D5.2.2 Project Progress Report (second half year)

ModelWriter

Text & Model-Synchronized Document Engineering Platform

Project number: ITEA 2 13028
Work Package: WP5
Task: T5.2 - Project Coordination and Reporting

Edited by:

Moharram Challenger <moharram.challenger@unitbilisim.com> (UNIT)

Date: 08-Dec-2015
Document version: 1.0.0

Document History

Version	Author(s)	Date	Remarks
0.1.0	Moharram Challenger	08-Dec-2015	

https://github.com/ModelWriter/Deliverables/issues/126

Table of Contents

1. Introduction

Role of the deliverable

This document is the first version of the project progress report covering 2015 semester 1.

The List of Technical Work Packages

UC Code	Requirements derived from
WP2	Semantic Parsing and Generation of Documents and Documents Components
WP3	Model to/from Knowledge Base (synchronization mechanism)
WP4	Knowledge Base Design and Implementation
WP6	Architecture, Integration and Evaluation

Structure of the document

This document is organized as follows:
- Chapter 1 introduces the document.
- Chapter 2 provides the PPR

Terms, abbreviations and definitions

Abbreviation	Definition
RDF	Resource Description Framework
WP	Work Package
UC	Use Case

Project key data

Project name	13028 ModelWriter
Full length title	Text & Model-Synchronized Document Engineering Platform

The project envisions an integrated authoring environment called "ModelWriter" for Technical Authors (such as Software or Systems Engineers etc.) which will combine a Semantic Word Processor (= the "Writer" part), looking like a usual word processor but capable to "understand" pieces of text and transparently create models of contents out of them; and a Knowledge Capture Tool (= the "Model" part), looking like familiar information modelling tools such as UML, BPMN, ReqIF, etc. ModelWriter will allow Technical Authors to freely move bi-directionally and interactively between text and model to enhance the quality (consistency and completeness) of the technical documents.

Call & project ID	ITEA 2 Call 8 - 13028
Time frame	start: 01-10-2014 end: 30-09-2017
Size	PY: 60.07 M€: 4.2
Project leader	Ferhat Erata (UNIT Information Technologies R&D Ltd.)
Involved countries	Belgium, France, Turkey

Belgium	Philippe Bureille (Sogeti Belg...	13 PY
France	Etienne Juliot (OBEO)	14 PY
Turkey	Aydin Can Polatkan (Mantis)	34 PY

PCA status	PCA has has not been signed yet
Project page	13028 ModelWriter
Latest FPP	Change Request (31-03-2015)
Latest PPR	Progress report in 2015 (semester 1)
Latest review	
Next review	ModelWriter #1 (a.m.) (24-09-2015)

STG evaluation

Submitted:

STG Reviewers

Project acronyms

KB (Knowledge-base), UC (Use Case), MW (ModelWriter), ALM (Application Lifecycle Management), EMF (Eclipse Modeling Framework), RDF (Resource Description Framework), QDMS (Quality Document Management System), MBSE (Model Base Software Engineering), BAFLING (Back and Forth Linguistic Processing), DL (Description Logic), Req. (Requirement), FORL (First Order Relational Logic)

Top 4 overall targeted innovations

1	Capability to maintain a readable textual document (using an editor) and relate its content to existing models' elements
Main contributors	Obeo, UNIT, LORIA
State-of-the-Art	> There are some Document annotation systems > A new capability is to annotate a text using an ontology > A new Recommendation system is addressed (synchronization links automatically proposed)

Top 4 overall targeted business impacts

1	MBSE development
Main contributors	All
Market / competitors	Challenge in MBSE development is how to maintain the coherence between multiple distributed models or between models and documentation.

#		Left column
2		Model / Text Synchronization Engine with iterative and interactive matching synchronization
	Main contributors	Obeo, UNIT, KocSistem
	State-of-the-Art	> Hand made synchronization available > Only doc generation or reverse engineering exist without interaction >There is no platform which extensively supports sync. between technical texts, models and knowledge base. Also, there is no Eclipse based platform capable of sync. text or model with a KB.
3		Formal Specification and Verification of Semantic Relationships between software and system artefacts
	Main contributors	UNIT
	State-of-the-Art	Formal specification and verification techniques have been used widely in industry, especially in the development of safe and secure systems. However, their use in the specification for the construction of semantic relationships between software/system artefacts need further investigation which will be addressed.
4		Semantic Annotation of Text with Model Elements
	Main contributors	CNRS/LORIA, Obeo, Airbus
	State-of-the-Art	Semantic Annotation is mostly restricted to the annotation of text with concepts or relations. ModelWriter will investigate how to make these approaches more generic and extend them to more complex items such as axioms and cardinality restrictions.

#		Right column
2		Reducing time to spend for Quality Control activities, by this way it provides manufacturers faster production
	Main contributors	HISBIM
	Market / competitors	QA QDMS, Microsoft Pinpoint QDMS
3		Expertise on document extraction
	Main contributors	OBEO
	Market / competitors	Some professional service will be proposed for these purposes: - creation of dedicated engines to detect structured data in documentation - creation of new connectors to support modeling designers consulting for usage and installation - customization of ModelWriter core feature for specific needs
4		More sell of Obeo Designer and Obeo SmartEA
	Main contributors	OBEO
	Market / competitors	> Support and Maintenance option in the contract "Obeo Designer" to support ModelWriter > Collaboration feature which support ModelWriter links with Obeo Designer Team > New feature in Obeo SmartEA to add documentation synchronization for strategic transformation and governance analysis

#		Left column
5		Reversible Semantic Processing
	Main contributors	CNRS/LORIA
	State-of-the-Art	Semantic parsing maps text to semantic representations. Natural language generation maps semantic representations to text. While these two processes are standardly treated independently, we aim to produce a system that both parse and generate a simple interaction between text-to-model and model-to-text conversion.
6		Natural Language Generation from KB Data
	Main contributors	CNRS/LORIA
	SotA	Existing approaches of generating from KBs generally use templates thereby yielding stilted text. There are also no approaches that can generate fluent text from arbitrary KBs. We aim to produce natural sounding text from Kbs using a grammar based approach.
7		Capability to ensure / manage the synchronization between the artefacts (documents and models) bi-directionally
	Main contributors	Obeo, UNIT
	State-of-the-Art	A synchronization link management is addressed as a synchronization engine.
8		Design and implement of a heterogeneous knowledge representation approach, which allows combination & migration between different representations in MW
	Main contributors	MANTIS

#		Right column
5		Supporting Traceability in ALM, e.g. by addressing impact analysis in the procedure of application development
	Main contributors	UNIT, Havelsan
	Market / competitors	> No similar offer in TFS / none announce
6		Synchronization between use case documents and business process models
	Main contributors	KocSistem, UNIT
	Market / competitors	The current tools in the market do not fully support automatic update on the process models when a change occurs in the use cases and vice versa.
7		
	Main contributors	
	Market / competitors	
8		
	Main contributors	

	State-of-the-Art	Representation of knowledge can vary in different abstraction levels. When a text like system description or req. definition is modeled as a single model that defines different aspects, there needs to be a coherence between definitions. MW will provide this coherence by an ontology used at different abstraction levels.

	Top 4 overall KPIs		Current	Target
1	Workload need to setup ModelWriter a new context		N/A	+ROI in 3 Months
	Metric description	Metric Description: Time to install ModelWriter, add new connector for existing tools and models, create the NLP resources, training of end users. This setup should made by non ModelWriter core developers. Note: this target is really dependent of the number of connectors, of ontologies, diversity of documents and tools. By using this target criteria, it will scale for small and large projects.		
2	Quality and precision of automatic synchronization detection		N/A	90% correct links
	Metric description	Number of synchronization links accepted by a user over number of synchronization links automatically proposed. Nielsen, Jakob (1990). Ten Usability Heuristics. (voir la liste plus bas) http://www.useit.com/papers/heuristic/heuristic_list.html Bastien, J.M.C., Scapin, D. (1993) Ergonomic Criteria for the Evaluation of Human-Computer interfaces. Institute National de recherché en informatique et en automatique, France (http://www.inria.fr)		
3	Automatic synchronization links number compared to manual synchronization links		N/A	30%
	Metric description	The number of manual links is based on the typical precision done by a user (while keeping a ROI). The automatic links counted are only the correct links. It will be useful to measure the missing synchronization links compared to manual annotation references.		
4	The performance of document related task from the end user point of view		X time	X/5 time
	Metric description	For Airbus case, the goal is to reduce the time of checking consistencies of a SIDP document at least 5 times in average.		
5	Number of supported Domain Specific notations; Supporting various user-visible modeling languages		0	>=2
	Metric description	At least two languages need to be supported such as: Ecore, OWL, Alloy and ReqIF		

	Market / competitors	

	Top 4 overall risks	S	P
1	Technical difficulty to deploy and integrate within existing frameworks/platforms	H	L
	Avoidance action	To consider most used technologies for both DSM and document edition	
	Back-up / mitigation plan	We have several frameworks as our target and if one of them has problem with integration, we will focus on the alternative one(s).	
2	Low performance and scalability	H	M
	Avoidance action	To create as early as possible a large data set to test the tool and its scalability, e.g. to test with real SIDP documents and models, typically more than 150 pages and 500 elements respectively.	
	Back-up / mitigation plan	We can change the architecture to client server and use background computation. Additional visualization techniques can be used to support the scalability of user interface, such as filtering, having different viewpoints and so on.	
3	Annotations and markers should be resistant to modification of input documents	M	M
	Avoidance action	This is the main technical challenge. e.g. if the user cut/paste a large paragraph, the link should be kept. Avoidance: This challenge need to be as the heart of the initial design of the KB.	
	Back-up / mitigation plan	If the engine can't keep automatically the link, an interactive UI should help the user to massively reconnect these links based on previous mapping. We could use Eclipse facilities to keep previous states of documents, markers. links. etc.	
4	Lack of data inside the consortium prevents the training of high quality Natural Language Processing Tools	M	M
	Avoidance action	Use external data	
	Back-up / mitigation plan	Demonstrate usefulness of the NLP tools on data that is external to the project (e.g., RDF data from DBPedia dn text data from WikiPedia).	
5	Semantic annotation is highly ambiguous and yields many possible annotations for a single text/model fragment (LORIA)	H	M
	Avoidance action	Apply disambiguation techniques (e.g., Lesk algorithm for word sense disambiguation)	
	Back-up / mitigation plan	Allow for interactive annotation where the user manually provides the required disambiguation	

6	Faster QDMS reporting/generating	2 per day	10 per day		6	The definition of KB is still in early stage and the API for the semantic services is not defined	M	M
	Metric description	In HISBIM pilot factory, Quality control department personnel is able to report 1 or 2 QDMS documents in a day. With the ModelWriter system 10 QDMS documents generation is expected.				Avoidance action	Mantis should work on this architecture and API for the next year.	
						Back-up / mitigation plan	The model / text synchronization can work without this semantic services connection.	
7	Covering different textual representatives in the project	0	>=2		7	Gendoc project has some features which overmap MW	M	M
	Metric description	Structured texts and semi structured texts need to be supported, e.g. Java code, Mark down, and so on.				Avoidance action	Discuss with Gendoc committer to align MW and Gendoc to avoid this overlapping.	
						Back-up / mitigation plan	Fork Gendoc or do everything in MW.	
8					8			
	Metric description					Avoidance action		
						Back-up / mitigation plan		

Changes in the technological and business relevance during the reporting period

> The main technical change during this period is that the technical writers who are using MS word are also addressed as ModelWriter's end users by developing a MS word processor plug-in in the scope of the project.
> Considering the Business relevance change during the period is the participation of Ford-Otosan, a large automotive industry in Turkey, in ModelWriter as a partner without fund. This will also improve the exploitation of the project results.

Project statement on progress during the reporting period

The first versions of the main components of the project are developed including: a semantic annotator, a semantic parser and a text generator, model and text synchronization, formal specification and configuration of the framework, visualization and consistency check of semantic relationships, and the graphical user interfaces. At the moment the integration plan is completed and the technical integration procedure is started for the 1st release.

Exploitation

Updates to partners' exploitation prospects

Partner	Type	Country	PY	Partner	Type	Country	PY
UNIT Information Technologies R&D	sme	TUR	11 PY	Havelsan	ind	TUR	0 PY
Mantis	sme	TUR	8 PY				
Hisbim Bilgi ve İletişim Teknolojileri	sme	TUR	7 PY				
Sogeti Belgium	ind	BEL	7 PY				
Centre National de la Recherche	res	FRA	8 PY				
KoçSistem	ifc	TUR	7 PY				
Katholieke Universiteit Leuven	uni	BEL	6 PY				
OBEO	sme	FRA	4 PY				
Airbus Group SAS	ifc	FRA	2 PY				
Ford Otosan	ind	TUR	0 PY				

Other updates

Top 8 overall partners' Exploitation Related Achievements

#					Status
1	Exploitation	New product	Implementation of the MW plug-in for MS Word		Planned
2	Exploitation	Enhancement	System Installation ontology_v1		Planned
3	Dissemination	Workshop	The 6th International ModelWriter Workshop in		Planned
4	Exploitation	New system	Exploitation of ModelWriter in ITEA3-ASSUME	T4B	Realised
5	Dissemination	Workshop	The 5th International ModelWriter Workshop		Realised
6	Exploitation	New product	CSV to OWL transformation program		Realised
7	Exploitation	Enhancement	Requirement Documents <-> ReqIF Standard	T4B T4I	Realised
8	Exploitation	Collaboration	Participation of FORD Otosan	T4B	Realised

Realized Exploitation Related Achievements statistics

Dissemination	Exploitation	Standardisation	New company	Patent	Human capital
Total: 10	Total: 10	Total: 5	Total: 0	Total: 0	Total: 0

Work progress during the reporting period

Top 4 technical achievements

1 Integration of the Semantic Annotator developed by LORIA/CNRS into the ModelWriter Prototype

Details: The annotator automatically produces links between text fragments and model elements.

2 Reversible Semantic Processing

Details: LORIA developed a prototype called BAFLING for parsing and generating text which integrates a semantic parser and a generator. The parser maps text to DL formulae and the generator generates text from DL formulae. The prototype was tested on Airbus data.

3 Formal specification of the key semantic relationships between software & system engineering artefacts

Details: A component for formal specification of the key semantic relationships between requirements, software & system engineering artefacts has been developed by UNIT.

4 Formal verification of the semantic relationships

Details: A component is developed to formally verify the semantic relationships specified between software and system artefacts by UNIT.

5 Visualization of semantic relationships

Details: KocSistem contributed in the development of the graphical user interface of ModelWriter Eclipse platform and in the implementation of the visualization component.

6

Details

7

Top 4 next technical targets

1 Extension of BAFLING to Airbus Data

Details: The prototype developed by LORIA will be extended to handle all normalized rules.

2 Using BAFLING for Synchronization

Details: The output of the semantic parser (a DL formula) will be compared to the existing model (KB) used by Airbus. Ontology enrichment techniques will be used to either add, remove or modify existing sync. links between model elements and text fragments.

3 Development of the knowledge base meta model

Details: The knowledge base meta model will be developed that can be used by all use cases defined in the project scope. The metamodel will serve as the backbone for all use cases aiming at model to text, text to model transformations and model synchronization.

4 Automated consistency checking

Details: UNIT aims to apply automated consistency checking on the formal specification defined on the platform.

5 ModelWriter plug-in for MS Word processor

Details: UNIT and HISBIM has begun to develop a MW plugin for MS Word which will enable a seamless integration with MW on Eclipse platform. User interface is almost ready, and now working on integration of Eclipse platform and Word plug-in.

6 Enhancing the visualization of semantic relationships

Details: KocSistem will improve the visualization to meet different requirements imposed by use cases.

7

	Details	
8		
	Details	

<table>
<tr><td colspan="2">Top 4 next technical targets from latest PPR</td></tr>
<tr><td>1</td><td>Conception and Implementation of a Semantic Parser</td></tr>
<tr><td>2</td><td>Conception and Implementation of a Text Generator</td></tr>
<tr><td>3</td><td>ModelWriter will be able to generate standard QDMS reports</td></tr>
<tr><td>4</td><td>A synchronization prototype with UI and engine together</td></tr>
</table>

	Top 4 issues		Impact	
1	Robustness of Semantic Parser		Lack of robustness may result in incorrect or missing synchronization links	
	Details	The same semantic content may be expressed in different ways. Additionally, text may be ill formed because of typos or grammatical mistakes.	Mitigation action	The semantic parser will integrate robustness mechanisms which allows (i) different formulation of the same content to be mapped to the same DL formulae and (ii) ill formed input to be handled.
2	Well formedness		ill formed text may be rejected by the end user	
	Details	While the parser must be robust to ill formed input, the generator should only generate well formed text.	Mitigation action	The generator will integrate well formedness constraints to ensure that its output is grammatical.
3	Separated use case related works		A meta model which is not generic enough for the project	
	Details	One issue we are having currently is that use case related works are realized separately without having an overall picture that utilize a knowledge base model that can be used globally.	Mitigation action	Once the use case activities are more mature we plan to come up with knowledge base sample usages that can be generalized into a knowledge base meta-model.
4	Scalability of state of the art reasoner engines		Reasoning on big models can be in-efficient	
	Details	The recent experiments show that the models with elements around 10000 items can be accepted in a reasonable time. Based on the provided use cases, it seems that the supported size is acceptable. However, the bigger models might be an issue.	Mitigation action	To mitigate this issue, we are planning to propose a new decision procedure which will incrementally use underlying backend reasoners.
5				
	Details		Mitigation action	
6				
	Details		Mitigation action	
7				
	Details		Mitigation action	

8				

Details					

Deliverables (overall status)		Due	Total
Number of deliverables (due / total)		50	112
Already finalized	42	84%	37%
Delayed (> 2 months) on due	8	8	100%

Mitigation action	

Actual-vs-planned (overall status)		Current	Total	%
Time elapsed (months)		15	36	42%
Planned effort consumption (PY)		23.5	60.1	39%
Reported actual effort (PY)	0%	20.7	61.1	34%

Details The project has a large number of deliverables which are inherited from the former leader of the project and their management is difficult due to their exhaustive number. As the STG also suggested in the previous PPR, we have provided a new change request to merge some of the deliverables. This new change request will be submitted to the ITEA to be considered after the voting of partners for removing Belgium partners (as their fund has not been confirmed yet). Anyway, 8 deliverables are delayed including: D6.5.1-1 , D6.5.2-1, D6.6.1-1, D4.6.2-1, D6.7.1-1. All of the above-mentioned five deliverables are delayed, since, they are planned for the 1st release of the project and the 1st release is ongoing now and will be finalized in early April 2016. Also, deliverables D4.5.1, D4.7.1, D4.3.1-1 are delayed due to the fact that the use case related works are realized separately without having an overall picture that utilizes a knowledge base model that can be used globally. As the use case activities are becoming more mature, we plan to come up with a knowledge base that can be generalized into a common meta-model. Then, we will finalize these deliverables.

Discrepancies explanation (partners) > Ford Otosan (-%): Since the participation of Ford Otosan was officially approved to begin with the year of 2016 by the upper management, no effort had been shown within this semester reporting.
> Centre National de la Recherche Scientifique (-53%): A PhD student employed on the project left for personal reasons. It took a while to recruit new staff but we eventually recruited 2 new persons on the project so that the budget should balance out in the end.
> Mantis (-33%): Due to slow start of activities related to development of knowledge base metamodel, we were not able to commit all the efforts planned. We plan to catch up in the following periods. Currently, all the partners are trying to carry out some implementations for their own use cases. We decided that we will merge the knowledge base usages at later stages of the project.

Delayed (> 2 months) on total	70	8	11%

Project technical progress (%)	88%	102%	30-40%

Details We are not expecting any cascading delay for the other deliverables.

Comments The first version of the main components are developed such as:
> a semantic annotator,
> a semantics parser and a text generator,
> writer enhancement including markers and mapping/link;
> formal specification and configuration of the framework,
> visualization and consistency check of semantic relationships,
> the graphical user interfaces.
The 1st integration is ongoing now and the first release is planned for the next month.

STG feedback on work progress

Feedback to previous STG remarks	STG responses

Considering the reviewers' remarks in the previous STG, the KPIs are now updated and all of them have current and target values along with more explanations. Regarding the risk analysis, proper mitigation actions are provided for the risks. For targeted business impacts, they are updated to state how the target innovations are employed by the partners to compete in the market. Finally, in this PPR, the technical achievements during the reporting period and the goals of the next technical targets are explained clearly and in detail. In addition to the STG remarks of previous PPR, the comments of the reviewers as the conclusion of the 1st project review are considered and related actions are taken by the partners. These actions and their results will be reported in the next project review meeting which is planned to take place in Istanbul on 29th Sep. 2016.