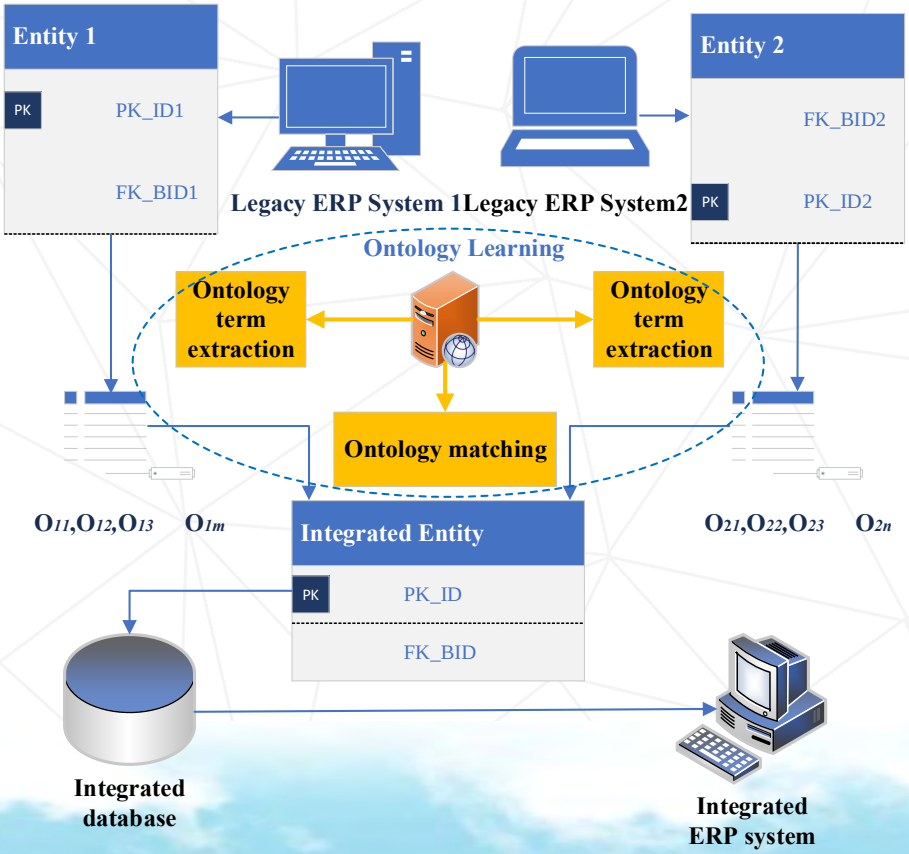
- Ontology-based semantic data integration.
- The heterogenous data will be integrated based on the interoperability of the ontologies and knowledge graph, the specific demo of data integration is depicted.



Proposed Solutions & Plan

☐ Legacy ERP system integration

- Evaluate the accuracy of integration result and integrate legacy ERP system for achieving the centralized management and decision.



- Check the consistency of the ontologies generated by ontology learning from SQL scripts.
- Evaluate the semantic accuracy of the data integration.

Conclusion & Future Work

In this project, the architecture of the ontology learning framework was proposed to integrate heterogenous data from various legacy ERP systems efficiently.

1

The approach for generating ontologies by ontology learning from the relational database SQL scripts is proposed, and the specified steps of the ontology learning are illustrated.

2

Conclusion & Future Work

- This project is in its initial exploration phase, there are some works that should be investigated and conducted in the future.

1 Knowledge extraction algorithm from SQL scripts

- The knowledge extraction algorithm based on NLP will be designed to extract the knowledge from SQL scripts.

2 Ontology generation approaches from RDF schema

- Ontology generation approaches from RDF schema will be studied to generate the ontology for the integration of heterogeneous data.

3 Design the tools to support data integration

- The data integration tool based on ontology learning from SQL scripts will be designed and developed to support the data integration.

I would welcome any question and suggestion.