D5.2.1 Project Progress Report (first 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:

Ferhat Erata <ferhat.erata@unitbilisim.com> (UNIT)
Moharram Challenger <moharram.challenger@unitbilisim.com> (UNIT)

Date: 07-June-2015
Document version: 1.0.0

Document History

Version	Author(s)	Date	Remarks
0.1.0	Ferhat Erata Moharram Challenger	07-June-2015	Draft

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

ITEA Office
High Tech Campus 69 - 3
5656 AG Eindhoven
The Netherlands

T + 31 88 003 6136
E info@itea3.org
W www.itea3.org

ITEA 3 is a EUREKA strategic ICT cluster programme

2. The Project Progress Report (2015 Semester 1)

Project key data				
Project name	13028 ModelWriter			
Full length title	Text & Model-Synchronized Document Engineering Platform			
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			
PCA status	PCA has not been signed yet			
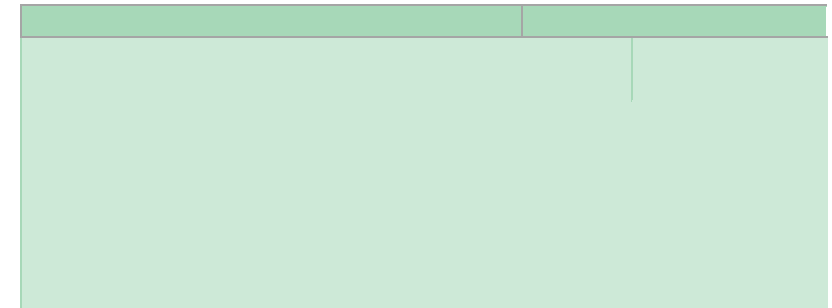
Project acronyms
KB (Knowledge-base), UC (Use Case), MW (ModelWriter), ALM (Application Lifecycle Management), EMF, RDF, QDMS (Quality Document Management System), MBSE (Model Base Software Engineering)

Top 4 overall targeted innovations				Top 4 overall targeted business impacts	
1	Capability to create a readable textual document (using an editor) and relate its content to existing models' elements		1	MBSE development	
	Main contributors	Obeo, UNIT, Loria		Main contributors	All
	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)		Market / competitors	Challenge in MBSE development is how to maintain the coherence between multiple distributed models or between models and documentation.
2	Capability to model some content graphically using a domain-specific notation and then generate some readable but structured text/ document		2	Reducing time to spend for Quality Control activities, by this way it provides manufacturers faster production	
	Main contributors	Obeo, UNIT		Main contributors	HISBIM
	State-of-the-Art	Preliminary template based document generation approaches are available		Market / competitors	QA QDMS, Microsoft Pinpoint QDMS
3	Capability to ensure / manage the synchronization between the artifacts (documents and models) bi-directionally		3	Expertise on document extraction	
	Main contributors	Obeo, UNIT		Main contributors	OBEO

	State-of-the-Art	A synchronization link management is addressed as a synchronization engine.
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 Knowledge or Database elements. ModelWriter will investigate how to make these approaches more generic and extend them to arbitrary models.
5	Reversible Semantic Processing	
	Main contributors	CNRS/LORIA
	State-of-the-Art	Semantic parsing maps text to semanticrepresentations. Natural language generation maps semantic representations to text. As 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.
7	Model / Text Synchronization Engine with iterative and interactive matching synchronization	
	Main contributors	Obeo, UNIT, KocSistem
	State-of-the-Art	- Handmade synchronization available - Only doc generation or reverse engineering exist without interaction. - There is no platform which extensively supports synchronization between technical texts, models and knowledge base. Also, there is no Eclipse based platform capable of synchronizing text or model with a KB.
8	Design and implement of a heterogeneous knowledge representation approach, which allows combination and migration between different representations in MW	

	Market / competitors	
4	New Open Source projects with professional service	
	Main contributors	OBEO
	Market / competitors	
5	More sell of Obeo Designer and Obeo SmartEA (OBEO)	
	Main contributors	OBEO
	Market / competitors	
6	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
7	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.
8		

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

	Top 4 overall KPIs		Current	Target	
1	Time to implement a new case				
	Metric description	Average time to implement a new case, including development of required NLP of Knowledge resources (such as ontology, NLP resources, and so on).			
2	Usability & automation degree				
	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	Number of supported Domain Specific notations; Supporting various user-visible modeling languages		0	>=2	
	Metric description	At least two of the following languages need to be supported: ADL, UML, ORB and BPMN			
4	Faster QDMS reporting/generating		2 per day	10 per day	
	Metric description				

	Main contributors	
	Market / competitors	

	Top 4 overall risks	S	P
1	Technical difficulty to deploy and integrate within existing framework/platforms	H	L
	Avoidance action	To consider most used technologies for both DSM and doc edition	
	Back-up / mitigation plan	Ability to offer support	
2	High complication and/or poor performance	H	M
	Avoidance action	Well identified capabilities level of automation	
	Back-up / mitigation plan		
3	ModelWriter Generated QDMS Reports may be inappropriate to vendor customer requirements	L	H
	Avoidance action	System development should be interactive with vendor customer quality department	
	Back-up / mitigation plan	If it is not possible, outsourcing quality experts who worked in a big vendor company at the past.	
4	Lack of data inside the consortium prevents the training of high quality Natural Language Processing Tools	M	M
	Avoidance action	Use external data	

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

| 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. | | 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	Number of different Open Source users detected on newsgroups	0	20		5	Semantic annotation is highly ambiguous and yields many possible annotations for a single text/model fragment (LORIA)	H	M
	Metric description					Avoidance action	Apply disambiguisation techniques (e.g., Lesk algorithm for word sense disambiguisation)	
						Back-up / mitigation plan	Allow for interactive annotation where the user manually provides the required disambiguisation	
6	Number of operational projects which is used in MW	0	6		6	The definition of KB is still in early stage and the API for the semantic services is not defined	M	M
	Metric description					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	Unique visitor on MW webpage per month	0	800		7	Annotations and markers should be resistant to modification of input documents	M	M
	Metric description					Avoidance action	It has been identified as a top challenge	
						Back-up / mitigation plan		
8	Covering different textual representatives in the project	0	>=2		8	Gendoc project has some features which overmap MW	M	H
	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.	

Changes in the technological and business relevance during the reporting period
One option during writing the FPP was to reuse all the existing code of the Open Source project called Intent. This option is no more adapted to the expectation outcomes of MW. We re-restructured the architecture and the new one is more reliable. The first prototype seems to validate our first choices and is positive.

Project statement on progress during the reporting period
A good architecture design has been setuped. A VERY good operational collaboration has occurred between Obeo and UNIT developers, which produces already several interesting Proof of Concept and a real sharing of knowledges. Thanks to knowledge acquired with ModelWriter, Obeo has sold 2 commercial contracts for professional services to extract any information from a document.

Updates to partners' exploitation prospects				
UNIT Information Technologies R&D Ltd.	sme	TUR	11 PY	
Mantis	sme	TUR	8 PY	
Sogeti Belgium	ind	BEL	7 PY	
Katholieke Universiteit Leuven	uni	BEL	6 PY	
OBEO	sme	FRA	4 PY	▮
Airbus Group SAS	ifc	FRA	2 PY	
Havelsan	ind	TUR	0 PY	▮

Top 8 overall partners' Exploitation Related Achievements					
1	Standardization	Open Source Software	Specification & Verification of ALM Platform	T4B T4I	Planned
2	Standardization	Open Source Software	Change Impact Analysis & Visualization	T4B T4I	Realized
3	Exploitation	Collaboration	Collaboration between UNIT and HAVELSAN	T4B	Realized
4	Dissemination	Workshop	The 2nd International ModelWriter Workshop		Realized
5	Standardization	De facto standard	System Installation Component Ontology	T4I	Realized
6	Dissemination	Publication	Parsing Text into RDF		Realized
7	Standardization	Open Source Software	Semantic Annotator	T4B T4I	Realized
8	Standardization	Open Source Software	Synchronization Engine Prototype	T4B T4I	Planned

Realized Exploitation Related Achievements statistics					
Dissemination	Exploitation	Standardization	New company	Patent	Human capital
Total: 7	Total: 3	Total: 3	Total: 0	Total: 0	Total: 0

Work progress during the reporting period

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

Top 4 technical achievements			Top 4 next technical targets		
1	Semantic Annotation		1	Conception and Implementation of a Semantic Parser	
	Details	A semantic annotator is generated which annotates text with model elements. In Airbus case, texts are System Installation Design Principles and the model is an OWL KB. In Obeo case, code documentation is annotated with Ecore concepts and Java procedure names.		Details	The mapping between text fragments and model elements produced by the semantic annotator aims to develop a semantic parser which can map text to model fragments. As the first step, we will test the parser on "simple texts" such as the semi-structured SIDP rules provided by Airbus.
2	QDMS system integration via plug-in		2	Conception and Implementation of a Text Generator	
	Details	ModelWriter system communicates with QDMS by XML files		Details	Using the mapping between text fragments and model elements produced by the semantic annotator, the aim is to develop a text parser which can maps model to text fragments. In a first step, we will test the generator on "simple input such as sets of RDFS triples.
3	Technical Architecture Design		3	ModelWriter will be able to generate standard QDMS reports	
	Details			Details	ModelWriter system might be automatically fill QDMS documents according to user basis inputs as short terms/keywords
4	First prototype of Model / Text synchronization		4	A synchronization prototype with UI and engine together	
	Details			Details	

5	Prototypes to select a reliable library for document parsing		5	First alpha release of ModelWriter integration	
	Details			Details	All of the model editor, writer part and knowledge base will be integrated.
6	Connectors for external writer part		6	Knowledge-based complete design	
	Details			Details	
7	Markers and their classification		7	Test plan and acceptance test	
	Details			Details	
8	Mapping of the markers		8		
	Details	The first version of Mapping is realized with regarding type of the relation for that mapping.		Details	

Top 4 next technical targets from latest PPR

1	Not application; this is the first review of ModelWriter and there is no last PPR

Top 4 issues			Impact	
1	Confidential Data delayed at the start of work on WP2		Work on the Conception and Implementation of the semantic annotator was delayed	
	Details	The Airbus data provides an interesting testbed for semantic parsing and text generation. However using it required the signature by the French partners of a Non-Disclosure Agreement which could only be finalized in June 2015.	Mitigation action	LORIA started work on a restricted dataset and implemented a first question/answering prototype using this dataset whose components were described in a conference paper. This allowed AIRBUS to develop a full blown version of the approach on their data.
2	Uncertainty about Belgium Participation means that work on Semantic Parsing was delayed		Work on Semantic Parsing was restricted	
	Details	The Belgium partners were responsible for work on Semantic parsing. This work was delayed to the second year.	Mitigation action	CNRS/LORIA investigated a first rule-based approach to semantic parsing which allowed for a first analysis of the linguistic issues raised by the data. Further work will build on automatic, statistical approaches.
3	Two ModelWriter technical leaders left Obeo		Knowledge transfer has been successfully operated	
	Details		Mitigation action	
4	Scope of Intent was mainly authoring which isn't the scope of MW		The architecture specification of MW helped to remove the part of Intent and to reorganize the project	
	Details		Mitigation action	

5	Space application drop		Losing the leadership at the beginning of the project	
	Details		Mitigation action	UNIT has accepted the leadership
6	Belgian consortium have not been funded yet		Some of the tasks are left without owner.	
	Details		Mitigation action	Re-allocation of tasks
7	Need for more large industries in Turkey for exploitation		Exploitation would be limited	
	Details		Mitigation action	- Havelsan participated as a large-scale use case provider - Ford-Otosan is under negotiation for participation in ModelWriter.

Deliverables (overall status)		
Total number of deliverables	134	
Already submitted	60	45%
Delayed (< 3 months)	3	2%

Details	The project has a large number of deliverables which are inherited from the previous leader of the project (Space Application) and their management is difficult due to their exhaustive number (some of the small tasks are defined as a separated deliverable). We may merge some of them and propose to ITEA in the next change request. Anyway, the only deliverables which are delayed are: - D6.5.1-1 Acceptance Test Plan (release 1) - D6.5.2-1Automated Acceptance Tests(release 1) - D6.4.3-1 IDE-integrated User Interface to handle Sync issues (major release 1) All of the abovementioned three deliverables are due at Month 12 of the project (Sep. 2015) and are delayed 1 month. The reason is that as their title indicates, they are planned for the 1st release of the project and the 1st release is postponed to month 16 of the project. Please note that all of the accomplished deliverables including software and document ones are considered in the submitted item. Also, please note that some of the deliverables are their first version and will be updated later.	
Delayed (> 3 months)		0%
Details		

Actual-vs-planned (overall status)			
Time consumption (months)	9	36	25%
Planned effort consumption (PY)	13.4	60.1	22%
Actual effort consumption (PY)	12.5	60.1	21%

Discrepancies explanation (partners)	> Centre National de la Recherché Scientifique (-55%): The first candidate identified for recruitment cancelled unexpectedly which lead to some delays in recruiting a PhD candidate to work on the project. > OBEO (-70%): The planned 2015 effort is not linear: we plan to work more on the end of the year than in the 1st semester. We also prefer to move some effort from 2015 to 2016: as the initial scenario of reusing Intent engine has been modified by recreating from scratch a new engine, we prefer to parallelize development after creating a strong core and having a clear idea of architecture and specification.
Project technical progress (%)	20-30%
Comments	Some of the components are developed such as Semantics parsing of the text; writer enhancement including markers and mapping/link (with their configurations); model-to-text linking and vice versa; However, some of the planned tasks for the first year of the project were not completed such as KB design and acceptance test due to the delay in the start of the project and the first major release is postponed to month 16 of the project.

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